TRANS

Boundaries / Permeability / Reification

Proceedings of the Australasian Computer Music Conference 2007

The Australian National University
Centre for New Media Arts
The ANU School of Music
Canberra, Australia

19th – 21st June 2007

australasian computer music association

NCH Swift Sound

THE AUSTRALIAN NATIONAL UNIVERSITY

Editors

Alistair Riddell

Alex Thorogood

Organizing Committee

Alistair Riddell

Alex Thorogood

Annette Marie

Mark Webber

Graphic Design

Annette Marie

Video Documentation

Winnie Yang

Technical Assistance

John Mills

Simon Price

Keynote Speakers

Warren Burt

Brad Garton

Paper Jury

Roger Alsop

Stephen Barrass

Ross Bencina

Andrew Brown

Warren Burt

Christian Haines

Ian Kamiskyj

Tim Kreger

Peter McIlwain

Gordon Monro

Greg Schiemer

Danielle Wilde

David Worrall

Thanks to the many volunteers that have fulfilled tasks in preparation for this conference.

Special thanks to NCH Swift Sound for their contribution to the 2007 conference.

Published by the Australasian Computer Music Association

P.O. Box 284
Fitzroy
VIC 3056
Australia
June 2007

ISSN 1448-7780

The paper refereeing process is conducted according to the specifications of the Australian Government for the collection of higher Education research data, and fully-refereed papers therefore meet the Australian Government requirements for fully-refereed research papers.

* Denote papers selected by peer review of full paper.

% Denotes papers selected by peer review of abstract.

Keynote Speakers

Warren Burt 4
Stories from the Cross-disciplinary Trenches

Brad Garton 11
I Can Be There

Papers

* Roger Alsop 12
The Biggest Shed in the World: an approach to computer enhanced creativity

* Lydia Ayers / Andrew Horner 18
Didgeridoo Synthesis Using Timbre Morphing

% Ross Bencina 26
AudioMulch 2006: a project review

* Warren Burt 27
Cellular Automata as Spectra: Beyond Sonification into Comparison

* Andrew Brown / Rene Wooller / Kate Thomas 34
The Morph Table: A collaborative interface for musical interaction

* Steven Campbell 40
Sensience: Electroacoustic Collaboration in a Reactive, Non-Tactile Ultrasonic Sensing Environment

% Poppi Narelle-Faye Doser 47
Synaesthesia and the Affect of Programming in SuperCollider

% Josh Dubrau / Mark Havryliv 49
P[a]ra[pra]xis: Poetry in Motion

* Toby Gifford / Andrew Brown 50
Polyphonic Listening: Real-time accompaniment of polyphonic audio

* Luke Harrald 59
Collaborative Music Making with Live Algorithms

* Mark Havryliv / Fazel Naghdy / Greg Schiemer 66
Synthesising Touch: Haptic-rendered Practice Carillon

* Alistair Riddell 74
Towards Interactive Control of Sound Patterns Using the Wiimote

* Greg Schiemer / Manuel Op de Coul 81
Pocket Gamelan: tuning microtonal applications in Pd using Scala

* Alex Thorogood 87
Chatter and Listening: A wireless multi-node sound sculpture

% Sebastian Tomczak 94
Handheld Console Comparisons: Lateral Consumer Machines as Musical Instruments

* Alexandra Uitdenbogerd / Ian Kaminskyj 95
The Mood Juke Box project: classifying and retrieving music by its perceived emotional content

* Daniel Vogrig / Stephen Patterson / Ian Kaminskyj 100
Guitar Playing Robot

* Danielle Wilde – *hipDisk* 107
an interactive sonic system inspired by core-body gesture

* Sonia Wilkie / Catherine Stevens 115
Psychoacoustic variables for temporal manipulation and spatialisation of the sound–induced illusory flash

Artist Talks

Ros Bandt 121
Artist talk

David Kirkpatrick / Paul Kopetko 122
Transmission

Tim Kreger 123
Life after the Cliff

Andrew Sorensen 124
Live Coding: New Perspectives on generative art

David Worrall 125
Sonification, Synthesis and SoniPy

Studio Reports

Stephen Barrass 126
Sonic Communications Research Group, University of Canberra

Greg Schiemer 128
The Haptic Carillon Project

Warren Burt 129
Illawara Institute of TAFE. Wollongong

Ros Bandt 130
Australian Sound Design Project Interactive Map and Website CD launch

Posters

Michael Bylstra 131
Cloudscape Radio: The Ambient Sonification of Clouds

Thorin Kerr 133
Computer Assisted Composition System for a Live Coding Environment

Keynote Speaker

Warren Burt
Illawarra Institute of TAFE, Wollongong Campus
PO Box U-27
Wollongong, NSW 2500
Australia
waburt@melbourne.dialix.com.au

Stories from the Cross-disciplinary Trenches

Abstract

Far from being a recent invention, cross-disciplinary thinking in the arts goes back at least to ancient Greece. The more recent history of cross-disciplinary thinking in music is referred to, and the author's own history of cross-disciplinary work is considered. The point is made that music and sound works should be co-equal partners in any collaborative relationship, and the necessity for new venues for this work is discussed.

Introduction

When Alistair first asked me to make this keynote address, and talk about "Trans" - that is trans-media, cross-media, interdisciplinary work, I was a bit bewildered. What did this idea of "Trans" have to do with me? Then it almost immediately dawned on me - I had been living in a cross-media, cross-genre, interdisciplinary world for so long that it seemed completely normal to me. Of course, as a composer, my colleagues would include chaos mathematicians, ecologists, experimental poets, video artists, computer programmers, post-modern dancers, performance artists, biologists, etcetera. Doesn't everyone? Oh, that's right! They don't. Once I'd recovered from my Homer Simpson "Doh!" moment, I began to think - computer.....music......electronic.....music. By their very names this discipline has cross media built right into it. Computer or electronics - I, for one, don't worship the digital as anything special - when you come right down to it, all those 1s and 0s are just low and high analog levels switched vewy vewy fast. In any case, "computer" or "electronic" both have a whole aura of science and technology about them. Music seems much more humanistic, wholistic, "soft" (even heavy metal!). On first glance, or first sloppy thought, we've got C. P. Snow's "two cultures" encapsulated right there, in the name of our discipline.

However, on second glance, this seems not to be the case. Music and technology have always been intertwined. In ancient Greece, music was taught as one of the quadrivium - the four essential disciplines - the others were astronomy, geometry and arithmetic. You learned your fractions, ratios and proportions by the study of music. You learned the balance of the solar system through studying music. (And don't be too hard on my main man Claudius Ptolemy if he got the structure of the solar system wrong - he got the maths right - so right that only in the early 20th century were the accuracy of his observations superceded by modern calculation methods. And if you want to find out what musical tunings were heard in the market place in Alexandria in 140 AD, or read cosmic poetry about the wonder of the universe, well then, Claude's your man.)

Some recent history

If we leapfrog ahead about 18 centuries, even before electronics became the mainstream technology, music was already becoming cross disciplinary. Hermann von Helmholtz' "On the Sensations of Tone," great classic of 19th century science that it is, stands as a model of cross-disciplinary thinking. His work inspired those two great cross-disciplinary musical thinkers of the early 20th century, Harry Partch and Edgard Varese - who both looked to psychoacoustics and science for information on tuning, and how sound works. And of course, in the mid-19th century, Wagner's idea of a total music theatre was already cross-disciplinarity personified, incorporating not just music, but text, lighting, stage design, acoustics, etc. And then of course, 19th century narrative theatre merges almost seamlessly with the emerging technology of motion photography to make the dominant cross-disciplinary narrative artform of the 20th century: cinema. Meanwhile, people like Kurt Schwitters with his Ursonate were making works that crossed the boundary between music and language. Composers like Arnold Schoenberg, Charles Ives, and Percy Grainger studied new developments in acoustics and sound technology very closely. By the 3rd decade of the 20th century, the connection between music and science, at least among advanced musical thinkers, was already firmly established. By the time the young John Cage was a student, in the 1930s, studying harmony with Arnold Schoenberg by day, and assisting Oskar Fischinger with his experimental animations by night, it was already possible to not only envision what kind of artistic future might be possible with all this new technology, it was also possible to be critically evaluative about what kind of future this might be, and what these new tools might mean.

A personal history

My own involvement with multi-disciplinary thinking began almost as soon as I started my undergraduate degree at the State University of New York at Albany in 1967. There was a course called something like "The Arts: 1600-1950," which all young composers were encouraged to take - it was taught by a composer, a sculptor and a writer. In it works of art, music and literature from each 50 year period were compared - structural commonalities were pointed out - for example, the rise of tonality, perspective, and narrative novels all about the same time, and conversely, the breakdown of all of those elements around the start of the 20th century. So very early on, we learned that connections between disciplines were not just fortuitous, they were there, and were important. As well, my undergraduate degree was what called at the time a "liberal arts" education. As well as a full load of music subjects, I took courses in mathematics, literature, history, politics, biology, comparative religion, geography, and studio arts. We were actively encouraged by our teachers to look at connections between the sciences, the arts, and the humanities. As well, this was the hippie era, and many of the institutions of that era, such as Stewart Brand's "Whole Earth Catalogue," as clear a predecessor to the world wide web as can be seen anywhere, actively promoted wholistic modes of thought.

On finishing my BA, I moved to California, and began studying at the University of California, San Diego, which was another place that was embued with the cross-disciplinary spirit. My composition teachers were two composers who were cross disciplinary themselves: Kenneth Gaburo, who not only worked with instrumental, vocal, and electronic music, but also with dance, video, extended vocal techniques, linguistics, and performance art; and Robert Erickson, who built instruments and conducted psychoacoustic research as well as composing music. At UCSD, while I was there, they even set up an institution called the Centre for Music Experiment, which had among its workers not only Erickson and Gaburo, but also Pauline Oliveros, whose work between music, meditation, karate and contemporary physics she described as "The Study of Attention," Roger Reynolds work with interactive electronics, and Jean-Charles Francois, John Silber and Keith Humble who explored improvisation, instrument building, and cross-media collaboration in the group KIVA, among many others. Also in other departments were a host of interesting artists and scientists, such as performance poet David Antin, Duchamp scholar Moira Roth, computer artist Harold Cohen, computer scientist Don Norman, philosophers Herbert Marcuse and Angela Davis, brain-researcher Manfred Clynes, and a host of people from such places as the Salk and Scripps Institutes. As well, my friends outside of the UCSD orbit were similarly oriented towards explorations among and between the arts and sciences. My two best friends were David Dunn, whose work, even at this early period, was drawing connections between ecology, advanced art, and music (he was already writing about "music's insufficiency as a self-contained discipline"); and Ronald Al Robboy, who combined Yiddish scholarship and music performance research with an almost 'pataphysical sense of connection between seemingly unrelated phenomena. Not coincidentally, both of them were assistants to Harry Partch, who was still alive and living in San Diego. It was in this heady intellectual climate that composers such as myself and Ron Nagorcka lived and thrived. And although I would like to wave the old school t-shirt in praise of my alma mater, it should be mentioned that of course UCSD was not the only cross-disciplinary music institution at this time. Other places that were similarly oriented included Stanford University, Mills College, and California Institute of the Arts in the US, York University in the UK; York University in Toronto; and eventually, IRCAM in Paris, which was consciously set up on the model of CME and Stanford, but with official government support, and no academic affiliation. (I was there at CME when Boulez, Globokar and Risset came to pick Roger, Pauline, Bob and Kenneth's brains.)

On moving to Melbourne in 1975, to help with the setup of the Music Department at La Trobe University, it seemed completely natural to both me and Keith Humble that the electronic music studio would also have video synthesis capabilities, and that the course would be designed to encourage cross-disciplinary collaborations. And in working with Ron Nagorcka, we quickly realized another "trans" - a social one - music had to leave its academic nest, and live in the community - Clifton Hill Community Music Centre was set up precisely to provide a home for experimental work that was outside the academy, and which encouraged artists to not only control their own means of production, but also their own means of artistic performance and dissemination as well. Interdisciplinary work was not just encouraged, it was regarded as the norm.

A real impetus for my own cross-disciplinary work came after I was given the heave-ho from academia in 1981. Suddenly thrown out into the so-called "real-world," being denied both the financial and technological support the academy had provided, I had to find ways of both making a living, and making my art. Fortunately, I was able to do so, and many of my projects that people now look on as models of cross-disciplinary work were actually a product of simple economic necessity. My work with the CSIRO in 1985-86, building microtonal musical instruments was one such example. I had returned from overseas in late 1984, and some friends told me about a new Australia Council program called "Artists and New Technologies." I applied for the program with the idea of doing computer graphics research at the CSIRO in Sydney. To my delight, I got the grant (no need to worry about making a living for *those* six months), and went off to Sydney to meet the CSIRO staff I would be working with, only to find that the computer graphics person I would have been working with had died, suddenly and unpredictably, a few days before my arrival. On being informed that they wouldn't be able to replace this person for some months, I asked what other CSIRO facilities were available, preferably in Melbourne. I was informed that the National Measurement Lab was at Monash University, and they had machine shop facilities. I quickly changed my project to one of instrument building and acoustic research, contacted the Melbourne lab, and was informed that the project could go ahead. That was the origin of my tuning forks - a product of economic necessity and quick thinking on my feet when the circumstances of the grant changed radically.

Some of the interesting cross-disciplinary projects I was involved in were the Serge Synthesizer project in California between 1973 and 1984; building my own electronics and small computer systems at both UCSD and LaTrobe in the late 70s and early 80s; working with Simon Veitch, and Perceptive Systems on the 3DIS system on several large scale projects in the late 1980s and early 1990s, several of which involved additional collaborations with dancers; working on a large scale video synthesis and sound project at the Los Angeles based art-science think tank International Synergy in the mid 1980s; collaborating with poet Chris Mann, and post-modern dancer Eva Karczag on a series of performances over a 30 year period from the late 70s to the present; collaborating with the actors, dancers, and performance artists at the Theatre of the Ordinary in Melbourne from 1992-2002, most notably with choreographer Al Wunder, and actor/director John Britton; and working with mathematician Henry Hunter on a series of pieces involving the application of chaos mathematics to music from the mid-80s until Hunter's death in 1992, and then continuing on that work and making it available as software resources for composers in collaboration with software designer John Dunn, a project which continues to the present day. A number of other projects could also be mentioned, but mentioning these should suffice to give the idea of the kind of projects I was involved in.

Lateral financial thinking

Some of the ways I found to fund these projects were quite bizarre - even the act of finding support for the work involved both cross-disciplinary and lateral thinking. As an example, consider my working at the Advanced Computer Graphics Centre at RMIT, Melbourne in 1994.

In 1993, I was getting frustrated because I didn't have access to video synthesis equipment. I knew that computer graphics was making great strides, but not being institutionalized, I didn't have access to the expensive equipment then used. I heard that RMIT had a place called the Advanced Computer Graphics Centre, with a room full of Silicon Graphics machines running Soft-Image, which was at the time, one of the state-of-the-art computer animation systems. I went to see the head of the ACGC to find out about getting access to the equipment as some kind of artist-in-residence. He told me that they would like to have me, but they couldn't apply for funding for me, nor could they sponsor me. If I wanted to work there, I would have to get an external source of funding in RMIT. I went to see Robert Owen, in the Visual Arts department, who taught sculpture. I told him about the ACGC, and offered to swap him a series of lectures on sound sculpture for his students in exchange for them sponsoring me as Artist-in-Residence, so they could second me to ACGC, so I could get access to the SGI machines. He thought t his was a good idea, if I could do all the application work myself. So I applied to the Music Board of the Australia Council to be Artist-in-Residence with Visual Arts / Sculpture at RMIT, in order to work with computer graphics. Amazingly, we got the grant - $5000. This would enable me to live for 5 months while I worked at ACGC. During 1994, I made three ninety second animations - which provided me with about 600 useable still images, which became the basis for the visual part of my installation "Dense Room", which played in Auckland, Louisiana, and Melbourne. The images were also recycled into "moving costumes" for the dancers in my 2000 opera "Lost and Abducted." That is, the dancers were dressed in all white, dancing in near darkness. The images were projected at an oblique angle onto the floor of the dance space. When the audience saw part of an image, it was because a dancer had moved between the projector and the floor. Fragments of abstract computer imagery were shaped by moving dancers' bodies.

I want to go over the funding of this again:

1) I was paid by the Music Board

2) To be in residence with a Sculpture department

3) In order to work with Computer Graphics

4) While living on a salary that was well below the poverty line.

This, I feel, encapsulates quite neatly the nature of trans-disciplinary work in Australia. Not only does the artist work between disciplines, they have to be clever enough to figure out how to manage funding sources between the disciplines!

A few stories from the world of cross-disciplinary arts as I experienced them might be revelatory of both the advantages of this way of working, and of its problems.

Catherine's story

An example of what happens to a trans-disciplinary artist in terms of the economic structure of society is shown by the career of my wife, Catherine Schieve. After getting a PhD in experimental music from the University of California, San Diego, she held a two year post-doctoral fellowship at the University of Melbourne, where she developed her interest in large scale graphic scores (33 meters long, for example). On returning to the US after that, she worked for a while as a book-binder in the library at the University of Texas, before getting a job in the Theatre Department at the University of Iowa. While working in the Theatre Department, she also got an MFA in Visual Art, specializing in video and multimedia art from the Art Dept at Iowa. She then moved to Bard College, where she taught in the experimental music course Music Program Zero, and also taught writing in the Bard Institute for Writing and Thinking. On leaving Bard, she became head of the digital arts program in the Visual Arts Department at South Eastern Lousiana University, specializing in digital printmaking. After that, she became the coordinator of media arts at a private high school, Escuola Graduada in Sao Paolo, Brazil, and then became digital arts coordinator at Vail Mountain School, in Vail, Colorado, while also teaching ethnomusicology at Colorado Mountain College, before moving to Australia in 2002. In one career, she's taught music composition, improvisation and history, theatre, visual art, writing, and technology. As she says, this cross-disciplinary approach has been necessary as much from economic necessity as it has been for any idea of a truly cross-disciplinary practice. Interdisciplinary artists, especially women, are often the last hired, and the first fired, and when the 3 years of the contract are up, or your department head gets the axe, it's on to the next institution or discipline. Notice that this is a Western Hemisphere story, mostly. In Australia, I suggest, her story would have also involved frequent stretches on the dole, and long stretches working for a variety of non-educational institutions, as academia here seems to be more and more calcifying into departments which teach traditional disciplines in a more and more structured manner.

Kenneth's scatter

Kenneth Gaburo's "scatter" technique is worth mentioning as an example of cross-disciplinary thinking, as much for its challenge to "rational" methods of making work as for the works of art produced by it. This process frequently involved him placing himself into a state of sensory deprivation, and then making some kind of physical gesture which would leave a trace. For example, for his orchestra, children, and electronics piece "Antiphony IX", he sat in a totally darkened room, in front of a drawing table, on which were taped multiple pads of graph paper. He began placing dots on the table, in total darkness, and did this until he felt every point on a particular page was visited. He then removed the page, and kept placing dots on the pages. He did this for several hours, until he felt he had reached a state of total exhaustion. At the end of that time he collected the sheets and placed them on the walls of his studio. For about a year, he lived with those drawings, occasionally circling particularly interesting constellations in coloured pencil. At the end of the year, with the drawings filled with interesting shapes circled by multiple coloured pencils, he drew a vertical axis for pitch on each drawing, and a horizontal axis for time, and these pages became the score for the orchestra. It's important to note that he did NOT transcribe the parts into some music notation program, but gave the graphic notation to the orchestra - learning to read the graphic notation was an essential part of the process. And for those who say that this was an impractical gesture, it might be noted that this piece had two performances in his lifetime, and has since been performed at least twice since his death in 1993.

In "scatter," then, a physical process leaves a trace. This trace is then analyzed, and results in a score for other people to perform. The interesting thing for those interested in algorithmic composition is that Gaburo was here using his body as a "random" information generator. By going through the sensory deprivation process, he tried to remove "surface" habits and "licks" of his, in order to reveal deeper underlying patterns. He frequently found surprising things - he made an electronic music piece "Re-Run" using a similar process with a Buchla synthesizer, which was not connected to the sound system - it was a silent instrument. At the edge of exhaustion, and the threshold of consciousness, he performed four tracks, not listening to anything during the process, but simply allowing his sense of physical gesture to dominate. After recovering from this work session, he listened to the tape, and found it had some of the most interesting counterpoint that he'd ever heard in any of his works. Having been involved in contrapuntal thinking all his life, he found it was now indeed firmly embedded in his bones.

And the idea of art as a tracing left by a process seems to me to be at the heart of one kind of transdisciplinary thinking. If a process can be applied to any kind of art (or other) material, then an artist such as Gaburo, or Schieve, or myself, can easily make verbal, sonic, visual, movement, or theatrical works. The viewpoint that art is about exploration of the results of the process, rather than primarily being the expression of personal emotions, can lead the artist into many different areas of science, art, sociology, etc.

Bob's stinging wisecrack – wisdom!

In computer music, we sometimes feel we are halfway between art and science. However, I remember an incident during one of Robert Erickson's psychoacoustics seminars that is indeed cautionary. Sometime around 1973, Erickson had arranged for Rainer Plomp, the reknowned Dutch psychoacoustician, to visit his seminar. In preparation for this, we all critically read Plomp's research, and during his presentation, we grilled him about this research, and its applicability to our experimental sound work. I remember my sense of cosmic disappointment with Plomp: he was being so careful, in the scientific sense of claiming nothing but what his result could empirically show, that for us as composers, his work became interesting but not useful. That is, he told us how people listened to older or pre-existing music, but his work did not lead to the "not yet existent." Perhaps that was a good revelation - science can show us things about sound, but it often can't provide guidance for us - it can't show us how to make choices - in fact, it might be said that unless, as composers, we're ahead of the development of new scientific ideas - that is - *they* have to study *our* work as much as *we* study *theirs*, then we're not doing our job, but are annexing ourselves to another church. Just as medieval musicians subordinated their work to the demands of the church, and commercial musicians of the 20th century subordinated their work to the demands of the market - so we have to beware of subordinating our work to the demands of science - OR it's evaluatory mechanisms. In fact, I remember a meeting with Erickson in the late 1980s. I showed him my work with making scores for my tuning forks based on transcriptions of Mandelbrot Matrices. His scathing put-down resonates with me still. He said "Are you still making that "Scientific American" music? I thought you would have outgrown that stuff by now." Wise words - are we, in computer music, simply composing demo pieces for the latest psychoacoustic theory? Do we really think the way to artistic salvation (and job security) lies in looking more and more like scientists, or in couching our work in terms inimical to its very nature? If we do, I would maintain, we are failing in our jobs.

A literary style?

One area of great failing for me in the field of computer music is the peer reviewed paper, the journal article, the conference paper. The paper is a great way of disseminating practical knowledge, but as a way of living with words (that is, what is known in the most profound sense as "writing") it stinks. I can't think of one ACMA paper in all the proceedings of the past decade or more, including my own, that I can read with pleasure. Many I have found very useful, but inspiring enjoyable uses of language, they were not. If we are creative people in sound, why can't we also be creative people in language? Why can't our means of communicating with each other be imbued with as much sense of fantasy and exploration as our music? Because we're afraid of losing DEST points? Because we feel we have to conform to the norms of scientific discourse? I hope my esteemed colleagues will forgive me if I say that if those are the reasons for our choice of modes of discourse, then we aren't the creative revolutionaries or explorers we fancy ourselves to be - we're wimps, and not particularly gracious wimps, either. When I see conference papers that are as fantastic in their imagery as the pieces they purport to describe, I will then rejoice.

The need for new contexts

One of the great areas of change that I think we, need to consider as cross-disciplinary thinkers and makers, is that of context. The question of where we place our work is, I think, paramount. This became very obvious to both me and Ron Nagorcka as soon as we arrived in Melbourne in 1975. If we would have had to rely on say, the world of classical music performance, or the world of pub rock for the basis for our explorations, our work probably would have died at the start. We started our own venues because we found that the existing worlds were not equipped to provide open minds, open contexts, and what has come to be known as an open source approach to creativity.

In fact, today I would go even further. I would maintain that if you think that making a different kind of music in the same places you made music before constitutes a revolution, then you don't understand the nature of revolution. This seems to be a mistake that everyone from Schoenberg to the Sex Pistols made. They tried to make revolutionary music, but in the same venues that older music had been made. Schoenberg's "Society for Private Musical Performances" was a start in the direction of trying to remake a new social space for a new music. The alternative spaces of the 70s-90s, such as Melbourne's Clifton Hill or Sydney's Performance Space were also a start. But they didn't go far enough. The free improv scene has continually tried to make new spaces for their work, as have the experimental dancers, and we quickly found that the nature of the space determined the work made in it. For example, the Make It Up Club from 1998 -2002 took place in a smoky, boozy place where people talked during the music. This encouraged a higher-faster-louder aesthetic. On the other block (literally 2 blocks away), the Theatre of the Ordinary had a non-smoking light and airy venue with a good quality sound system and a large dance space. The tradition here was one of sitting quietly, paying exquisite attention to the work. The work made here was much quieter, more subtle, and more oriented to interaction with the audience. It IS hard work, but it seems to me that we really do need to keep searching for homes for our work, and in maintaining those homes, as well as making the work itself.

It's not for nothing that I continually joke with students that the role of the composer in Australia is to invent the instrument, build the instrument, write music for the instrument, train performers to play the instrument, organize the gig, find a venue for the gig, advertise the gig, sell tickets for the gig, perform the gig, record the gig, edit the recording for the CD of the gig, maintain the website about the gig, post the recording of the gig on the website, and then write up the documentation about the gig and disseminate that documentation in both print and electronic media.

New tools are not enough

I have read a lot of writing in which the author enthuses about the current abundance of free and sophisticated digital tools for art-making. While I, too, am wildly enthusiastic about this, I would suggest that the mythology that the availability of radical tools will automatically produce a proliferation of radical art is indeed just that - a myth. It could be that our tools don't transform us *enough.* That is, the radically transformative implications of new computer music tools can be blithely ignored even by its practitioners. A tough aesthetic stance – one that pushes beyond the known - must be developed, cultivated, and sustained - it won't simply be produced by the availability of tools.

The need for musicality

If we, as musicians, have anything to offer all of our colleagues in the other arts, the sciences, the humanities, I would suggest that what we have to offer is our musicality. And I would further suggest that we need to make them aware of this quality, and how this quality can be of benefit to them. Rather than conforming to the norms of another artform, or discipline, I think we need to ask: to what extent does our musicality affect our non-musical work? To what extent does our sense of phrasing, of density, of swing, of structure pervade our writing, visual work, dance work, dramatic work, etc? One can speak of a very musical writing style, or a very musical way of moving. Can we, in intermedia works, bring a different sensitivity to the table - one that might make critics or commentators on those artforms change their terms of reference in writing about the work, acknowledging an influence of musicality on those other artforms? Even more, could we make work that would, in some way, convince word-oriented critics and writers of the absolutely equal importance of those OTHER non-verbal forms of human intelligence - the sonic, the kinesthetic, the tactile, the visual?

A reality check

A bit of a reality check might be in order. Just so we don't think that our revolution (or whatever it is) is won, consider this incident that happened to me just a couple of weeks ago. I mentioned to a high-school music teacher the absolutely non-controversial fact that music is based on numbers - notes being vibrations at certain numbers of cycles per second, and intervals being constant ratios between two different vibrating frequencies. She was not only amazed at this news, she was scandalized, feeling that this kind of information had no place in a high school curriculum, much less in the minds of her students. This, it might be mentioned was a young teacher about 2 years out of music school! All of us need to have experiences like this continually, I think, to keep reminding us that we are, indeed, a tiny minority not only in the society at large, but also in the arts as well. There is, still, a very long way to go, but the work needed to get there is exciting, and filled with possibility. At times, in an era of diminishing funding opportunities, calcifying educational institutions, and incompetent administration, it may seem that cross-disciplinary work was a brave idea that never quite caught on. However, I would maintain that it is indeed not only the “way of the future” which institutions will eventually need to adopt for their own survival, it is, and for over a century and a half, already has been, the absolutely normal way that art has been made - and is the basis for any new understandings that the arts will be able to give us.

Keynote Performer

I Can Be There

Brad Garton
Professor of Music
Director of Computer Music Center
Columbia University
New York City
USA
brad@music.columbia.edu

Abstract

The early notion of the Internet nurturing the growth of trans-geographical "communities" that could replace traditional local human social groups hasn't quite worked in the naive way initially imagined. However, aspects of the Internet have allowed the development of new kinds of communal activity, leading to new forms of shared social bonding. This "keynote event" will describe and demonstrate unanticipated characteristics of these alternative social networks, particularly as they relate to artistic production and creative imagination.

Roger Alsop
VCA, University of Melbourne
234 St Kilda Rd
Southbank, 3006
Australia
ralsop@unimelb.edu.au

The biggest shed in the world: an approach to computer enhanced creativity

Abstract

When I was growing up we had a shed in the back yard, it was the place where any and everything we did not have an immediate use for was stored; there was no method to what was kept or where it was put. This created an environment where anything could be found with anything else, everything was familiar and nothing was sacred or precious: a true cornucopia from which much was made.
This paper looks at a role of the computer as an assistant to creativity through being a resource of information, opinions and interpretations and the conceptual spaces they inhabit, where the digitized object or idea is available to the artistic creative processes at a variety of levels and in a variety of ways.

Introduction

"Artistic expression forms an experiential bridge, crossing boundaries -- between people, communities, nations. The expression reifies the relationship between transmitter and receiver -- the relationship reifies the expression. (29 jan)" (author unidentified, http://www.acmc07.org/)

I made a cursory search for a definition of the verb *reify*, revealing two quite different interpretations, one from the 1974 Concise Oxford Dictionary(Fowler 1974), the other from Encarta World English Dictionary(Bloomsbury Publishing Plc. 1999), both common arbiters of word definitions in their respective times,

According to the Concise Oxford it means to "convert (person, abstract concept) into thing, materialize", and according to Encarta it means "to think of or treat something abstract as if it existed as a real and tangible object". Both dictionaries were and are well used in their different time frames, and both provide an interpretation of the respective world views of their times.

The distinction between these two interpretations resides in the phrases "to *think of* or *treat* something abstract *as if* … real", and to "convert into thing … [to] *materialize*" something; one is *conceptually* creative, the other *actually* creative. This demonstrates a change of meaning over time, the interpretation evolving appropriately to the technological and cultural developments within the societies using it. The Oxford definition indicates a time in which reality was less mutable and virtual realities were not considered as real possibilities, something was real or it wasn't and it was possible to *convert* the unreal into reality, perhaps through a creative act. The Encarta definition allows for a virtual construction, something can as easily be real as it can be *as if* real.

These two definitions, rooted in their times, indicate that definition(s), through which an understanding of the world is achieved, are dependent and mutable, and may be generated and used reactively and creatively when interpreting the world.[1]

The advent of the computer as a common tool was possibly the event that allowed the concept "*as If … real*" to become common, it is only recently that a virtual reality can be considered as a valid reality[2]. It is this tool that has redefined much of the processes used to interpret, influence, understand and be creative. For the musician the computer has had many influences, it provides ways to collate and organize sounds, ways to generate sounds and, most importantly here, ways to reorient practices.

Artistic creativity

The computer-oriented musician/composer as belongs to a subset of the larger set 'artist'. It can be assumed that the 'artist' is essentially a communicator, who sees publicly expressing their ideas as implicit in their role. Leo Tolstoy saw that "Art begins when one person, with the object of joining another or others to himself in one and the same feeling, expresses that feeling by certain external indications." (Camp 1998)[3]. Of course the reasons for making art are as many as there are artists, and it is possible to substitute other words such as, perhaps 'concepts' or 'ideas', for "feeling", making his statement more suitable and accurate in some cases.

An artist creates something that may be historically new, in that the work, and possibly the concepts that allowed it, have not previously existed, and it may also be personally new, in that the artist has created something that they have not created before and based on concepts new to themselves. In the act of creation, there is often a need to redefine approaches and systems assumed as part of the creative act, and this is a significant part of the artist's role.

The Art(ist) as re-definer

When making something new, a *re*defining of what is known or believed to be known is often essential; the case above indicates a *re*defining of a word resulting from broad societal changes.

One of the functions of the artist is to be a re-definer, reinterpreting the world for themselves and subsequently for others through their work. For example: by painting a pipe and stating within the painting that it is not a pipe Magritte confronted the notion of representative painting; by renaming and relocation a urinal Duchamp created the 'readymade', requiring the viewer to see objects differently; and by having a musician play silence in a concert hall Cage caused the world to become musical for 4'33". By reinterpreting and redefining traditions in their respective practices these artists questioned many traditionally held beliefs and understandings of art and creativity. In each of these cases the creative works caused an iconoclasm of some sort, forcing a broader reinterpreting and redefining of viewpoints within the various arts practices and perhaps further.

Within the creative process, the artist also often reinterprets and redefines themselves, their relationship to their work, and to their world. This can occur as the artists discovers new paradigms through self reflection of and through examining their own work. It may also occur as new technologies, and the new conceptual frames that they invoke, become available.

The ubiquity of computer-based technology in technologically developed societies offers the artist residing within those societies opportunities unthinkable a generation ago. Being able to access intellectual, conceptual, physical materials from other societies offers a plethora of opportunities for redefining and reinterpreting, either of those materials or through those materials. It also offers opportunities to generate wholly new conceptual frames and the subsequent works that may either be developed from within or as a result of those frames. This is one of the computers greatest contributions to the creative artist.

The Artist as creative agent

Margaret Boden sees creativity as fitting two possible areas, H(istorical)-creativity, "creating a concept which has never been created before at all [and P(sychological)-creativity], a concept which has never been created before by a specific creator." (Boden 1995; Wiggins 2001). Both of these areas are of interest to the computer-based artist, in that they are working with technology where H-creativity is occurring at a meta level and P-creativity is occurring a more local/personal level. Of course, there is considerable overlap between these two modes, they are often looping to equally influence conceptual and technological frames coincidentally.

In Boden's terms a more creative result is one that transforms a conceptual space, not one that just the explores or exploits a conceptual space, however expert or novel that exploration or exploration may be. According to her "dropping a constraint is a general heuristic for transforming conceptual spaces." (Boden 1995) An example she gives of this is Friedrich August Kekulé redefining his view of the compound benzene in order to better explain it[4] (Boden 1995); perhaps it is finding the right constraint to drop that creative genius resides.

The creativity of people such as Kekulé, Albert Einstein or John Nash lay in their ability to reinterpret and then redefine; having impact by generating a change in conceptual spaces that for the most part did not have immediate tangible outcomes. However, what these changes allowed for was the exploitation and reification of and within the newly created conceptual spaces, providing for and resulting in new understandings and tangible outcomes. Artistic creativity requires a tangible outcome; it requires that something unreal be first be formulated and considered as possibly real by the artist, who then makes it real.

The expression of the creative artist is through being actually creative, the reification, in the 1974 Oxford sense, of the artistic impulse. By bringing abstract things into a state that can be witnessed, the artist shares the impulses that caused the creative act, making their creativity tangible[5].

When embarking on the creative process the artist may have a notion that they think is worth sharing with others, and the reasons for this desire are manifold. On another hand, the artist may be experimenting within their expertise such as Jackson Pollock's development of 'drip' painting into his technique, and through doing so develop or discover something that they believe is worth sharing Or the artist may be endeavoring to broaden their expertise in new or allied fields and in doing so generate works that are novel to themselves and to the audience of that field, composer/choreographer/installation artist Meredith Monk is an example here.

The audience as creative agent

It is also very much the case that an artist's, or otherwise creative person's, work may not be fully appreciated or understood in their times. Vincent van Gogh is an example of an artist whose work was not appreciated in his lifetime, selling one or two paintings before his death. Other artists such as Anton Webern or Raymond Scott, while perhaps slightly more successful in their lifetimes than Van Gogh, still were not recognized for the contributions they made to their respective fields during their lifetimes.

Another example of creativity not being recognized is Heron of Alexandria's use of steam power somewhere between 150 BC and 250 AD. The majority of his work with hydraulics was used in the development of entertainments, creating fountains, raising platforms, magically opening doors, and so on. Perhaps the most potent, but unrecognized, of his designs was the

> "aeolipile … a hollow sphere mounted so that it could turn on a pair of hollow tubes that provided steam to the sphere from a cauldron. The steam escaped from the sphere from one or more bent tubes projecting from its equator, causing the sphere to revolve. The aeolipile is the first known device to transform steam into rotary motion." (O'Connor and Robertson 1999)

While the aeolipile was seen as a toy at the time of its invention it is now easy to see that it is actually a steam-powered engine; made useful by simply attaching an axle to the revolving sphere.

That Heron and his contemporaries did not see this possibility shows, perhaps, a lack of understanding of the conceptual frameworks required to make this connection.

In 1690 Isaac Newton explained the Laws of Thermodynamics, the third, "Every action produces a reaction equal in force and opposite in direction", having particular relevance to Heron's aeolipile, perhaps providing the required conceptual frames for its exploitation. It was shortly after this understanding of nature's behavior that Thomas Savery was able to exploit Heron's toy in creating the steam engine(O'Connor and Robertson 1999; Lahanas 2004; Forrester 2005).

These are examples of the contemporary viewer not valuing a creative work simply because they do not have the tools through which to access and assess the work. Unfortunately, and possibly most profoundly, the predominantly market driven society all audiences exist in defines the immediate value of any creative work, artistic or not, and to a large extent, how the work should be perceived.

It is also possible that the creative act can reside within the viewer, making them as much a part of the creative process as an author. This collaboration in the creative process can be expressed through a variety of interpretations; photography is an example where the often viewer creates the artwork. In this case the viewer captures and interprets on film an existing object at a 'decisive moment' in such a way as to make it an artwork, the works of Henri Cartier-Bresson, who has exceptional skill in capturing and interpreting the 'right', unedited, "readymade", moment, such as *Rue Mouffetard, Paris, 1954*, are an example. (Mraz 2003)This is a case where the viewer created an artwork simply through observing and capturing an unrehearsed or un-devised event.

The more traditional audience plays a more passive but effective role in creating an artwork simply through interpreting the artwork from their own perspective(s). This means that the artist loses their position as soul author and arbiter of their work through the very act of expressing and sharing their creative impulse(s).

An example of this process is my (re)interpretation of Cage's *Imaginary Landscape No. 4.* The piece as performed now would vary significantly the premier performance in New York in 1951. "When one listens to [or performs] the work, it is obvious that one cannot predict what will be heard, <u>which is exactly what Cage was aiming at with this composition</u>. [my underline]"(Unknown 2006) Obviously the content would be vastly different in a contemporary performance from what Cage could possibly have imagined when composing it; but it is also possible that a contemporary interpretation of the work will be significantly different from what he might have assumed or predicted.

My understanding of the work shifted dramatically while driving from Phoenix to San Diego in 1993. On that trip I set the car radio to scan and it took a long time to get from one end of the dial to the other, spending about ten seconds on each station. This led me to reinterpret Cage's work to be about social, temporal and locational change, where a four-minute performance would give a very precise indicator of each of those three elements. In this way I, as audience, captured, redefined and reinterpreted another artists work, personalizing it and claiming it as my own.[6]

Computer creativity

The computer forms a robust fulcrum around which conceptual spaces can spin or teeter. As an artist the computer allows an interrogation and exploitation of very many conceptual spaces, and the myriad offshoots that they can engender, with comparably less effort than was required in pre-computer times. This is potentially one of the most valuable attributes of the computer to the creative person. It offers similar meta-explorations to that undertaken in the field conceptual art, which

> "sought to analyze the ideas underlying the creation and reception of art, rather than to elaborate another stylistic convention in the historical succession of modernist avant-garde movements. Investigations by conceptual artists into networks of signification and structures of knowledge (which enable art to have meaning) … examine the interstice between visual and verbal [or sonic] languages as semiotic systems." (Shanken 2002)

This exploration creates an area where creativity itself is being explored and forms a framework in which researchers, such as Boden and Minsky (Minsky 1982), have made inquires. There is a sense that the pathway to developing effective AI creativity is through gaining a clear understanding of human creativity and modeling it. This approach has broadened over years, the call for papers for the MUSIC-AI 2007 International Workshop on Artificial Intelligence and Music gives a list of methods used of AI approaches to musical creativity including:

> "cognitive modeling, data mining and classification, expert systems, generative systems, grammars, fuzzy logic, genetic algorithms, hidden Markov models, intelligent agents, inductive logic programming, knowledge representation, knowledge-based systems, machine learning, neural networks, constraint satisfaction, and planning".(Unknown 2006)

This list indicates the breadth of techniques used and available to get a creative outcome from a computer. In exploring each of these approaches, an understanding of "networks of signification and structures of knowledge (which enable art to have meaning)" may be gained.

The Computer as re-definer

The computer as re-definer is not really that different from the artist as re-definer. The home movie maker filtering their movies, relocating the images and removing anything unwanted is an example of redefining and reinterpreting the original material made easily available through computer technology.

This kind of rebuilding is prevalent in sound, for example: sounds can have the place in which they were captured redefined as another place, real or not; they can be splintered and rebuilt into completely different sounds that would be unrecognisable when compared to the original. Sounds can be turned into visuals and visuals into sounds; movement can be re-defined as any of the above. In these examples the artist is the generator and cause of the changes.

The computer as creative agent

The computer can assist the artist in the creative process and in doing so generate outcomes that are personally new and potentially historically new in many ways. These include, perhaps most potently but often less recognized, the extraordinary access to and rapid transference of information afforded by computer technology. It is now possible for an artist to quickly develop, or become part of, a tightly focused yet large community, in which ideas are quickly discussed, interrogated and evaluated. Up until the advent of computers, and subsequently the Internet, this kind of intercourse was difficult and time consuming.

Use of computers as an assistant to the creative process, or as a creative machine themselves can be seen in the work of programmers, scientists and engineers such as Steven Thaler, whose "Creativity Engines are neural network systems that display creative behaviour; this effect is achieved by introducing a precise degree of random perturbation into the neural networks"(DeGracia 1999). This is a process that attempts to replicate an aspect of the creative process; the jumbling of data within set constraints and allowing new logics to appear. (Thaler 1996; Hesman 2004)

It is also possible to use the computer as a tool that reconstitutes the work of certain composers into new compositions reminiscent of those composers. (Cope 1999; Cope 2003; Cope 1992), or that generates entirely new compositions via algorithms created by composer.

There are many programs that will generate what may be called creative artworks such as Harold Cohen's painting program AARON and Ray Kurzweil's poetry generating program (Kurzweil 2000; Cohen 2001; Kurzweil 2001); and the music generating programs of very many computer oriented composers[7]. Bruce Buchanan provides a list of creative programs that expands those above to include the areas historical science, chemistry and mathematics. (Buchanan 2001)

One area of creative artistic activity that claims, with some justification, to have integrated computer technologies into the entire creative process is dance, Johannes Birringer claims that

> "dance has taken the lead, among the theatrical arts, in absorbing technology as a creative tool, affording dancers and technologists the opportunity to explore interactive environments, virtual places, and integrated methods that have shifted artistic process." (Birringer 2002)

Here the technology allows for an, occasionally seamless, integration between the modalities of movement and sound requiring the composer create outside of traditional musical paradigms.

Problems with computer creativity

The perpetual nascence of computer technology makes it difficult for the computer based art form to reach a maturity of expression similar to that of other art form, oil painting for example. Not being static long enough to develop a maturity similar to that of more well established practises may simultaneously be the computer's most valuable and frustrating attribute.

There is also, among some computer musicians, an infatuation and focus with the technology, this can be quite appropriate as the technology in some ways bounds and expands the possibilities of the practice. However, there is an infatuation with the next and newest does not allow development of expertise similar to that of an instrumentalist, and that can occasionally mask or subvert the reason for the art making. An extreme metaphor would be architects focussing on building equipment as an integral part of their artistic discourse.

My work

Three recent works, *5*, *20* and *41*, signify developments in my own art making that result from new interpretations of developments in computer technology. The initial notions that caused the works related to the supposed 'impersonalness' of digital technology, particularly when used as a tool in art making. It is easy to see that the digital artist must have a foot in two camps, that of the 'artist' who invents the cause of the work, and of the 'technician', who implements the work. These positions are mutually informing and overlapping, causing a practise built on shifting sands. Personally, I find these shifting sands one of the most intriguing and fascinating aspects of computer based art making, creating a continued sense of displacement and wonder.

The area of artistic expression I have been looking into recently, and wanting to explore in these works, is the portrait. This is an area of art making well explored in visual arts but perhaps less so in the computer based arts. Two other areas of interest are the 'readymade', which is very much explored in the digital arts, evidenced by the prodigious use of sampling in music and freely available images in the visually oriented arts; and finally glitch, deliberately creating situations where the computer behaves in an unpredictable way due to inabilities of the hardware and/or software.

When beginning the exploration I worked with images and sounds of my immediate material and virtual environments. I also looked for digital portraits on the Internet, choosing those available in Flickr (2007) under the creative commons license as being easy to access and having much variety.

What I found most interesting was the vast number of people making images of themselves and others available for use by anyone else[8]. My predisposition regarding photographs I have of friends and family is that they are viewed in highly controlled circumstances to invited viewers; I found the willingness to make what I considered intimate photos, such as wedding or photos of the children strangely disquieting and intriguing, automatically thinking that these images were for a select audience. The other thing that I found as intriguing was the number of posed 'art' portraits available; people were obviously being creative with the medium and broadcasting their creative efforts widely.

Over January 2007 I gathered a number of images of faces from Flickr and began to adjust them with the Max/MSP/Jitter (Zicarelli 2007) program using various processes, finally centering on the magnifying object "jit.rota" and reducing images to areas around one hundred thousands of their size. This resulted in significant, and unpredictable, alterations in the images. Occasionally the images would recognizable but appear granulated or pixilated in various ways. The unpredictability of the outcomes was, I am assuming, due to the interactions of the hardware and software of the computer. In order to test this I ran the same patcher on two very different computers, a 2GHz Intel Core 2 Duo iMac with 2 GB of RAM running OS 10.4.7 and Max 4.6.2 and Jitter 1.6.2, and an 867 MHz G4 PowerPC laptop with 640 MG of RAM running OS 10.4.8 and Max 4.5.7 and Jitter 1.2.4. The program worked on both machines but the outcomes, while quite similar, showed some differences. The most profound differences were when the works were rendered to DVD.

These differences were mainly in the visual frames being presented in performance and when being recorded. Importantly these differences did not significantly affect the ideas that formed the artistic intentions.

Conclusion

Using computers in the creative process can follow many paths, each as idiosyncratic, individual and personal, as there are composers. By using methods such as those discussed in the MUSIC-AI 2007 call the composer can be introduced to new ideas, approaches, musical paradigms and so on. However there is as much to be gained by using broader, less esoteric, computer technology to inspire these things within the individual creator.

References

(2007). Fickr, Yahoo Inc.

Birringer, J. (2002). "Dance and Media Technologies." PAJ 70: 84-93.

Bloomsbury Publishing Plc. (1999). Encarta World English Dictionary, Microsoft Corporation. 2005.

Boden, M. (1995). creativity and unpredictability. Stanford Electronic Humanities Review. 4.

Boden, M. A. (1995). "Modelling creativity: reply to reviewers." Artificial Intelligence, Volume 79, 79(Issue 1, November): 161-182.

Buchanan, B. G. (2001). "Creativity at the Metalevel:AAAI-2000 Presidential Address." AI Magazine Fall.

Camp, J. V. (1998). "What Is Art?" (excerpts) by Leo Tolstoy. 2004.

Cohen, H. (2001). AARON: A Product of Kurzweil CyberArt Technologies.

Cope, D. (1999). "Facing the Music: Perspectives on Machine-Composed Music." Leonardo Music Journal Vol. 9, (Power and Responsibility: Politics, Identity and Technology in Music): 79-87.

Cope, D. (2003). "Computer Analysis of Musical Allusions." Computer Music Journal 27(1): 11-28.

Cope, D. (1992). "Computer Modeling of Musical Intelligence in EMI." Computer Music Journal 16(2): 69-83.

DeGracia, D. J. (1999). In the Theater of Dreams: Global Workspace Theory, Dreaming, and Consciousness.

Forrester, R. (2005). The Discovery of Steam Power, Rochelle Forrester. 2005.

Fowler, H. W., F.G (1974). The Concise Oxford Dictionary. Oxford, Clarendon Press.

Hesman, T. (2004). Stephen Thaler's Computer Creativity Machine Simulates the Human Brain, St. Louis Post-Dispatch. 2005.

Kurzweil, R. (2000). Ray Kurzweil's Cybernetic Poet: HOW IT WORKS. 2005.

Kurzweil, R. (2001). AARON: A Product of Kurzweil CyberArt Technologies, Kurzweil CyberArt Technologies, Inc. 2005.

Lahanas, M. (2004). Heron of Alexandria, Michael Lahanas. 2005.

Minsky, M. (1982). WHY PEOPLE THINK COMPUTERS CAN'T, First published in AI Magazine, vol. 3 no. 4, Fall 1982.

Mraz, J. (2003). What's documentary about photography? From directed to digital photojournalism.

O'Connor, J. J. and E. F. Robertson (1999). Heron of Alexandria, School of Mathematics and Statistics University of St Andrews, Scotland. 2005.

Shanken, E. A. (2002). "Art in the Information Age: Technology and Conceptual Art." LEONARDO 35(4): 433-438.

Thaler, S. L. (1996). Neural Networks That Autonomously Create and Discover (US Patent 5,659,666).
Unknown (2006). Imaginary Landscape No.4 (March No.2).
Unknown (2006). MUSIC-AI 2007 International Workshop on Artificial Intelligence and Music.
Wiggins, G. A. (2001). Towards a more precise characterisation of creativity in AI.
Zicarelli, D. (2007). Max/MSP/Jitter. San Francisco, cycling74.

Endnotes

[1] In fact, the word does not exist in the 2003 Oxford Australian Integrated School Dictionary, used by current secondary school students.

[2] People now spend real money for virtual clothes in *as if* realities such as 'Second Life'

[3] Tolstoy had many other comments on "Art" that are succinctly listed by Camp.

[4] Boden recounts that Kekulé came across his method of understanding in a half waking dream, where apparently unrelated images caused the required insight.

[5] By tangible I mean something that can be witnessed, and include works that have an ephemeral existence, such as a dance or musical performance or an installation.

[6] Without negating Cage's authorship.

[7] Obviously, this list would be enormous were I to include all of the computer-based artists who use the tool in their creative works and explorations.

[8] It must be noted here that the open creative commons license is the default when posting images on Flickr.

Lydia Ayers
Andrew Horner
The Hong Kong University of Science and Technology
Clear Water Bay, Kowloon
Hong Kong
layers@cs.ust.hk, horner@cs.ust.hk

Didgeridoo Synthesis Using Timbre Morphing

Abstract

This research describes an additive synthesis design for the ***didgeridoo*** *which captures many of its subtle timbral and expressive characteristics. The design uses timbre morphing to model the dynamic spectra of the instrument in Csound. The design is very flexible, and allows vocal sounds, harmonic sweeps and flutter tonguing. Finally, a composed musical excerpt illustrates the expressive features of the design.*

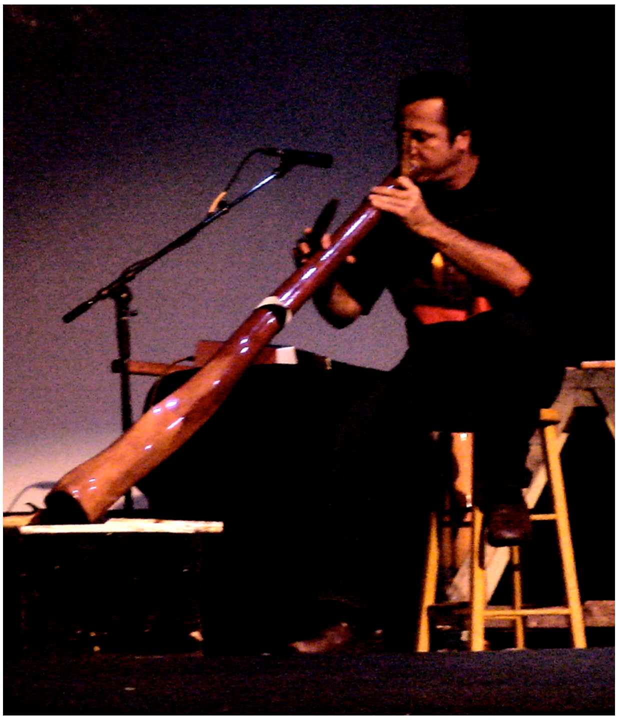


Figure 1. Ash Dargan Playing the Didgeridoo

Introduction

We modelled the *didgeridoo* (see Figure 1) using additive synthesis in Csound (Vercoe 1992), which captures many of its subtle timbral and expressive characteristics. This paper first gives some background for the *didgeridoo* and then describes the source material we used in this research. Following a brief summary of some additive synthesis customizations we have used to model other instruments, we discuss the features in a new design using timbre morphing to model the dynamic spectra of the *didgeridoo*. We then examine musical expression on the *didgeridoo*, including vocal sounds, harmonic sweeps and flutter tonguing. Finally, we conclude with an excerpt from a composition illustrating the expressiveness of the design.

Background

Previous work on modelling instruments has primarily focused on simulating western instruments, especially wind instruments (Risset 1969, Risset and Matthews 1969, Morrill 1977, Horner and Beauchamp 1996, Horner and Ayers 2002). Rodet and Lefèvre (1997) connected the frequencies of two notes with a line segment and morphed the transition using parameter interpolation to give a smooth slur. We previously designed wavetable synthesis models (*i.e.*, group additive synthesis models) of many wind instrument tones that are easy to use and sound like the original individual tones (Horner, Ayers and Law 1999, Horner and Ayers 1998a) and synthesized timbre tremolos and flutter tonguing on wind instruments (Ayers 2004).

Researchers have also studied Asian instruments, including the Chinese *dizi* (Tsai 2003, 2004), the Asian free reed instruments (Cottingham 2004) and large Chinese bells (Braun 2003). We modelled the Chinese wind instruments using a wavetable model similar to the one we used for western wind instruments, and slurred varying numbers of notes in trills for the Chinese *dizi* and Chinese folk wind instruments (Ayers 2003, Ayers and Horner 2004a and b). We have synthesized the Chinese flutes and the *suona* using additive synthesis in Csound (Ayers 2005, Ayers and Horner, 2006).

Researchers have studied the acoustical properties of the didgeridoo (Fletcher 1983, Wiggins 1988) and morphed the sound of a double bass and a *didgeridoo* (Kuchera-Morin 2002).

The *Didgeridoo*

The *didgeridoo* is an Australian buzzed lip instrument traditionally made from the trunk of a stringy bark or eucalyptus tree that has been hollowed out by termites (Hudson 1994). Every tree is uniquely shaped, and the meandering termite pathways in the walls of the instrument help determine the character of its sound. Some instruments have a large bell at the end of the tube where the tree root flares out which adds a rich color to the sound. The instrument is played with the lips buzzing into one open end of the tube (see Figure 2), rather than with the lips in a small mouthpiece, as with the western brass instruments. The lips are much looser than for the brass instruments, making the sound quite rich and deep. In addition, various mouth sounds affect the timbre. The *didgeridoo* plays a drone tone colored with vocal sounds and harmonics. The *didgeridoo* comes in several sizes which produce different fundamentals for the drones. This paper examines synthesis of representative sounds produced on a *didgeridoo* with a tone pitched at B1 (62.5 Hz).

Figure 2. Didgeridoo Mouthpiece.

The Source Material

We worked from CD recordings by Ash Dargan (Larrikia tribe, Darwin, NT), *Ash, Dust & Dirt* (1999) and *Didgeridoo Made Easy: A Beginner's Guide* (2001). (We obtained his permission to analyze timbres from his CDs to create the designs for this research. We have also corresponded with him regarding the results of this research.) Since the *didgeridoo* doesn't typically play scales, we began the research using *didgeridoo* sounds on just one pitch. Because we wanted to use the resulting instrument in a composition, we started with *Wangurra* (sleep) (Dargan 1999), one of the pieces that used a *didgeridoo* in B1, a key that fit the composition. We then chose a collection of interesting sounds to analyze for the instrument model.

Usable Features of the Previous Model

We found that the additive synthesis model used in our previous work (Ayers and Horner 2006) was a good starting point for this project, but it needed some refinement to produce the expressive playing techniques that are characteristic of the *didgeridoo*. The following summarizes the features that we reused in the new model.

The design uses 63 sine waves to model the rich spectra of the *didgeridoo* timbres (see Figure 3). Each code block produces a sine wave at the required harmonic frequency, with a slight random inharmonicity. Each harmonic has a large variation in its spectral average amplitude which makes the tones sound more lifelike and unique. The attack and decay times become longer as the harmonics get higher, which makes the overall tone get brighter as it gets louder and less bright as it decays. Each harmonic's amplitude envelope is multiplied by a slow random noise and a fast random noise (jitter) and the harmonics are added together. The design avoids aliasing by omitting any harmonics above the Nyquist frequency. We used a time-varying FM vibrato on the frequency of the tone paired with an amplitude modulation vibrato.

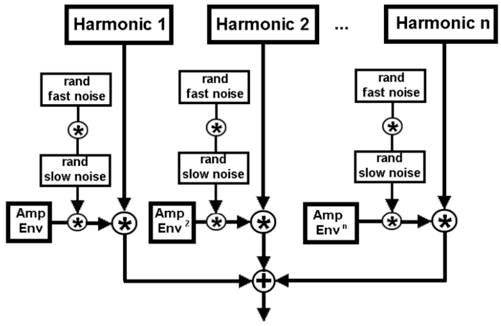


Figure 3. The Additive Synthesis Model.

Synthesizing the *Didgeridoo*

We found the features of the additive synthesis model summarized in the previous section useful in creating a design to model the *didgeridoo*. However, the concept for the new model is different. Since the *didgeridoo* plays basically a single fundamental drone tone and its harmonics, colored by the activities of the voice, we focused on morphing together a string of timbres, rather than slurring a group of pitches.

The following discussion will consider a 2.344 second group of three sounds using the voice to produce "ah-ee-oo too wah," from the beginning of *Wangurra* (Dargan 1999). A spectrogram (Horne 2001) of the segment shows a harmonic sweep at the beginning, followed by a percussive "t" at the beginning of "too" and then a "wah" (see Figure 4).

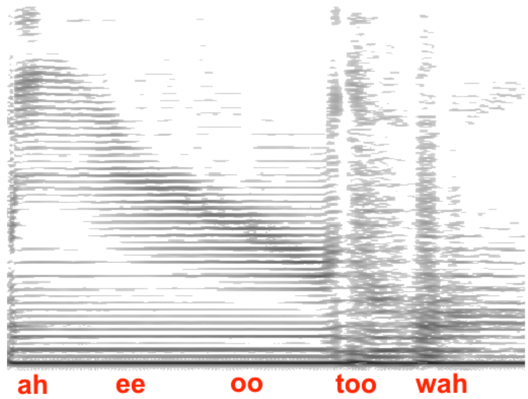


Figure 4. Spectrogram of Original Didgeridoo B1 "ah-ee-oo too wah" Sound Segment.

Figure 5 shows a 63-harmonic time-varying phase vocoder analysis (Dolson 1986, Wun 2000) of the segment. The segment contains a medium amount of instability. The harmonics sweep from the high harmonics at the lower left to the lower harmonics at about time 1.5, when the noisy "t" interrupts the sound, followed by the bright spectrum of the "w" in the "wah."

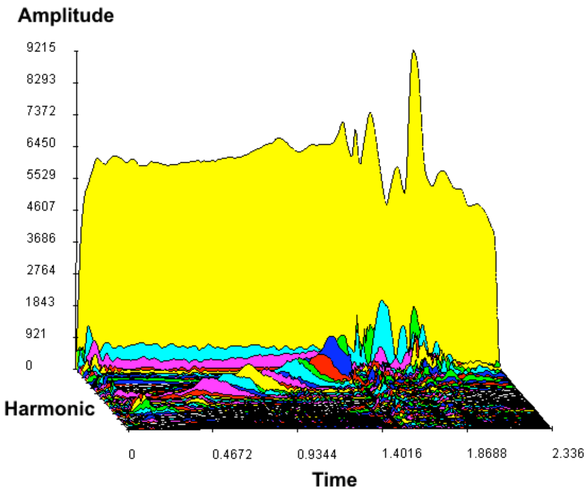


Figure 5. Time-Varying Phase Vocoder Analysis of Original Didgeridoo B1 "ah-ee-oo too wah" Sound Segment.

Figure 6 shows the average spectrum, which has a strong fundamental and a small formant centered on the third harmonic. There is a smaller formant around the 15th harmonic. We found, however, that because the infinite variety of vocal sounds causes the formant to vary considerably, we cannot use the average spectrum to model the *didgeridoo* as effectively as we modelled other wind instruments in our previous additive synthesis designs.

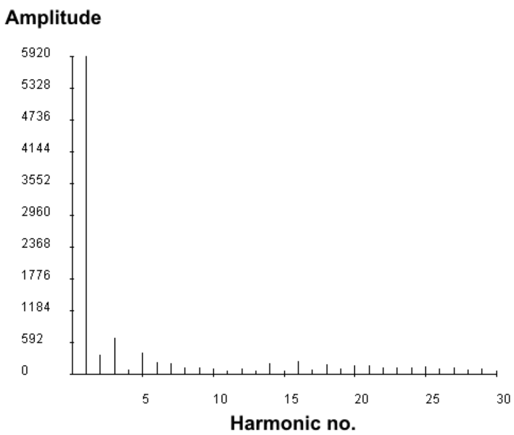


Figure 6. Average Spectrum of Original Didgeridoo B1 "ah-ee-oo too wah" Sound Segment.

For our instrument design, we divided this sound segment into three sounds, a harmonic sweep (the changing vowel sound) and the syllables "too" and "wah." We describe the two syllables in the following sections, as they provide good illustrations for how we modelled the *didgeridoo*. We modelled the harmonic sweep as an expressive technique and we describe it in the section on *Didgeridoo Musical Expression*.

Modelling the "Too" Sound

Figure 7 shows the time-varying phase vocoder analysis of the vocal "too" syllable mixed with the *didgeridoo* tone. This sound has two phonemes, the noisy "t" in the attack and the "oo" vowel which is most of the sound. As our modelling technique involves capturing the spectra of the phonemes and then morphing between them, we might like to call this sound a "morpheme," but as morphemes are sounds with "meaning" and these sounds may be without independent meaning, we will refer to these sounds as "syllables."

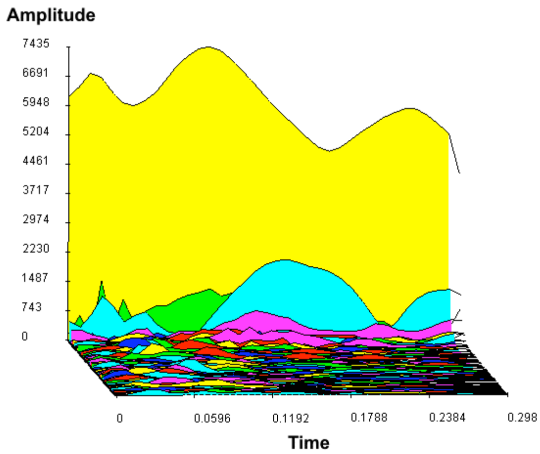


Figure 7. Time-Varying Phase Vocoder Analysis of Original Didgeridoo B1 "Too."

We took two snapshots of this sound (see Figure 8), one for the "t" at time .035 and one for the "oo" at time .267. We chose times from the amplitude peaks in the middle of the section of the waveform that represented the phonemes.

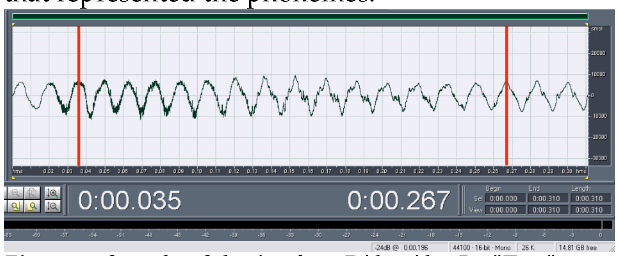


Figure 8. Snapshot Selection from Didgeridoo B1 "Too."

The Spectral Snapshots

After choosing the snapshot times, we examined the phase vocoder analysis graphs (Figures 9-10).

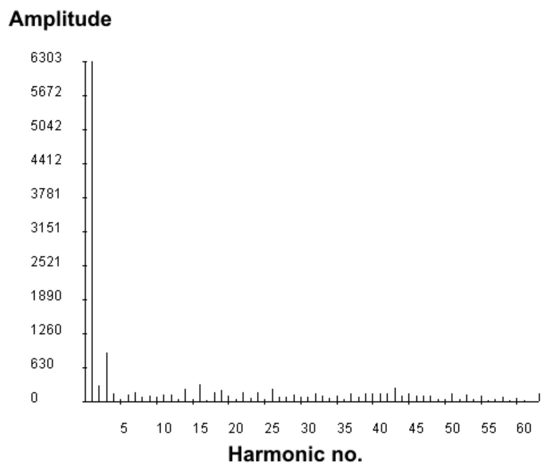


Figure 9. Spectral Snapshot of "t" at Time .035.

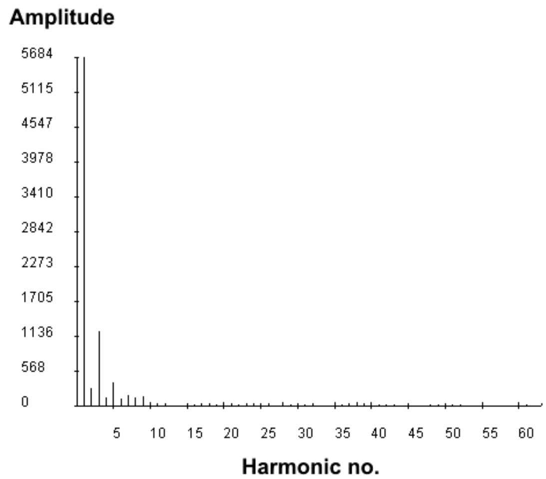


Figure 10. Spectral Snapshot of "oo" at Time .267.

We stored the amplitude data from the graphs in two Csound score function tables. We reserved the number at index 0 in the table for an amplitude at an inharmonic frequency (see the next section, *The Wah Sound*).

f51	0	128	-2	h0(v)	h1	h2	h3	…
f101	0	128	-2	h0(v)	h1	h2	h3	…

Although the synthesized "too" sound has a different time-varying spectrum (see Figure 11), the "oo" part does sound reasonably like the original sound. Using this method with one spectral snapshot is unlikely to capture the full quality of the noise in the "t" sound, but the result is sufficient for our musical purposes. An important feature of our design is that repeated performances of the same sound will have varied spectra in order to make the performance expressive.

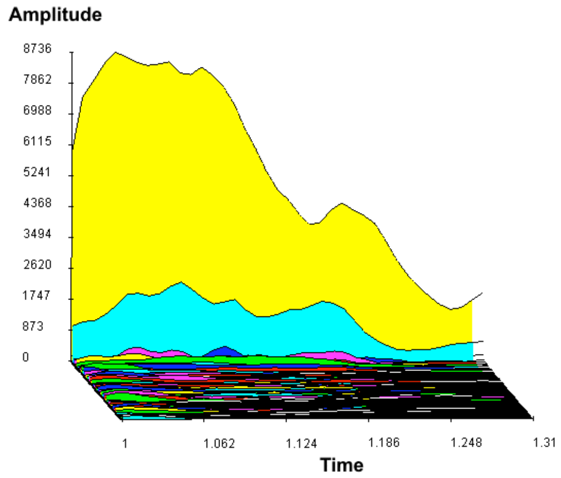


Figure 11. Time-Varying Phase Vocoder Analysis of Synthesized Didgeridoo B1 "too."

The Wah Sound

We divided the "wah" sound into two phonemes, a short "oo" for the "w" and an "ah" (see Figure 12). We chose the snapshot time for the "w" in the unstable beginning at time .049 seconds (see Figure 13). We chose the time for the "ah" near the end at time .659 seconds (see Figure 14) because this part of the sound becomes richer as it begins to morph into the next sound, another harmonic sweep.

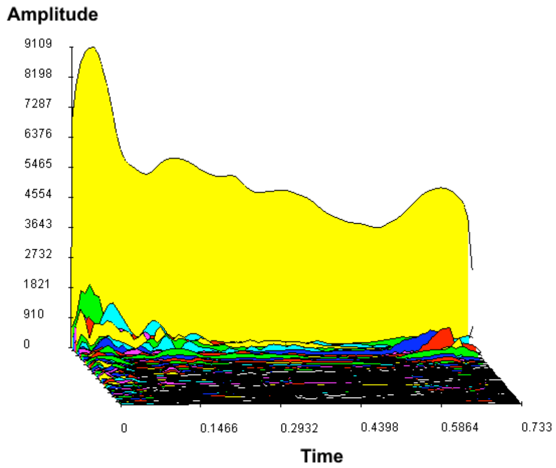


Figure 12. Time-Varying Phase Vocoder Analysis of Original Didgeridoo B1 "wah."

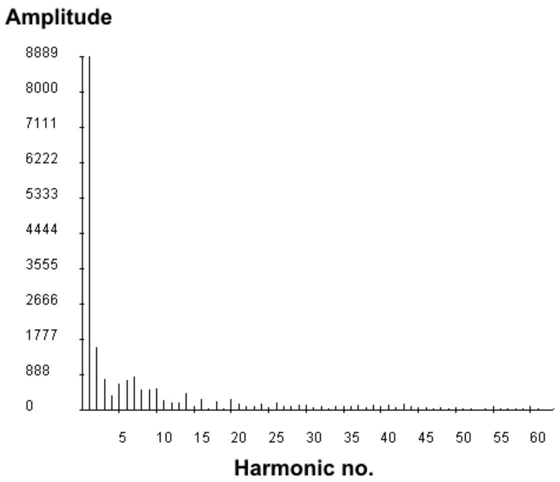


Figure 13. Spectral Snapshot of "w" at Time .049

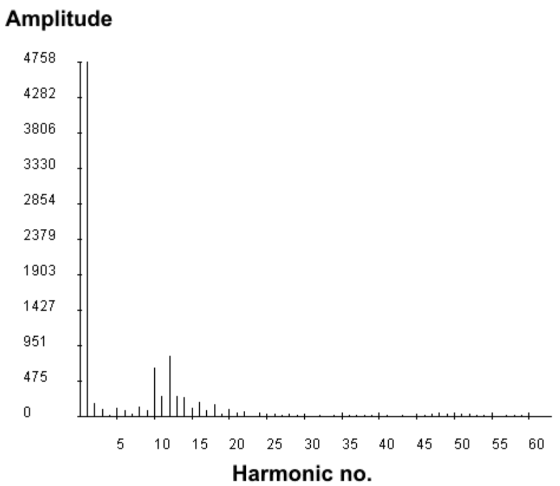


Figure 14. Spectral Snapshot of "ah" at Time .659

The waveform analysis (Hibbert 2003) summary showed that this sound has an extra inharmonic frequency of 174.5 Hz, which is 13 Hz flatter than the third harmonic of 187.5 Hz (see Figure 15).

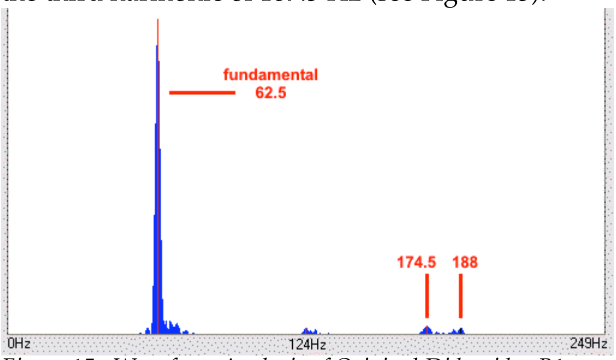


Figure 15. Waveform Analysis of Original Didgeridoo B1 "wah."

Interestingly, however, although the summary doesn't show it, the waveform analysis graph shows the expected harmonic at about 188 Hz and its amplitude is only slightly weaker than the amplitude of the harmonic at 174.5. We added the extra harmonic into the function table at time 0, after calculating its amplitude in relation to the amplitude of the fundamental. We set a slightly varying frequency of about 170-175 for this harmonic to model the natural frequency variations of a performer's voice. Although the synthesized "wah" sound has a different time-varying spectrum (see Figure 16), it sounds more like the original sound than the synthesized "too."

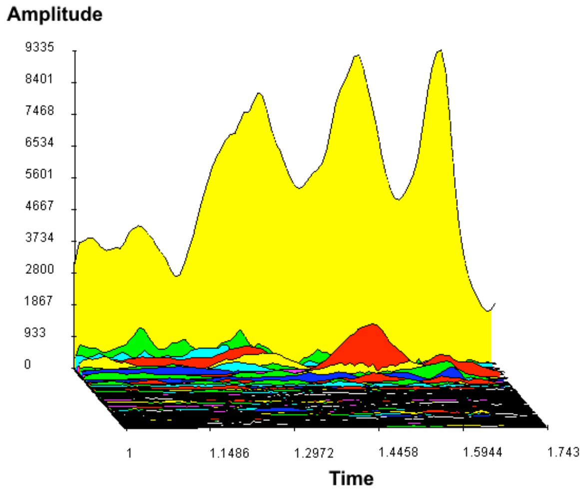


Figure 16. Time-Varying Phase Vocoder Analysis of Resynthesized Didgeridoo B1 "wah."

Musical Expression on the *Didgeridoo*

This section describes vocal sounds, harmonic sweeps and flutter tonguing, expressive techniques that increase the musical potential of the design.

Vocal Sounds

Didgeridoo players use a variety of vocal imitations of animal sounds in their performances. Some of these sounds are the owl, the dingo, the cockatoo parrot, the crow, the buffalo, the kangaroo hop, the kookaburra (laughing jackass), the brolgar (large crane), the emu (Australian ostrich) and the mopoke (boobook owl) (Hudson, 1994, Schellberg 1997, Dargan 2001). The imitation of the buffalo is a particularly gritty sound which uses a combination of vocalization and flutter tongue.

The Harmonic Sweep

The sound segment begins with a vocal sound similar to a smooth owl-like "who" descending from about B4 to E4. The *didgeridoo* resonance pulls the voice toward the *didgeridoo* harmonics similar to the stepped effect a sweeping band-pass filter produces. The resulting sound is more or less similar to the phonemes, "ah," "ee" and "oo." We divided the sound into one pair of slices for each of these phonemes. We chose the snapshot times at .066, .114, .323, .611, 1.136 and 1.487. After we produced a tone by morphing the six sounds, we added a band-pass filter sweeping from the 40th harmonic to the 20th harmonic. This method successfully captures this effect.

Morphing the Sweep Too Wah Sound

We morphed the sounds together to make the combined original sound, "ah-ee-oo too wah." In our previous additive synthesis instruments, we used line segments to connect the note parameters for slurs and morph the transition. In the case of the *didgeridoo*, we don't need to slur "notes," but the code for the slur feature is useful in connecting sound segments of different timbres. Each harmonic is one component signal for the spectrum, with its own frequency and amplitude line segments for the group of morphed notes. Figure 17 shows a Spectrogram analysis of the synthesized "ah ee oo too wah" sound.

Figure 17. Synthesized Harmonic Sweep and "too wah."

Another Harmonic Sweep

Another harmonic sweep moves up and down twice (see Figures 18-19). We found that for musical use of the sweeping band-pass filter effect, it was enough to set the envelopes determining the harmonic sweeping without concern for the exact original vowel sounds. That is, we got a pretty similar effect no matter which original sound we applied it to. Musically, we incorporated the sweeping band-pass filter into the piece with up to 5 (mostly) randomly determined moving segments.

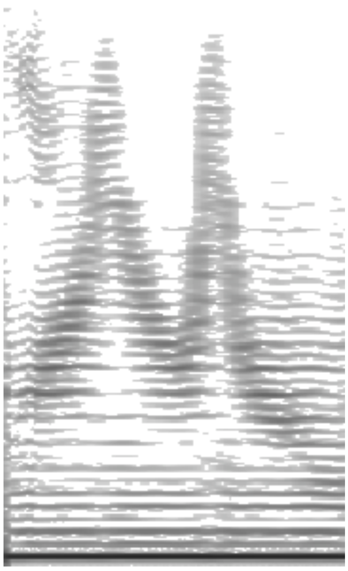


Figure 18. Multiple Harmonic Sweep in Original Didgeridoo B1 Sound.

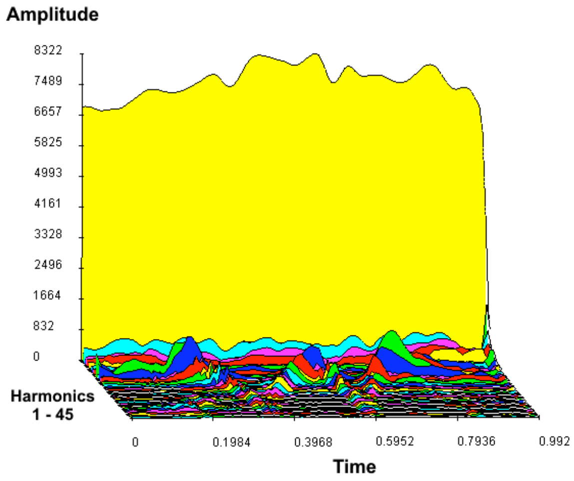


Figure 19. Time-Varying Phase Vocoder Analysis of Multiple Harmonic Sweep on Original Didgeridoo B1 Sound.

Flutter Tonguing

Didgeridoo players typically accomplish flutter tonguing by gargling in the back of the throat, which opens and closes the air stream. The amplitude modulation that we used in the previous wind instrument designs produces a close approximation to this gritty sound.

Examination of Ash Dargan's recorded flutter-tongued *didgeridoo* tones showed that the frequency of the amplitude modulation is approximately 20 Hz. The next step was finding a function to represent the average amplitude envelope of the individual modulations. We found that a simple amplitude envelope would suffice, with a minimum amplitude of 0% of the total amplitude (see Figure 20). The repeating amplitude function models the amplitude change for the flutter. The function does not need to model change of speed or jitter, so it can represent one average cycle of the flutter, and adjusting parameters such as the flutter rate randomly within their typical ranges can vary each cycle. Several score parameters control the flutter rate using a line segment. For example, our design uses four score parameters to control an initial, middle and final flutter rate, and the time required to change from the first flutter rate to the second.

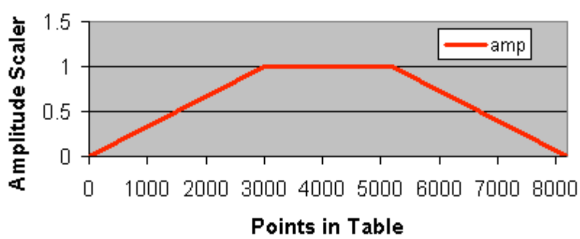


Figure 20. Flutter Tongue Amplitude Envelope Function.

Figures 21 and 22 compare the time-varying phase vocoder analyses of an original B1 *didgeridoo* flutter-tongued tone and a synthesized tone. Although the synthesized tone is less rich than the original, the perceived effect is still quite realistic.

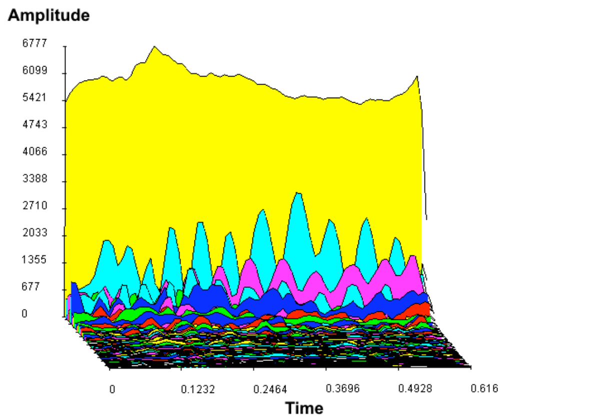


Figure 21. Time-Varying Phase Vocoder Analysis of Original Didgeridoo B1 Flutter-Tongued Tone.

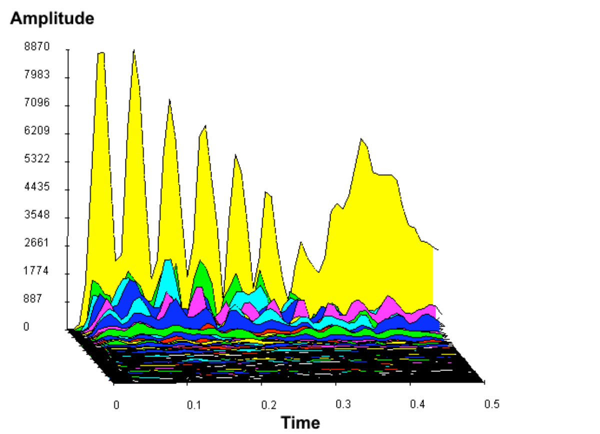


Figure 22. Time-Varying Phase Vocoder Analysis of Synthesized Didgeridoo B1 Flutter-Tongued Tone.

We can either apply the amplitude modulation to the correct part of the sound segment using a delay, or apply it to the entire segment. The instrument easily combines a flutter-tongued timbre with a group of timbres. We use a delay on the amplitude modulator to cause the modulation to begin at the same time as the sound requiring the flutter-tongue occurs in the group of sounds. Using the delay instead of just an amplitude envelope ensures that the flutter begins on the desired timbre.

Use in a Composition

We modelled the *didgeridoo* so we could use it in a composition. Figure 23 shows excerpted *didgeridoo* phrases from the recorded soundscape accompanying a live flute in Ayers' piece, *The Chalky Desert Where Nothing Grows*.

The design currently allows morphing up to 12 pairs of timbres in a connected group. The slurs in the figure indicate typical rhythmic groupings of timbres. As previously described, the flutter tongue may be applied to a single timbre within a group or to a whole group of timbres. We used random cells in a spreadsheet program to choose the *didgeridoo* sounds and vary their parameters.

We then copy and paste chunks of Csound score to increase the input speed. Repeated performances of the same sound will have varied spectra, making the performance more expressive.

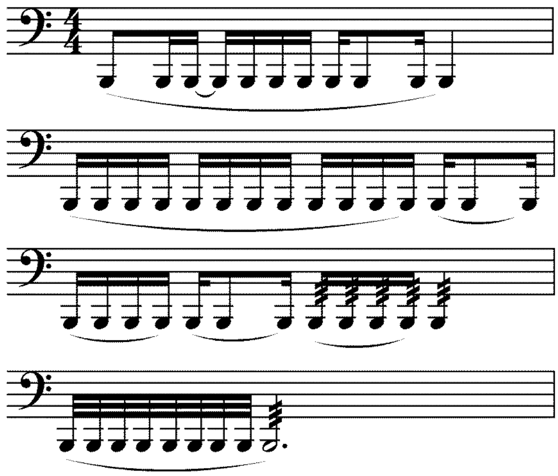


Figure 23. Excerpted didgeridoo phrases from Ayers' piece, ***The Chalky Desert Where Nothing Grows***

Conclusion and Future Work

We used additive synthesis with timbre morphing to model the dynamic spectra and musical expression of the *didgeridoo*. The design captures the subtle characteristics of many of its timbres, including vocal sounds, harmonic sweeps and flutter tonguing. Finally, an excerpt from a composition illustrated the expressiveness of the Csound design. Future work includes extending the model to synthesize more vocal sounds combined with *didgeridoo* sounds, along with modelling *didgeridoos* in other keys.

Acknowledgements

We would like to express our appreciation to Ash Dargan for giving permission to use his photo and to analyze timbres from his *didgeridoo* recordings to create the designs for this research. We also appreciate his input in personal correspondence regarding the results of this research. This work was supported by the RGC Competitive Earmarked Research Grant 613505.

References

Ayers, L. 2003. "Synthesizing Trills for the Chinese Dizi," *Proceedings of the International Computer Music Conference*, Singapore.

Ayers, L. 2004. "Synthesizing Timbre Tremolos and Flutter Tonguing on Wind Instruments." *Proceedings of the International Computer Music Conference,* Miami, Florida, USA.

Ayers, L. 2005. "Synthesizing Chinese Flutes Using Csound," *Organized Sound.* 10:1.

Ayers, L. and Horner, A. 2004(a). "Synthesis of Chinese Dizi Ornaments," *Ghost in the Machine: Performance Practice in Electronic Music*, Australasian Computer Music Conference. Wellington, New Zealand.

Ayers, L. and Horner, A. 2004(b). "Expressive modeling of Chinese folk wind instruments," *Journal of the Acoustical Society of America,* San Diego, California, USA, 116:4, part 2.

Ayers, L. and Horner, A. 2006. "Synthesizing the Chinese Suona," *Medi(t)ations, Australasian Computer Music Conference 2006*, Australasian Computer Music Conference. Adelaide, Australia, pp 15-22.

Braun, M. 2003. "Bell tuning in ancient China: a six-tone scale in a 12-tone system based on fifths and thirds," web page http://w1.570.telia.com/~u57011259/Zengbells.htm

Cottingham, J. 2004. "The acoustics of the khaen, bawu and gourd pipe," *Journal of the Acoustical Society of America,* San Diego, California, 116:4, part 2.

Dargan, A. 1999. *Wangurra* (Sleep) on *Ash, Dust & Dirt*. Indigenous Australia.

Dargan, A. 2001. *Didgeridoo Made Easy: A Beginner's Guide*. Indigenous Australia.

Dolson, M. 1986. "The Phase Vocoder: A Tutorial," *Computer Music Journal.* 10:4.

Fletcher, N.H. 1983. "Acoustics of the Australian didgeridoo," *Australian Aboriginal Studies*, 1

Hibbert, B. 2003. *Bell Waveform Analysis Program*. Great Brookham, Surrey, UK. www.hibberts.co.uk.

Horne, R. 2001. *Spectrogram*. Visualization Software. www.visualizationsoftware.com/gram.html

Horner, A., Ayers, L. and Law, D. 1999. "Synthesis Modeling of the Chinese Dizi, Bawu, and Sheng," *Journal of the Audio Engineering Society.* 47:12.

Horner, A. and Ayers, L. 1998(a). "Modeling Acoustic Wind Instruments with Contiguous Group Synthesis," *Journal of the Audio Engineering Society.* 46:10.

Horner, A. and Ayers, L. 1998(b). "Modeling Chinese Musical Instruments," *Proceedings of the 135th Meeting of the Acoustical Society of America*. Seattle, WA, 4.

Horner, A. and Ayers, L. 2002. *Cooking with CSound, Part 1: Woodwind and Brass Recipes*. AR Editions.

Horner, A. and Beauchamp, B. 1996. "Piecewise Linear Approximation of Additive Synthesis Envelopes: A Comparison of Various Methods," *Computer Music Journal.* 20:2.

Hudson, D. 1994. *Making and Playing the Didgeridoo*. Indigenous Australia. 1A2021DVD.

Kuchera-Morin, J. 2002. *Paleo* for Double Bass and Computer Generated Tape, Discordia Music, Chicago, Illinois.

Liu, W. 1989. *Entanglement of the Fenyang Tune and Baban* on *Chinese Wind Instrumental Music Vol. 2*. Hugo Productions (HK) Ltd.

Morrill, D. 1977. "Trumpet Algorithms for Computer Composition," *Computer Music Journal*, 1:1.

Neuenfeldt, K. (ed.) 1997. *The Didjeridu: From Arnhem Land to Internet*. John Libbey & Co., Sydney, NSW, Australia

Risset, J.-C. 1969. *Introductory Catalogue of Computer-Synthesized Sounds*. Bell Telephone Laboratories, Murray Hill, N.J.

Risset, J. C., and Matthews, M. "Analysis of Instrument Tones," *Physics Today*, **22**(2): 23-30, 1969.

Rodet, X. and Lefèvre, A. 1997. "The Diphone program: New features, new synthesis methods and experience of musical use," *Proceedings of the International Computer Music Conference.* Thessaloniki, Greece.

Schellberg, D. 1997. *Didgeridoo: Ritual Origins and Playing Techniques*. Binkey Kok Publications, Diever, Holland.

Tsai, C-G. 2003. *The Chinese Membrane Flute (dizi): Physics and Perception of its Tones.* Ph.D. dissertation under the supervision of Prof. Dr. Wolfgang Auhagen and Dr. René Caussé, Humboldt-University Berlin, Germany.

Tsai, C-G. 2004. "The timbre space of the Chinese membrane flute (dizi): Physical basis and psychoacoustical effects," *Journal of the Acoustical Society of America,* San Diego, California, 116:4, part 2.

Vercoe, B. 1992. *Csound: A Manual for the Audio Processing System and Supporting Programs with Tutorials.* MIT Media Lab, Cambridge, MA.

Wiggins, G.C. 1988. "The physics of the didgeridoo," *Physics Bulletin*, 39.

Wun, S. 2000. *PVan*. www.cs.ust.hk/~simon/pvan

AudioMulch 2006: a project review

Ross Bencina
www.audiomulch.com
rossb@audiomulch.com

Abstract

AudioMulch is software for real-time sound synthesis, music composition and performance-oriented audio processing. Version 1.0 of the software was released in February 2006 after an extended beta testing period spanning over 8 years. Following this release a new development cycle was commenced under the "AudioMulch 2006" project moniker.

This report presents as-yet unreleased work undertaken since version 1.0, including the port to Trolltech's Qt GUI toolkit, a new interactive patching user interface, and microtonal tuning table support. Successes and challenges encountered during the process of development are examined. During the project period collaborations with performers, composers, designers and audio engineers were undertaken to improve the software in various ways. The results of these collaborations are also discussed.

AudioMulch 2006 development activities were partially supported by The Australia Council for the Arts' "Sounding Out" instrument builders initiative, and by Arts Victoria.

Warren Burt
University of Wollongong; Illawarra Institute of TAFE
PO Box U-27
Wollongong, NSW 2500
Australia
waburt@melbourne.dialix.com.au

Cellular Automata as Spectra: Beyond Sonification into Composition

Abstract

Although Cellular Automata (CA) have been used in music composition for a number of years, most of the writings on them have either concentrated on the mechanics of the CA themselves, or else have dealt with implementations of CA in particular methods of sound synthesis. We instead concentrate here on the use of CA in "Scrabbles," a six-minute long musical composition, dealing mostly with the way the composition was structured using CA materials, and the algorithmic program that was written to choose parameters for the individual sounds of the piece.

Introduction

Although Cellular Automata (CA) have been used in music composition for a number of years, most of the writings on them have either concentrated on the mechanics of the CA themselves, or else have dealt with implementations of CA in particular methods of sound synthesis. Here we will instead concentrate on the use of CA in one particular musical composition, and talk mostly about the way in which the composition was structured using CA materials. Part of the maturation of areas of computer music is when the understanding of a technique is taken for granted, and is augmented by a deeper consideration of how and why particular techniques are used in particular compositions. Consider all the papers written on, for example, frequency modulation in the 70s and 80s. Today, fm is simply one more technique in the synthesis toolkit: the compositional use of the technique is now of primary concern.

History

There is an extensive literature on Cellular Automata and their use in music. The most extensive consideration of the inner structure of CA and musical applications is Dave Burraston's recently completed PhD thesis, *Generative Music and Cellular Automata.* (Burraston, 2006). Stephen Wolfram's *A New Kind of Science* is an exhaustive survey of CA (Wolfram 2002), and his research has recently been incorporated into an interactive website, *WolframTones,* in which one can hear musical realizations of various CA (Wolfram Research 2007). Unfortunately, I find most of the music generated with the algorithms on this site to be disappointing. An early software synthesizer which used CA in its granular synthesis engine was Eduardo Miranda's *Chaossynth*. Currently being rewritten, the earlier version is now available free from www.nyrsound.com (Miranda 1999). These are but 3 references to an enormous field. For more information, refer to the massive References section of Burraston's thesis.

As interesting as this research is, it did not engage me for this project. I was not so much interested in the mechanisms of CA, other than understanding them well enough to be able to know what was going on inside them, but was more concerned with using them as spectra, in order to get sounds with a certain degree of complexity. I have used CA before in a number of compositional contexts. They were used to choose pitches, durations and timbres, making clouds of slowly moving sound in *Voices, Tuning Forks, and Accordion* (1986), a composition for choir with microtonal tuning forks and microtonal accordion. The CA used for this was a mutant version of Conway's *Game of Life* written by my friend Dr. Henry Hunter. I also used CA to assemble extremely slowly changing inharmonic spectra in the computer composition *The Easy Beauty of Parallel Lives*, a movement of *Graphic Descriptions* (2003). The CA used for this piece was Peter Meyer's Q-Life. I also incorporated a one-dimensional CA into my early DOS interactive program *Randie,* described in "Interactive Improvisations with Electronic Music Systems," and with which I performed real-time interactive music on a laptop from 1991 on (Burt 1991). *Randie* produced MIDI output, which was performed with a variety of instrumental, electronic, and sampled timbres.

Sounds with Complexity – A Priority

For the current work, however, I was interested in using CA to assemble sounds with more complex, constantly changing timbres. This would entail finding CA which would have the right kind of changing graphics to produce sounds of the sort I was interested in. The CA produced would then be converted to sound using a graphic synthesis program. There are a number of these, such as the freeware *Coagula* for the PC (Ekman, 2003) and the commercial programs *Atmogen* (Sonorous Codes, 2004) for the PC, and *Metasynth* (Wenger, 2006) for the Mac. They all work on a similar principle, which is an extension of Western Music notation's up-and-down for frequency and left-to-right for time paradigm. That is, lines drawn closer to the top of the graphic produce higher frequencies, while lines drawn more to the right of the graphic produce later sounds. Additionally these programs add colour for placement in stereo space, and brightness for amplitude. All the programs generate sound with sine waves, but some, such as *Metasynth*, allow other samples to be used, while *Coagula* adds the feature that while pixels on the red-green continuum produce sine waves, the addition of blue to any colour adds a proportional amount of band-limited noise to the sound. For this composition, *Coagula* was used.

Notice that the conception of the kind of sound I wanted came first. I wanted to create sounds with rapidly changing, complex spectra. It occurred to me that CA might produce diagrams that, when synthesized with *Coagula*, could do this. *Sound* came before *technique*. Computer music, in general, has been, is, and, for the foreseeable future, as long as we are (rightfully) concerned with the developments of new tools and techniques, will continue to be, stuck in a scientistic rut. That is, as a field, computer music has the trappings of science (peer-reviewed papers, an emphasis on new uses of new technologies, etc), but lacks the ability to have rigorous proof that constitutes the essence of science. There is no way to prove music. (Success or failure in any conceivable market does NOT constitute proof of any kind whatever.) We should not pretend we are scientists. Nor do we need their approval (although their support would be nice). We are artists, and we operate with a different set of criteria. Only the quality of our compositional thought – that is, how we use our new tools – will make our efforts worth listening to.

Palette Choice as a Compositional Act

To see if any CA would produce the kind of diagrams that might be useful to me, I used *Mirek's Cellebration*, a freeware CA explorer for the PC (Wojtowicz, 2001). I wanted to synthesize spectra – that is, I wasn't interested here in synthesizing broadband noise, but sounds with discrete partials which changed frequency and amplitude rapidly. For just as most matter is empty space between fundamental particles, so most spectra (whether harmonic or inharmonic) consist of empty "frequency space" between partials. That is, a spectrum usually consists of a small number of partials at specific frequencies, not of partials at all available frequencies. This would mean that my CA diagrams would have to consist of mostly black, the colour used by graphic synthesis programs to indicate silence. Unfortunately, most CA programs completely fill in the image, as in Figure 1.

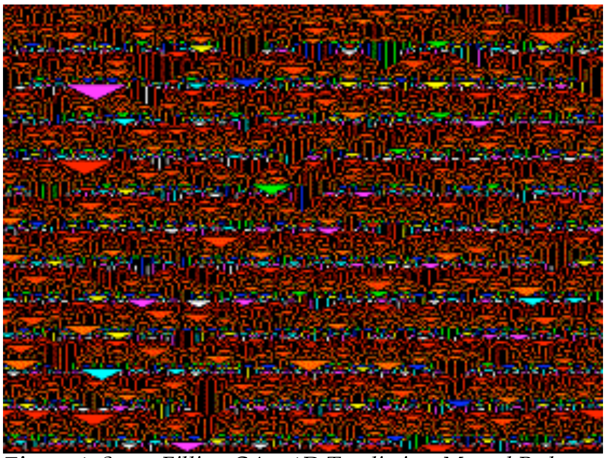


Figure 1: *Space Filling CA – 1D Totalistic – Marvel Rule*

In order to get diagrams that would be useful to me, I would have to either find a CA rule that generated mostly black, or else invent a colour palette that would "filter" the CA so that broad areas of black were generated. The "1D Totalistic" rules produced diagrams that consisted of mostly bands of activity occurring in horizontal lines, and I made a palette which consisted of alternating areas of black and colour, so that successive generations of the CA would be either sounding (coloured) or silent (black). By making the older generations all black, I guaranteed that most of the diagram would be black, with activity constrained to a certain number of horizontal regions. This would give me the kind of drawing I wanted. The "Coagula01" palette consisted of the following values and colours.

Generation	Colour
0	Black
1	Light Blue: R=128; G=255; B=255
2	Black
3	Red: R=252; G=0; B=0
4	Black
5	Green: R=0; G=255; B=0
6	Black
7	Light Yellow: R=255; G=255; B=128
8-24	Black

Table 1: *Colour listing of "Coagula01" palette*

Notice that the Light Blue and Light Yellow colours both have blue in them. This means that both those colours will produce some level of band limited noise. Applying this palette to the CA shown above in Figure 1 produced the result shown in Figure 2.

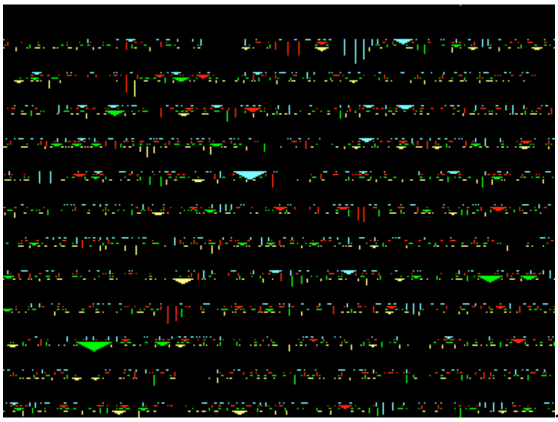


Figure 2: *1D Totalistic Marvel CA with Coagula01 Palette*

In this diagram, activity is restricted to horizontal bands, yet the kind of activity in each band is rapidly changing. This looked like it could produce the kind of sounds I was interested in. And in fact, when quickly processed by *Coagula*, it did indeed produce the kind of sonic complexity I had been searching for. My first impression on hearing the sound was that this was a sonic world I'd like to live in for hours. I quickly realized, however, that if I wanted the piece heard in the context of an academic conference, with its crowded schedules and severe time constraints, I would have to limit the duration of any piece made with these sounds. I therefore decided to make a short piece with many contrasting sounds, so that the range of sounds made with these graphics could be quickly heard. In my case, this is actually a welcome contrast, as my previous piece, *Proliferating Infinities* (2006), is just over 13 hours in length. This piece will probably remain unheard by all but a few friends until I can figure out a context in which it can be presented in such a way that its contemplative nature can be fully experienced. And as an aside, I wonder if it will ever be possible for ACMA, or similar organizations, to organize an event in which compositions of longer duration could be experienced in a non-pressured manner.

Exploration of the available 1D Totalistic CA rules gave me eight different rules which produced diagrams such as these. These rules, in Mirek's program, were called Class 4C, Forest, Gears 1, Marvel, No Name 1, Porridge, The City, and Tulips. All of these were different from each other, but all had activity occurring in parallel bands with changing activity within the bands. Of these, Class 4 C produced kinds of activity that changed over time, with alternating areas of straight vertical lines and smaller textures, as shown in Figure 3, in which a complex texture expands "out of" the area of vertical lines in the middle of the graphic. Each of the rules had their own intrinsic characteristics, and each produced differently detailed sonic textures with the larger family of sounds.

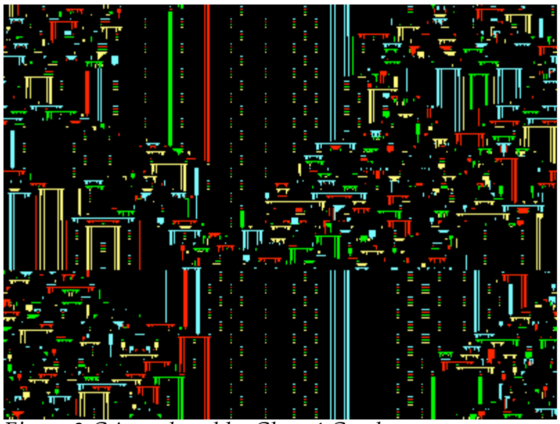


Figure 3 CA produced by Class 4 C rule

The City is similar to Class 4C. There are more areas of horizontal lines with small vertical "stalks", and less areas of large horizontal lines and empty space, but it has the same ability to have moving shaped frequency bands as Class 4C and Gears 1. The many long horizontal lines produce a series of rapidly sputtering very short noisebands.

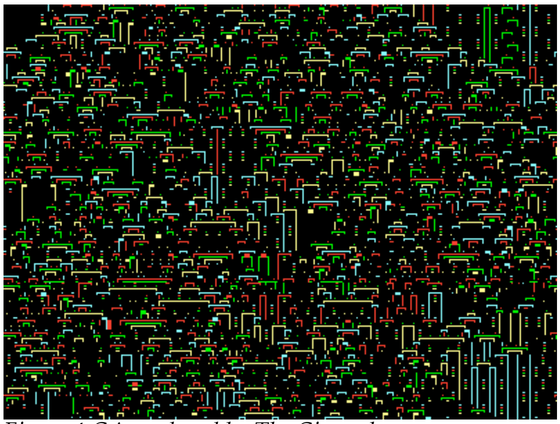


Figure 4 CA produced by The City rule

No Name 1 is similar to Marvel (see Figure 2), but has horizontal bands that are more narrowly spaced, and there are longer vertical lines which sometimes connect two adjacent horizontal bands. This make the spectrum thicker – more like a buzzing, dissonant tone cluster, than a well spread out chord with active spectral bands.

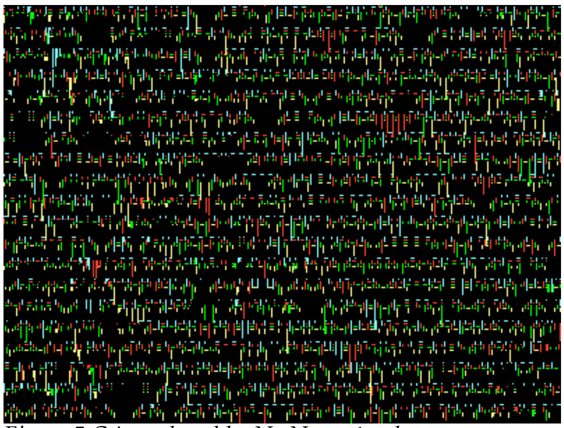


Figure 5 CA produced by No Name 1 rule

Forest is again, similar to Marvel, but extremely long vertical lines can appear. These make extremely short, sharp noise bursts, which make pleasing ornaments over the top of the overall active sonic texture. Forest can also have "holes" which by creating silence, give gestural shape to the overall texture.

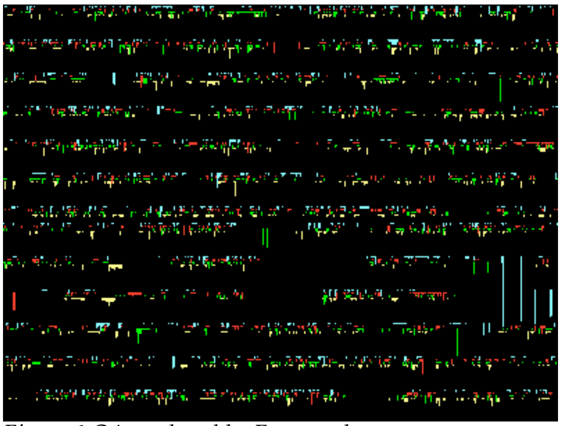


Figure 6 CA produced by Forest rule

Gears 1 produces bands of downward pointing triangles. This makes a quite surprising texture of rapidly attacking narrow noisebands that begin and change pitch unpredictably, while still retaining the sense of a sustained chord that occurred with Marvel. (It was this "sustained chord with great inner life" sound that first got me interested in exploring these particular CA further.) Very large "holes" can also appear in the Gears 1 diagrams, and these give a very pronounced sense of shaping of frequency and gesture to the overall spectra.

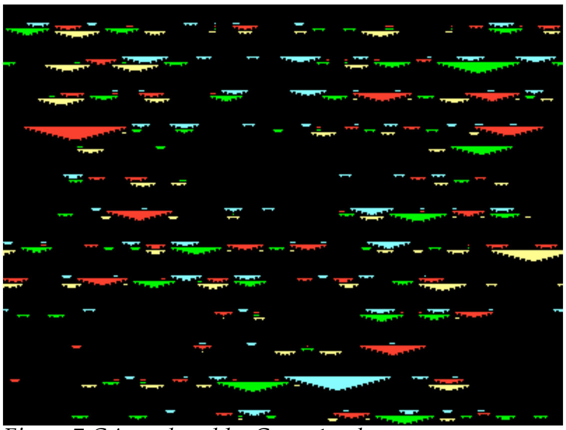


Figure 7 CA produced by Gears 1 rule

Porridge is a thicker, more regular version of Gears 1. Here the rapidly attacking narrow noisebands are very narrow, much more closely spaced, and the occur with much more regularity and faster than in Gears 1. The sound produced is somewhat of a cross between the sustained chord of Marvel, the sputtering noisebands of The City, and the tiny, expanding-then-contracting noisebands of Gears 1.

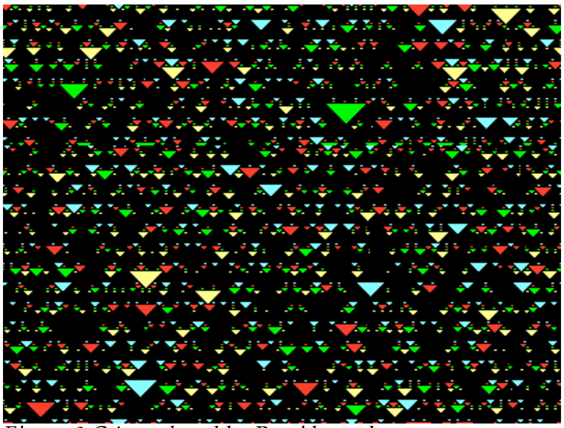


Figure 8 CA produced by Porridge rule

Tulips has very wide descending triangle shapes broken up by black vertical bands. These, surprisingly, only contribute a sense of wider noisebands to a texture that is basically very similar to Porridge or Gears 1. It should be mentioned again that all of these diagrams produce a similar family of sounds. The sonic details differ from rule to rule, but the overall sense of gesture – a sustained sound with a very rich and lively internal life – is similar over the whole family. This similarity is why these rules were chosen, and many other 1D Totalistic rules were rejected. I chose particular rules based on their ability to give me the kind of sound I wanted, trusting the mathematics of each CA to provide interesting internal details within my chosen overall sonic idea.

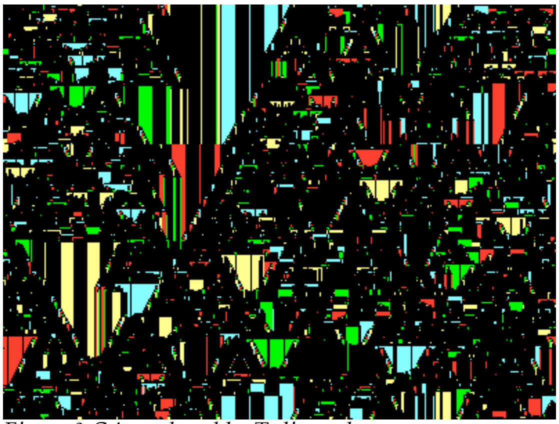


Figure 9 CA produced by Tulips rule

Choosing Parameters

Coagula turns graphics into sound, but has certain parameters that must be set. Among these are the duration of the sound, and the high and low frequencies that define the outer limits of the synthesis. Different settings of these parameters will radically change the nature of the sound. Additionally, I experimented with tilting the graphics slightly, by up to 5 degrees away from horizontal, to see if a sense of rising or falling pitch would be present in the spectra. It was, so I decided that some of the drawings would be tilted slightly to produce this as well.

After much experimentation and listening, I decided that there would be sequences of sounds made by these spectra, and that two sequences at a time would be mixed to make a more complex final result. Each sound would be made by a graphic produced by one of the eight possible rules. 70% of the time, the sound would be made by a straight graphic, while 30% of the time, the graphic would be tilted between 1 and 5 degrees in either direction. Each sound would last between 3.0 and 20.0 seconds, graduated in 10ths of a second. The selection of durations would be biased differently for each sequence of sounds. In some, short durations would be favored, in others, longer durations would be prominent. The use of graphics 400 pixels wide mean that at the longest duration, there would still be 20 pixels each second, thus avoiding the problem of the "rhythm of the pixels" becoming a predominant aspect of the sound. The lowest frequency at the start would be between 50 and 800 hz, while the highest frequency at the beginning would be between 800 and 12800 hz. The CA height of 300 pixels would insure that the spacing of possible frequencies would always be less than 35 cents per step. This would avoid the problem of inappropriate melodic material being formed through too wide spacing of frequencies. The ranges of the parameters are summarized below.

CA type	Class 4C, Forest, Gears1, Marvel, No Name 1, Porridge, The City, Tulips – chosen with equally-weighted random numbers
Straight or tilt?	70% - straight; 30% tilted – 3% each 1 to 5 degrees clockwise or 1 to 5 degrees counter-clockwise.
Duration	From 3.0 to 20.0 seconds. Biased random choice – linear 2x and linear 4x distributions (favoring low values), and linear inverse 2x and linear 4x distributions (favoring high values).
Low frequency at start	From 50 to 800 hz, chosen in octaves with equally-weighted random numbers
High frequency at start	From 800 to 12800 hz, chosen in octaves with equally-weighted random numbers

Table 2: *Parameter choices and ranges*

To make the decisions for each individual sound, an algorithmic gesture generating program was written in John Dunn's *ArtWonk* (Dunn 2007). I knew very well the range of sounds I wanted. Using algorithmic methods to choose the exact details of the sound allowed me to experience a sense of surprise as I observed the structure of the piece emerging, sound by sound. It also allowed me to get results I couldn't have gotten by picking parameters by instinct. And I have worked with algorithmic compositional methods so long that by now, my compositional instincts for choosing which method of choice to use for each parameter of a sound are highly developed. This led, for example, to using a different random distribution for choosing durations in each of the four lists. I knew the kind of spread of durations that would be generated by each distribution, and I chose each one accordingly.

Table 3 gives a listing of the durations produced by each of the four different Linear Nx distributions. Durations for each list were chosen from between 3.0 and 20.0 seconds, and were totaled up for each new sound. When the total duration exceeded 180 seconds, or 3 minutes, that list was ended. Reflecting the increasing average durations of the sounds in each list, the four lists had 30, 20, 14 and 11 sounds respectively, as the average duration increased from short to long.

Looking down each list, one can quickly see how, for example, Linear 4x chooses mostly durations at the very low end of the range, while Linear 2x chooses a low weighted selection but one more biased towards the middle, while the inverse distributions simply reverse this bias (I was delighted, by the way, that the first list began with the highly improbably long 18.3 second duration).

It might be mentioned that these distributions were part of a package of about 30 different random distributions, functions and attractors I wrote for John Dunn's *ArtWonk,* which have since been incorporated into the program. Since designing these functions, I have spent a considerable time playing with each, until I now almost intuitively know which distribution is going to give me what kind of control for a given parameter.

List #	1	2	3	4
Random Distribution	Linear 4x	Linear 2x	Linear Inv. 2x	Linear Inv. 4x
Durations	18.3	6.4	18.3	19.6
	3.4	6.2	17.6	11.2
	8.4	3.4	18.5	19.6
	8.1	11.1	16.1	15.7
	5.2	8.4	7.9	16.2
	4.3	14.7	16.8	15.1
	10.2	8.1	11.6	15.5
	4.7	14.4	9.8	16.3
	6	10.9	7.3	16.4
	6.9	6.8	10.6	19.8
	4	14.4	6.8	11.4
	3.8	5.2	14.4	
	3.5	6.3	10.2	
	3.7	14.3	16.6	
	3.7	4.3		
	3	10.2		
	3.8	10.8		

	10	10.8		
	3.2	4.7		
	5.6	13.8		
	5.3			
	6.2			
	4.3			
	5			
	8.3			
	5.7			
	14.2			
	3.6			
	4.6			
	4			

Table 3: *Durations chosen by different random distributions*

In each subsequent list, as well, increasing frequency ranges were chosen. Frequency was chosen in a two stage manner. In the first stage, a random number, 1, 2, 4, or 8, is chosen with an equally weighted random distribution. In the second stage, a random number between either 50 and 100 (for low frequencies) or 800 and 1600 (for high frequencies) is chosen. These two numbers are multiplied together to get the required low or high frequency limit for the synthesis. This two stage method generates frequencies with an even spread of frequencies among all the octaves. Simply choosing a random number, even with equal-weighting, between the low and the high frequencies of a particular band would tend to not give frequencies evenly distributed among all the possible octaves (especially in the short term, when only 30 or less values are being chosen). To insure that all octaves were visited, the two stage method described here was made. In Lists 2, 3, and 4, the high frequency band was increased by one octave. In Lists 3 and 4, the low frequency band was dropped by one octave. In list 4 as well, choices were made from only the top two and the bottom two octaves. This meant that each subsequent list covered, on average, a wider pitch band than its predecessor. When the lists were mixed into the final sections, this meant that the piece would not only slow down, it would cover a wider frequency range at the end than it did at the beginning.

In the *ArtWonk* patch, pressing a button gave new values for each new sound. As shown, these were written into lists. The interface for the program is shown in Figure 10.

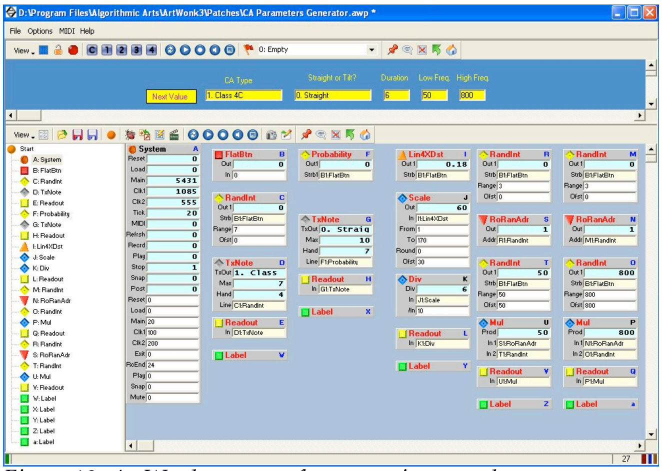


Figure 10: ArtWonk program for generating sound parameters

To sum up the overall process of making the piece: the specified CA was first generated in *Mirek's Cellabration*. The CA graphic was then processed and, if necessary, tilted in a graphics program. The processed graphic was then converted into sound with *Coagula*, using the parameters generated by *ArtWonk*. The generated sounds were spliced together end to end in a sound editing program. Lists 1 and 2 were mixed together to make a 3 minute section. Lists 3 and 4 were also mixed together to make another 3 minute section. These two 3 minute sections were then spliced together to make the final composition. Since the durations of Lists 1 and 2 are shorter than the durations for Lists 3 and 4, this means that the section 1 has a faster composite rhythm than section 2. This gives a slight sense of deceleration to the complete work. This sense of deceleration is one which I seem to be exploring in a number of current works. Maybe it reflects my fervent desire to slow down the pace of my life.

Conclusion

At each stage of composing the piece, I first made decisions as to what kind of sound I wanted, or the range of values wanted, or the desired distribution of values. Parameter ranges or distributions were then evaluated by working interactively. Knowledge of the kind of results produced by each selection method allowed me to pick methods appropriate for each parameter, and gave me the ability to proceed with composing rapidly and intuitively.

While composing, I was afraid that the "event-oriented" nature of my decision making would make a piece that would seem too "bitsy" or too rapidly changing. On listening to the completed piece, I realized that my fears were groundless. The different frequency ranges, rhythms, and durations of each sound simply contribute to a sense of rapid variation within an overall highly active sound, all the composite sounds and frequencies fusing into a single lively texture.

The final composition, "Scrabbles," achieves my desire – a science of complexity is used to make a complex sound object, one with great inner life and dynamism. Its immediate predecessors might be pieces such as Iannis Xenakis' *Bohor* (1962) and Larry Austin's *Caritas* (1969), although with its short duration and rapid textural changes, it perhaps can be seen as a rapid slideshow of snapshots in contrast with the vast frescos of those two magisterial compositions.

References

Austin, Larry (1969) Caritas, computer music on tape. Source Records, SR17, CPE, 1970.

Burraston, D. (2006) Generative Music and Cellular Automata. PhD Thesis, Univ. Technology Sydney, Australia. http://www.noyzelab.com/research/research/thesis.zip

Burt, W. (1991) Interactive Improvisations with Electronic Music Systems. *Sounds Australian*, Summer 1991-92.

Burt, W. (1986) Voices, Tuning Forks and Accordion, on *Music for Tuning Forks*, Tropicapricorn CD, www.tropicapricorn.com.

Burt, W. (2003) The Easy Beauty of Parallel Lives from Graphic Descriptions, on *Road Works – Some Numbers – Graphic Descriptions – Stretti,* Tropicapricorn CD, www.tropicapricorn.com

Burt, W. (2006) *Proliferating Infinities*. Unpublished 13 hour composition for harp samples in 264 different microtonal scales.

Dunn, J. (2007) ArtWonk (3.0 beta) http://algoart.com

Ekman, Rasmus (2003) Coagula – graphic sound synthesis freeware http://hem.passagen.se/rasmuse/Coagula.htm

Miranda, E. (1999) Chaossynth, software synthesizer. http://www.nyrsound.com/

Sonorous Codes (2004) Atmogen – graphic sound synthesis software. http://www.sonorouscodes.net

Wenger, E. (2006) MetaSynth4 – graphic sound synthesis software. http://uisoftware.com/MetaSynth/

Wojtowicz, M. (2001) Mirek's Cellabration – freeware cellular automata explorer. http://www.mirwoj.opus.chelm.pl/ca/index.html

Wolfram Research (2007) WolframTones: An Experiment in a New Kind of Music – interactive website. http://tones.wolfram.com/

Wolfram, S. (2002) A New Kind of Science. Wolfram Media, Champaign, Il.

Xenakis, Iannis (1962) Bohor, musique concrete on 8-channel tape on *Xenakis: Electronic Music*, EMF CD, EM102 www.cdemusic.org

The Morph Table: A collaborative interface for musical interaction

Andrew R. Brown ^*
René Wooller ^
Kate Thomas ^
^ Queensland University of Technology
2 George Street
Brisbane, 4000.
Australia

* Australasian CRC for Interaction Design (ACID)
"The Works" Musk Avenue, Kelvin Grove, 4059
Australia

r.wooller@qut.edu.au
a.brown@qut.edu.au
kj.thomas@qut.edu.au

Abstract

The Morph Table is a new music interface designed for collaborative music making. It comprises a software system that generates transitions and variations (morphs) between MIDI-based musical material, and a table-top hardware design on which cubes representing algorithmic parameters are moved around to control generative music. Like other table-top interfaces the size and multiple objects afford social interaction. The generative music system of the Morph Table makes it particularly suited to installations and use by inexperienced users. This paper outlines the design and usage features of the Morph Table.

Introduction

The Morph Table is a collaborative musical system which allows participants to morph between predefined sets of music through physical moving cubes on a table top surface. The design of the Morph Table involved some new developments, both in the areas of table-top interface and note-level morphing algorithms. Firstly, we will outline some of the previous research into both table-top interfaces and morphing algorithms. Following this is an overview of the musical morphing software, LEMorpheus and an outline of the design of the Morph Table. The design of the table and cubes is given priority, as the morphing algorithms have been described previously (Wooller and Brown 2005). Finally our informal observations and experiences of interacting with the Morph Table will be provided.

Previous research

Blaine and Fels (2003) claim there has been a history of electronically mediated musical collaboration from the earliest days of electronic music, citing Stockhausen's *Mikrophonie I and II* (1963/1964) experiments with professional percussionists, choirs and ring modulators. In the fifty years since these initial experiments involving electronics, the rapid and radical development of computer technology has had significant consequences and implications for musical composition and interaction potential.

There has been a strong movement in the past two decades towards developing new ways for people to interact with music via computers and non-traditional interfaces; evident internationally in the popularity of the NIME conference series, and locally with the Australia Council's SoundingOut grant scheme that resulted in web sites such as Digital Instrument Building (Brown et al. 2003) and Clatterbox (Bridgeman 2004) and numerous new instrument designs. This trend has seen the development of a wide range of new musical interface designs, many based on table-top and/or video-tracking interfaces like the Morph Table.

Figure 1. The Morph Table in use.

Table-top interfaces

There have been a range of previous table-top interfaces for music, and the target audiences and interface designs have varied widely amongst them. Kaltenbrunner and Bencina (2007) provided a comprehensive review of these, however, we will summarise some of the work to provide a context in which to situate our Morph Table project.

Tabletop interfaces have been quite popular in recent years and they generally shared the feature of direct manipulation of objects on the table to control music or sound structures or parameters. The types of objects and methods for monitoring the movement of them have varied somewhat and include video tracking of objects, imbedding buttons and triggers in the table surface, such as in Composition on the Table and Jamodrum (Blaine et. al., 2003).

One of the earlier table-top interfaces was the Audiopad (Patten et al. 2001) that used Radio Frequency tracking of objects on a tabletop surface. A projector mounted above the table provided graphical animations that showed the effect of and relationship between the objects. A variety of synthesis controls were simulated by the interface, as it controlled the playing of electronic dance music.

Music Table (Berry et al. 2003) used an overhead camera to track markers on cards laid on the table using the Augmented Reality (AR) Toolkit system. A computer monitor displayed the video camera image with graphics overlaying the card markers. Each marker represented a musical pattern. The user could structure the work by adding, deleting and arranging cards on the table.

The ReacTable featured a translucent round table, on which are positioned objects of different shapes. A video camera situated beneath the table tracked the objects position and orientation. Object position was mapped to the "topological structure and the parameters of a sound synthesizer" that controled connections between and parameters of various audio signal processing patches (Jorda et al. 2005).

While these projects may have similar goals or use similar technologies to ours, the Morph Table uniquely combines musical morphing algorithms and a table-top interface for users with limited previous experience. While these systems often used projections to provide user feedback, the Morph Table relies on simple mapping and object labelling to enable the object position to be self-descriptive. This also simplifies the system installation, and cube lighting effects were are an alternate visual stimulus that enable operation in dimly lit environments.

Both the Morphing Table and the ReacTable utilise the ReacTIVision (Kaltenbrunner & Bencina 2007) video tracking system, working with fiducial markers and live video data streaming. As a result they have similar potential gestural approaches, however the Morph Table differs from the ReacTable on a number of key factors.

Music generation: The ReacTable is an interface into an audio DSP patching environment whereas the morph-table operates on the note-level and enables separate parts within the music to be morphed independently between pre-composed patterns.

Cubes: The ReacTable uses 'pucks' – each puck displaying a single fiducial while the Morph Table uses cubes which have at least four fiducials on different sides of the cube allowing for change and variation by rotating the cube.

User Feedback systems: The ReacTable uses a highly complex system for user feedback involving a frosted, curved tabletop to display projected graphics generated directly from the players gestures. The Morphing table, however, relies purely on the placement of cubes and aural feedback.

An example of a table top interface that uses blocks but does not use video camera to track them is the BlockJam project (Newton-Dunn et al. 2003). In this system one block was connected directly to a computer and there were physical connections between blocks as a way of structuring musical sequences. Blocks were functionally homogenous on a broad level (either a 'play' or 'path' block), and heterogeneous in their potential musical role as the user could 'dial' or scroll through different instruments or sequential directions in each block.

BlockJam requires blocks to be physically connected, created a series of binary states between blocks. It relies on sequences and encourages the user to think sequentially and two dimensionally. The function and state of the blocks, and the musical paths created by their arrangement were visible in the positioning blocks relative to each other on the table.

The Morph Table software directly addresses one of the key challenges unable to be addressed with the BlockJam system; continuous control of musical expression. Unlike block jam, which is "an alternative means of controlling a sequencer" (Newton-Dunn et. al, p. 392) the morphing software is able to work with continuous control and gesture to allow for greater musical sensitivity and expression.

Previous note-level morphing algorithms

The Morph Table is used to control note-level musical morphing algorithms. We have published a comprehensive survey of compositional morphing algorithms at ACMC (Wooller 2005), however, a short summary of this will be provided here.

A number of morphing and morph-like algorithms have been developed previously, the most significant projects are: GRIN (Mathews and Rosler, 1967), HMSL (Burke et. al. 2005), DMorph (Oppenheim 1995) and The Musifier (Edlund 2007).

GRIN is of particular historical significance, being the first attempt at algorithmically morphing between one note sequence and another. Mathews used a pen interface to draw envelopes representing note parameters. Polansky, using HSML, developed a number of specialized morphing algorithms for various compositions

(Polansky and McKinney 1991; Polansky 1987; ibid. 1991; 1992; 1996; 1996b).

Oppenheim's DMorph had some unique features such as the separate morphing of rhythm and pitch as well as n-source morphing. Interestingly, DMorph uses a 2D Cartesian plane (like a table) to control the morph index between four different sources (Oppenheim 1995; 1997; 2006).

The Musifier introduced the ability to morph between two different MIDI instruments by sending note information to each and fading the source out while fading the target in. Another significant feature is fact that the tonal representation is morphed separately to the note-level information.

The LEMorpheus morphing system used for the Morph Table builds on many of the features of the previous approaches, while also making some new contributions. Crucially, LEMorpheus utilizes an evolutionary approaches to morphing, which has remained relatively unexplored. The morph table is the first of these systems to incorporate a physical interface for controlling a morphing algorithm in realtime, and to allow collaborative control of the morphing algorithms.

Musical Morphing on the table

Morphing is defined as a hybrid transition between two pieces of music, called the *source* and *target*. In the morph table, there are four different voices in the music – drums, bass, lead and pads – and each of these can be morphed independently by moving the cube that relates to that part along the length-wise axis of the table. For any particular voice, when the cube is on one side of the table the *source* pattern is played, while on the other side of the table, the *target* is played. Moving the cube from one side to another will elicit a note-level or "compositional" morph between the two patterns, a process typically involves automated key-modulation, cross-fading of the different timbres as well as mutation, selection and substitution of note patterns. Moving the cube breadth-wise across the table controls the level of audio effects that are applied to that voice. Other morphing algorithms may also be used, including linear interpolation between pitch, duration, note-onset and velocity envelopes; as well as probabilistic generation of material based on analysis of source and target.

There are four different sets of music, each comprising different pre-composed source and target patterns and pre-rendered morphs. Each of the four sets relates to one of four fiducial positioned on the sides of the cube. Each voice can be switched to any particular set independently by flipping the cube onto the side that (arbitrarily) relates to that set. This changes the timbre, source and target pattern and morph for that voice.

With the possibilities of each cube being either on the source, target or off the table, combined with four different sets of music, there is a total of 6561 discrete musical combinations. This is arrived at because for each cube there are 9 discreet states:

1. not on the table
2. pattern 1, left side (morph index = 0)
3. pattern 1, right side (morph index = 1)
4. pat. 2, L
5. pat. 2, R

Etc.

Imagine if there were two cubes - for each of the 9 states of one cube, we could put the other cube in any of it's 9 states, thus 9^2. Therefore with 4 cubes we have 9^4 possible discreet combinations, 9^4 = 6561.

When the near continuous morphing between source and target is factored in, the musical possibilities are dramatically increased further, allowing a substantial degree of flexibility, especially considering the pre-composed nature of the music.

It is not within the scope of this paper to provide a detailed explanation of the morphing algorithms; more detail is available from (Wooller and Brown 2004) and (Wooller 2005).

Morph table design

The Morphing Table is a combination of several interface components. It consists of a flat surface, the tabletop, upon which three-dimensional cubes are positioned. Four of the cube's six sides are covered with fiducials (high contrasts pattern markers) and the remaining two sides are used for cube labelling (bass, drums, and so on). The tabletop is Perspex, and a web camera is placed underneath the table such that it can capture the movement of the cubes on the table surface. The table needs to be of a certain height to be both comfortable for the user and to optimise the working range of the web camera. The camera covers an area up to approximately 20 cm inside the perimeter. This area, we call the "beer-resting" zone, can (also) be used for printed instructions, logos, and as a "dead" zone for cubes not currently in play.

The table was also designed with performance and storage in mind, and therefore was able to be disassembled by removing the legs. The screws on the inside of each corner do this and as they fall within the "beer-resting" zone and do not interfere with the camera view.

The cubes are also made of Perspex, and are internally lit by a fully contained LED system. Four faces of the cubes display four different fiducials, and the remaining two faces display text to identify which instrument set the cube is manipulating.

A Mac mini computer runs the reacTIVision software to track the location of the cubes and communicates their positions via Open Sound Control (OSC) to a Windows computer running the LEMorpheus morphing software. From the position of the cubes, the required morph indices and sound-FX levels for each part are extracted. The morph indices influence the note-stream that is generated by the morphing algorithm. Both the note stream and the sound-FX levels are sent to a software synthesiser which renders them into audio. The synthesizer patch itself is quite sophisticated, incorporating 32 separate synth modules (Reason 3 "combinators"), each with a specialised patch of sound-FX.

Table Lighting

Variations in lighting had significant impact on the ability of the web camera to accurately detect and track the fiducials. Tabletop reflectivity required consideration of how external lighting would affect the camera tracking; light from directly above the table could interfere with the camera, light from underneath the table could reflect off the surface making it more difficult to read the fiducials. We decided that lighting the cubes internally would help in making them more durable in a range of lighting conditions. The success of this approach was clearly evident at performances which began in the afternoon and carried through until past sundown. The changing light throughout the afternoon at these venues had little effect on the successful functioning of the table; with only a few minor adjustments being required to the camera settings.

Video tracking

The reacTVision system, used for cube tracking, utilises black and white markers developed by Ross Bencina that he calls amoeba fiducials (Bencina and Kaltenbrunner 2005).

Figure 2. A fiducial marker identifies cube surfaces.

Fiducial Size

For greater accuracy (detection and tracking response) we found that it was best to use quite large fiducials so that the variations were more pronounced, because they were then less susceptible to changes in lighting, and were able to operate over a wider range on the table with ease. The large fiducials also compensated for problems with recognition that were exaggerated by the "fisheye" effect of the *creative Live! Ultra*'s wide angle (72 degrees) web camera lens used to capture more of the rectangular tabletop area.

Cube design

The Morphing Table cubes have fiducials displayed on four faces, affording more musical potential to each physical object than single sides objects used in most previous table-top systems.

Figure 3. Morph Table cubes.

Cubes operate on multiple faces, and the downward facing fiducial indicates the current state of the block. In contrast to Block Jam which uses two types of blocks ('play' or 'path' blocks) which have different functions, each cube in The Morph Table is functionally the same, and is mapped to a different instrument. Each cube represents a different musical part (instrument) and each face of a cube represents a different musical riff for that part. When cubes are in view the riff plays, when lifted above or to the edge of the table the part ceases at the end of the currently playing riff.

Cube Lighting

Several obstacles were encountered when designing the internal lighting system for the cubes. A system of LEDs was used after much experimentation. This component of the interface construction was considerably more time consuming than any other aspect, as the design was highly sensitive to the diffusion of light across the surface. Uneven lighting lead to areas of over- and under-exposure which, in turn, resulted in a cube being misidentified, or unable to be detected at all. Experiments to rectify this included printing the fiducials onto paper of varying translucencies, adding a variety of foams and packing materials between the LEDs and paper, using aluminium foil to reflect light from over-exposed areas to darker areas, and adjusting the angles of the LEDs. The most satisfactory result involved angling the LEDs towards the edge of the card, and printing the fiducials onto tracing paper. The fiducials were then darkened with a marker to create maximum contrast, and the 'dots' were slightly enlarged to compensate for over-exposure.

The delicate electronic circuit needed to be durable, compact, and light weight in order to be contained within the cube, and easily serviceable. The final system consisted of an LED circuit on Vero Board, mounted on thick card and was powered by a two AAA batteries. Each side was lit by 9 5mm LEDs, making for a total of 36 LEDs running in parallel. The cardboard provided stability by reinforcing the circuit board. The shape of the card allowed the circuitry system to be fitted snugly into the cube without the use of adhesives or other attachments. It also aided with diffusion, a major consideration, as the white surface of the card reflected the light.

Interacting with the Morph Table

We installed the Morph Table in several public venues around Brisbane, Australia during 2006. In these venues the morph table was used as both a performance instrument and as an interactive installation for the public. Informal usability observations were taken at these events from which we conclude the following preliminary usability comments.

Blaine and Fels (2003) suggest initial negotiation of user roles and protocol that may be meditated by a more experienced player or instructor, particularly in the case of musical novice, however they also highlight that the novelty of the interface may be unfamiliar to 'expert' musicians also, and hence require some basic instruction. In the case of the morph table, it was decided that instruction should be minimal to observe transparency of operation, learn-ability and playability of interface. This placement of the camera under the table, as with the reacTable, allowed for more free-flowing user interaction, eliminating the need for the user to consider interference with the video tracking. The use of large, clearly labelled, cubes made distinctions between cubes quite apparent. There were some difficulty in differentiating between sides on the cube; hard to tell which patterns/combinations they had already played.

It was clear that most users enjoyed the experience of playing the Morph Table. Participation was quite intuitive and, with a limited set of instructions mostly relating to how the table axis were mapped to musical parameters and the effect of flipping the cubes, users interacted with the table for up to half an hour. However, we also observed that the length of time spent with the interface varied between users, and that people who were more familiar with the computational arts processes at the heart of the morphing algorithms spent longer with it. This indicates that even more clear instructions and more obvious musical changes may assist novice users. The interaction between users, especially with regard to passing on learned performance "tricks," was repeatedly facilitated by the Morph Table.

Conclusion

In this paper we have outlined the morph table project. We have surveyed the research context for table interfaces, and provided a review of the musical morphing software. We have detailed the design and construction of the Morph Table and shown how we have balanced ease of use and simplicity of construction and operation in order to maximise user engagement with music making.

Future avenues for research with the Morph Table include redesigning the project for new contexts and audiences, including installation, performance and possibly as a component of an ensemble. New contexts and audiences could call for further experimentation with musical content and the application of the morphing software, as well as interface considerations.

The development of the physical interface design holds much potential. Refining it towards a more economical and durable model, which would involve further material and lighting experimentation, could create new application possibilities in the educational and commercial arenas.

There is video documentation of the Morph Table available on YouTube. http://www.youtube.com/watch?v=nKXhfApKCms

References

Bencina, R. & Kaltenbrunner, M. (2005) "The Design and Evolution of Fiducials for the reacTIVision System," Proceedings of the Third International Conference on Generative Systems in the Electronic Arts (Third Iteration), Melbourne, pp. 97 - 106.

Berry, R., Makino, M., Hikawa, N. and Suzuki, M. (2003) The Augmented Composer Project: The Music Table. *Proceedings of the 2nd IEEE and ACM International Symposium on Mixed and Augmented Reality*. Washington, DC: IEEE Computer Society, p. 338.

Blaine, T., Fels, S. (2003) Collaborative Musical Experiences for Novices. *Journal of New Music Research*, Vol. 32(4), pp. 411-428.

Bridgeman, S. (2004) *Clatterbox*. http://www.clatterbox.net.au/

Brown, A. R., Doornbusch, P., Sorensen, A., McIlwain, P. and Riddell, A. (2003) *Digital Instrument Building*. http://digitalinstruments.ci.qut.edu.au/

Davidson, P. and Han, J. (2006) Synthesis and Control on Large Scale Multi-touch Sensing Displays. *Proceedings of the 2006 International Conference on New Interfaces for Musical Expression*, Paris France: Ircam, pp. 216-219.

Edlund, J. 2007. Music Morphing, InterAmus Music Systems. accessed on 4th of April, 2007. site: http://www.interamus.com

Flety, E. (2005) The WiSe Box: A Multi-performer Wireless Sensor Interface using WiFi and OSC. *Proceedings of the 2005 International Conference on New Interfaces for Musical Expression*. Vancouver, BC, Canada, pp. 266-267.
http://mtg.upf.edu/reactable/?related
Jorda, S., Kaltenbrunner, M., Geiger, G., and Bencina, R. (2005) “The Reactable” *Proceedings of* of *the International Computer Music Conference (ICMC2005)*, Barcelona, Spain
Kaltenbrunner, M. and Bencina, R. (2007) “reacTIVision: A Computer-Vision Framework for Table-Based Tangible Interaction.” *Proceedings of the 1st international conference on Tangible and embedded interaction*. Baton Rouge, Louisiana: ACM Press, pp. 69-74.
Mathews, M. V. and L.Rosler 1969. "Graphical Language for the Scores of Computer-generated Sounds". Music by Computers, H. V. Foerster and J. W. Beauchamp. New York, John Wiley and Sons, Inc.: 84-114.
Newton-Dunn, H., Nakano, H., Gibson, J. *Block Jam: A Tangible Interface for Interactive Music.* Journal of New Music Research 2003, Vol. 32, No 4, pp. 393-393.
Oppenheim, D. 1997. "Interactive system for compositional morphing of music in realtime". USA, IBM. Patent no: 533636
Oppenheim, D. 2006. Pers. Comm. MMorph music demos, R. Wooller.
Oppenheim, D. 2006. Personal Correspondence with R. Wooller. MMorph music demos.
Oppenheim, D. V. 1995. "Demonstrating MMorph: A System for Morphing Music in Real-time". ICMC95 International Computer Music Conference. Banff Canada, ICMA: 479-480.
Patten, J., Recht, B., and Ishii, H. 2002. Audiopad: a tag-based interface for musical performance. In, E. Brazil, Ed. *Proceedings of the 2002 Conference on New interfaces For Musical Expression* (Dublin, Ireland, May 24 - 26, 2002). Singapore: New Interfaces For Musical Expression, pp. 1-6.
Polansky, L. 1987. "Distance Music I-VI for any number of programmer/performers and live, programmable computer music systems". Perspectives of New Music. 25: 537-544.
Polansky, L. 1991. "51 Melodies", Artifact Recordings/Frog Peak Music.
Polansky, L. 1992. "More on Morphological Mutations:Recent Techniques and Developments". ICMC. San Jose: 57-60.
Polansky, L. 1996. "Morphological Metrics". Journal of New Music Research. 25: 289-368.
Polansky, L. 1996b. "Bedhaya Guthrie/Bedhaya Sadra for Voices, Kemanak, Melody Instruments, and Accompanimental Javanese Gamelan". Perspectives of New Music. 34: 28-55.
Polansky, L. and McKinney, M. 1991. "Morphological Mutation Functions: Applications to Motivic Transformations and to a New Class of Cross-Synthesis Techniques". ICMC. Montreal.
Wooller, R. 2005. “Morph Music Demos,” Queensland University of Technology. accessed on 4th of November, 2005. site: http://www.lemu.org/download.html#l2sw
Wooller, R. 2006. "Review of compositional morphing: works, techniques, applications and possibilities". *Australasian Computer Music Conference* (ACMC). Adelaide: ACMA, pp. 172-177
Wooller, R. and A. R. Brown. 2005. “Investigating Morphing Algorithms for Generative Music.” In, Innocent, T. (Ed.) *Third Iteration*, Melbourne: CEMA, pp. 189-198.

Steven Campbell
School of Creative Arts
James Cook University
Townsville, 4811
Australia
steven.campbell@jcu.edu.au

Sensience: Electroacoustic Collaboration in a Reactive, Non-Tactile Ultrasonic Sensing Environment

Abstract

Sensience *is the title of a collaborative work realised with a proprietary ultrasonic sensing system and live tenor saxophone. The collaboration allows the acoustic saxophonist to move freely within an ultrasonic performance space and to trigger a range of samples pre-recorded by the performer. The aesthetic intention, the musical form and technical design of the work are discussed, along with an overview of the work's performance. The work is briefly contextualized with regard to interactive computer music practice.*

BACKGROUND

The first part of the *Sensience* collaboration is an ultrasonic system entitled PLaY+SPaCE. The system uses up to eight ultrasonic sensors to detect positions of people or objects moving within a detection space of up to 100sqm, the technical attributes and design of the system detailed previously in Campbell 2003 and 2005. The system allows an acoustic performer to move within the space unencumbered by physically attached sensing devices and hence the system is non-tactile. The system has previously been used in a wide range of applications including dance, installations, disabilities workshops and multimedia performances.

Prior to this collaboration, a similar work was developed for Sydney-based jazz saxophonist Andrew Robertson, forming the groundwork for the technical approach to *Sensience*. In this work, a live improvisation by Robertson was recorded and numerous samples of various lengths were selected from the recording. The samples were then assigned in the PLaY+SPaCE software environment to specific points within the system's physical sensing space, allowing Robertson to move within the space to trigger and react to samples of his own earlier improvisation. Additionally, Robertson's movement through specific points within the space allowed the performer to add and control effects such as delays and granular techniques to the triggered samples.

The second contributor to the *Sensience* collaboration was Berlin-based saxophonist Ulrich Krieger, an accomplished performer with a high level of skill in extended saxophone techniques and considerable experience in a wide variety of musical styles (Krieger 2006). During an artist residency in Australia in 2006, the collaboration for *Sensience* was established, the saxophonist provided with details of the PLaY+SPaCE system and shown video footage of the former Robertson collaboration. Subsequent meetings were held to establish directions for the new collaboration.

AESTHETICS

An outcome of initial meetings was that a similar technical model to that utilised in the Robertson work would be followed, however the samples would be in line with Krieger's own interests in noise, ambience, and noise on the verge of silence (Krieger 2006). Kreiger's highly developed instrumental technique would enable the sampling of multiphonics and breath noise through the saxophone at very low volume levels, the mergence of pitch into and out of such noise, very low volume pitches, percussive and pitched key clicks and tongue slaps. Additionally, Krieger has an interest in generating timbres from his instrument that are not generally associated with the saxophone: electronic/synthetic sounding timbres that he is able to produce at low volume levels, a technique he labels 'acoustic electronics' (Krieger 2006).

The possibilities of combinations of samples of the above techniques, along with live/acoustic performance were explored, as well as the use of effects and DSP. A decision to present an ambient and quiet work was made, and DSP effects, though of interest to both parties, were to be limited to basic delays and reverb. The work would thus rely on Krieger's triggering of the samples based on his extended saxophone techniques, in combination with his improvisational abilities to acoustically react to those samples through layering and contrast.

INTERACTIVITY/REACTIVITY/FORM

Discussion on interactivity generally has raised numerous questions as to what constitutes a truly interactive system within new media arts. Manovich (2001) dismisses the term "interactivity" as being too broad, relevant to any Human Computer Interface (HCI). Rather, Manovich provides a range of concepts and sub-classifications, for example "variability", "branching-type interactivity", "closed interactivity" and "open interactivity". Within *branching-type* interactivity the user may select items (e.g. buttons or hyperlinks) that enable movement along particular branches or pathways of a website or multi-media presentation. In a *closed interactive* system elements along a branch are predefined and fixed. Conversely, in an *open interactive* system both the elements and structure of the overall work are "modified or generated…in response to the user's interaction with the program" (p.40).

In the area of computer music the discourse has been a focus in monographs eg. Rowe (1992) and Winkler (1998), and in numerous papers, eg. Bongers (1999). Bongers outlines interactivity between the performer and the HCI as complete when a full interactive loop occurs, wherein the system and the user both contribute to human and machine cognition in the realisation of a work. Reactivity (as opposed to interactivity) occurs when the user alone controls the output of the system, without "cognitive" returns or responses from the computer.

Winkler (1998) provides a further definition, a reactive system in his terms labeled as a "Conductor Model" wherein one entity (the conductor) directly controls musical output (the orchestra). A more fully interactive system is in Winkler's "Improvisation Model", equated to a jazz ensemble wherein the individual player's solos "alter and influence the surrounding accompaniment".

The PLAY+SPaCE system allows for various levels of interactivity, however from the outset, the *Sensience* collaboration did not seek to utilise any level of *open interactivity*. Rather, within the limited timeframe the work was created, such possibilities for interactivity were stripped down so that, in the terms of Manovich (2001), a closed interactive system was designed. As the performer has complete control of the output of the system, this design falls under Winkler's (1998) "Conductor Model", a system that is essentially, and in Bonger's (1999) terms, *reactive*.

Reactivity in *Sensience* is achieved through the performer triggering pre-determined samples within the space and reacting to these, for example playing live and sustained saxophone pitches and multiphonics over multiple sustained samples to create richly layered timbres and textures. The work is structured around the improvisational abilities of the performer, variable reactions to the system made possible by simple and constrained randomisation of triggered sample output, as described below. Whilst the system outputs are indeed finite, a combination of deterministic and constrained randomisation of sample triggering results in a range of possible musical environments that are less constrained by pure determinism.

The work is designed to have six individual and linear sections (scenes), each scene containing various sonic possibilities for the user to explore. Unlike even the most basic of reactive or interactive media, the user here has no navigation options to freely move between scenes, nor to step backward. Rather, there is a single navigation function assigned to a single trigger point in the ultrasonic sensing space, this point acting as a "Next" button to move to the following scene. Figure 1 illustrates this basic structure of the work.

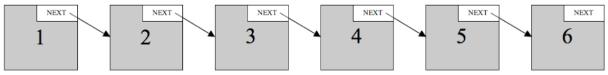


Figure 1 – Structure

The strict limitations on interactive possibilities within this structure are relevant to the compositional aesthetic selected for the work. Limited numbers of samples are available in each scene, requiring the performer to acoustically react to, and work with, repetitive sampled elements as per minimalist and ambient musical styles. The limited sonic possibilities resulting from this simple reactive system, and the elementary form utilised in the work, were nevertheless sufficient to provide the performer with materials with which he could engage to such an extent as to form a live performance of c. 20 minutes in duration.

SAMPLING

A single recording session was held, with Krieger providing c. 20 minutes of audio appropriate to the project. From the recording 23 samples were derived and grouped into six categorised sample sets. Six samples were single pitched notes played at very low volume and categorised as "Straights", five samples were single pitched notes dissipating into breath sounds and categorised as "To Noise", two samples were pure breath sounds and categorised as "Noise", five samples were key clicks including both percussive and pitched sounds and categorised "Clicks", and five samples were tongue slaps, categorised as "Slaps".

A disadvantage of the PLaY+SPaCE system is an audible click emitted from each sensor on sending an ultrasonic signal. In the system the sensors are timed to emit sequentially and hence there is a continuous audible clicking. Normally the intrusion of the clicks is masked by sounds/music triggered by users within the space. In *Sensience* however, the exploration of low volume levels resulted in the clicks being intrusive. A workaround was devised wherein the clicks themselves would become integral to the work, the sound of the clicks sampled and assigned to the space for triggering by the performer. The sampled clicks were categorised as "Sensors", the sample assigned to eight consecutive keys within the software sampler used for the work to result in variation via pitch shifting.

In each of the six scenes of the work trigger points in the PLaY+SPaCE sensing space are assigned to varying sample categories and combinations of the sample categories.

SENSING ENVIRONMENT

The PLaY+SPaCE hardware configuration for *Sensience* utilises four ultrasonic sensors, labeled S1 to S4 in the Figure 2 map. Each sensor is assigned to detect eight trigger points, a total of 32 trigger points available in the overall space. The PLaY+SPaCE sensing space can be up to 100sqm in size, but for this work a size of c. 50sqm is used, leaving room for an audience beyond the lower edge of the grid in Figure 2. The trigger point numbered 1 on Sensor 1 is assigned throughout the work to trigger changes of scene (the "Next" button), indicated by the red arrow.

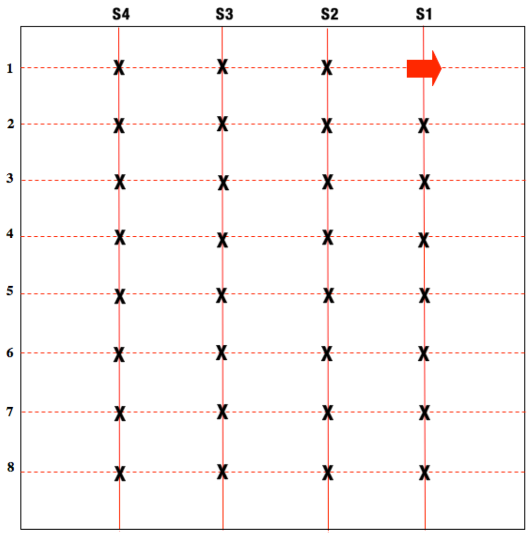


Figure 2 - Sensor Configuration

Figure 3 provides a visual perspective of the space, with Krieger in performance. The carpeted area in the image is the sensing space, the four sensors placed on stands at the edge of the carpet at the height of the performer's waist.

Figure 3 - Performance Space. Photo used with permission.

SOFTWARE ENVIRONMENT

PLaY+SPaCE utilises a proprietary software program for interfacing with the sensor system, as described in Campbell 2005. Each individual work created for the system then uses a patch designed in MAX/MSP to map incoming sensor data, to provide the user with navigation options and to output audio. Generally third party plug-ins are utilised, including samplers and effects.

Figure 4 shows the main MAX/MSP patch used for *Sensience*. Subpatch objects within the main patch contain programming relevant to each of the six scenes of the work. For third-party items, vst~ objects are used and here show the software sampler used (Steinberg's Halion2) and basic reverb and delay effects added prior to audio output.

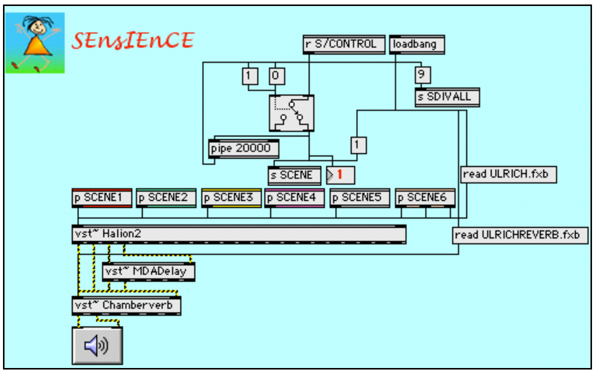


Figure 4 - Main MAX/MSP Patch

SCENE MAPPING

The general mapping of sample categories in each scene is shown in Figure 5, this figure forming the basis of the following overview of a performance of the work.

Figure 5 - Scene Sample Assignment

As an example, scene mappings for the opening scene are discussed in detail. Figure 6 shows the mapping for the scene, wherein the "Clicks" sample category is used, the five "Clicks" samples assigned to ten trigger points in the space. In addition to the key clicks, the "Sensors" sample category is also used in this scene. All 32 trigger points within the space trigger the "Sensors" sample, one of the eight pitch variations (resulting from pitch shifting) randomly selected at each triggering. This results in considerable layering of the sampled sensor clicks over the audible/acoustic clicks of the sensors themselves, the samples triggered as the performer moves throughout the entire sensing space.

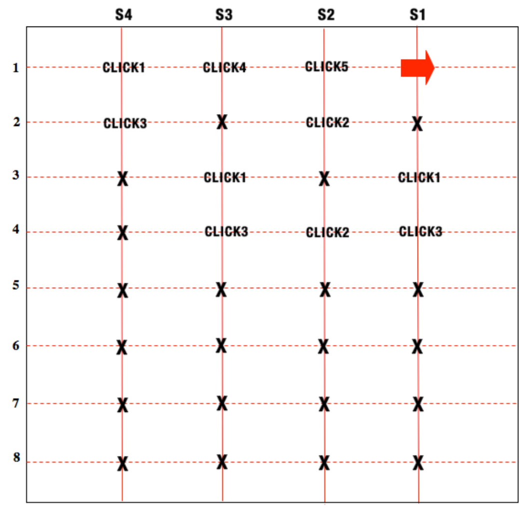


Figure 6 - Scene 1 Mapping

Figure 7 shows part of the MAX/MSP subpatch for Scene 1 and represents the mapping of the "Sensors" sample within the space. Input from all eight trigger points of the four sensors (S1 to S4) is allowed to generate a random number from 1 to 8. This number is then mapped to appropriate MIDI keyboard numbers as assigned in the software sampler.

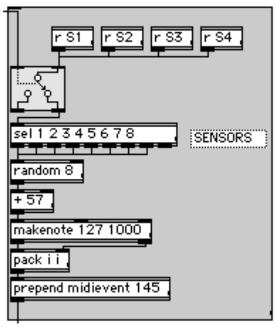


Figure 7 - Scene 1 MAX/MSP Subpatch

Inputs from Sensors 1 to 4 in the remainder of the subpatch (Figure 8) are mapped to trigger key click samples as seen in Figures 5 and 6. For example, trigger points 3 and 4 are selected from the Sensor 1 input to trigger MIDI note numbers 72 and 76 respectively, these in turn assigned to the "CLICK 1" and "CLICK 3" samples in the software sampler.

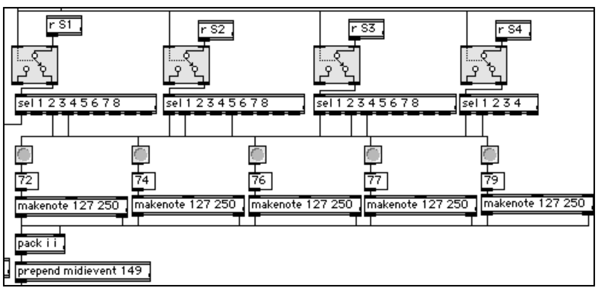


Figure 8 - Scene 1 MAX/MSP Subpatch

The sample assignments in the remaining five scenes each use a similar approach in triggering to that of Scene 1, i.e. the samples in each scene are either set to limited and specific points in the sensing space or are triggered randomly from a range of trigger points. The widespread triggering approach as used for the "Sensors" sample in Scene 1 is utilised again in Scenes 3 and 4. The 32 trigger points used here may be considered as a random triggering 'zone' that encapsulates the entire sensing space.

In Scene 5 of the work, smaller zones are employed, as shown in the mapping in Figure 9. Here two random triggering zones are used for the "Noise" category samples, triggered by movement through points 2 to 8 on Sensors 1 and 4. A subset of the "Clicks" category is used (six of the ten trigger points as used in Scene 1), these treated with a simple delay effect.

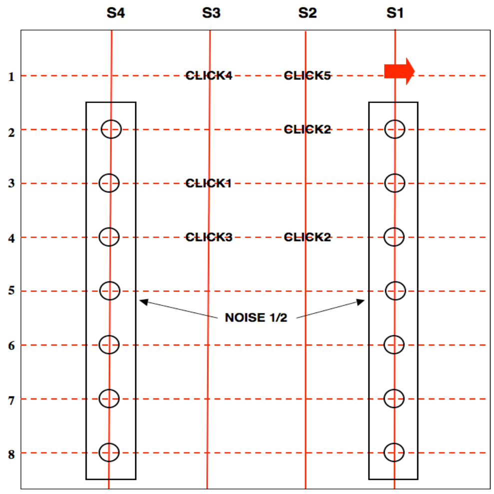


Figure 9 - Scene 5 Mapping

In Scenes 3 and 4, two different samples are triggered within a full-space zone of 32 trigger points. the mapping of two different samples to a trigger point resulting in a polyphony, or layering, of single-note "Straight" samples. Within the scene, four different dyads are randomly triggered, these shown in Figure 10.

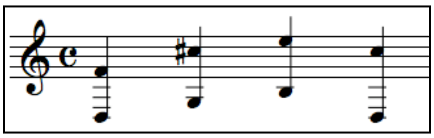



Figure 10 – Scene 3 Polyphony

Considerable variation occurs within Scene 3 as lengths of individual samples differ (ranging from 15 to 24 seconds), and a pitch cannot be retriggered until it has ended. Figure 11 illustrates with the first dyad, the upper pitch able to be retriggered prior to the end of the lower pitch. Within the scene this results in a continuous layering of the six samples in the "Straights" category.

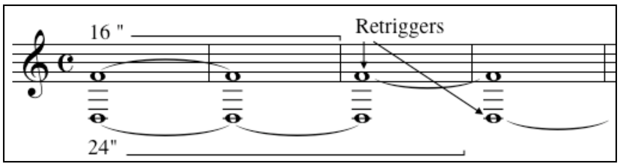


Figure 11 – Scene 3 Sample Layering

In Scene 4 a similar mapping is used, though here the samples are from the "To Noise" category resulting in continual layering of both pitch and breath noise.

Scene 6 of the work combines the "Straights" samples mapped as per Scene 3, i.e. over the entire sensing space. As shown in Figure 12, the "Sensors" sample is mapped as per the "Noise" sample on Sensor 1 in Scene 5, and the "Slaps" samples are assigned to a random triggering zone.

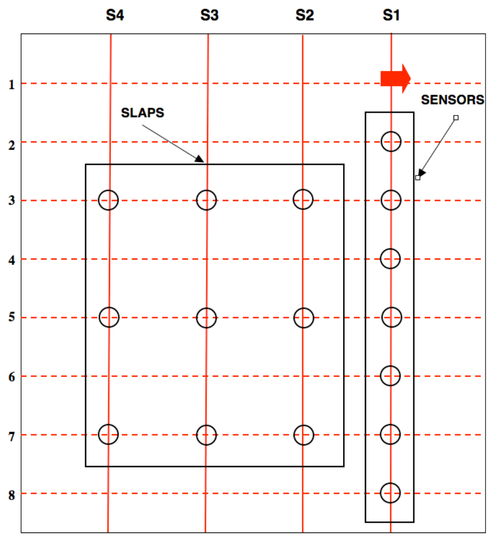


Figure 12 – Scene 6 Mapping

A final feature in the design of the work is an overlap of samples between sections. Rather than have an abrupt shift from one section to the next following a trigger to change scenes, a short period occurs in which trigger points in the space are allowed to trigger samples from both the old and the new scenes. In the opening two scenes this period is ten seconds, due to these scenes being quite sparse sonically, and in the remaining scenes the period is reduced to five seconds. Though these time periods are relatively short, they enable an effective transition between scenes by introducing new samples in conjunction with those previously heard.

PERFORMANCE

The preceding discussion has focused on the 'electro' portion of the *Sensience* electroacoustic collaboration, however this provides little detail regarding the musical outcomes achieved through Krieger's acoustic reactions to his triggering within the space. The following overview shows the manner in which the performer utilised the combination of his instrument and the triggered samples to achieve an ambient and quiet work of varying textures and contrasts.

The performance took place in October 2006, with Krieger having several rehearsals in which he adopted strategies for his performance before giving careful consideration to a slow choreography that he would utilise to further enhance the ambience of the work.

Figure 13 provides the structure, sample categories and timings of the six sections.

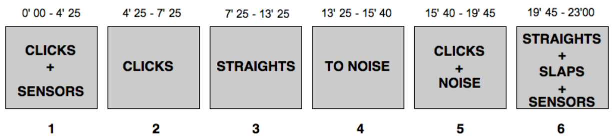


Figure 13 - Sections/Timings

The work begins with the integral clicking of the PLaY+SPaCE sensors, the system started prior to the performer entering the space. Krieger enters and immediately begins the layering of the "Sensors" sample, occasionally adding percussive acoustic sounds including fingernails on the bell and barely audible keyclicks, thus imitating the sampled and acoustic sensor sounds.

The space is slowly explored, the occasional triggering of the "Clicks" samples imitated and augmented acoustically. Over the course of the scene, Krieger moves slowly from a left-hand entry point to the right hand Sensor 1, exploring the "Clicks" samples in his progress. The trigger to Scene 2 is activated after 4'25" of this slow exploration.

Scene 2, with its focus on the "Clicks" samples allows Krieger to introduce similarly textured tongue-slaps, centred around Ab1. Following this introduction, the performer briefly focuses on repeatedly triggering the Click 4 sample on Sensor 3. This sample has a combination of pitched (Bb 1) sounds and a rapid percussive series, approximated in Figure 14.

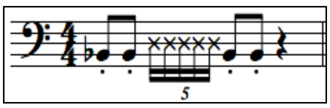


Figure 14 - Click 4 Sample

Through the repeated triggering of the sample the performer establishes a rhythmic episode, simultaneously developing a tonal interplay between the sample's Bb and his acoustic Ab.

A similarly brief interplay occurs on the Click 1 sample on Sensor 4 before the performer introduces a pitched and sustained D3, moving back to Sensor 3 to trigger "Clicks" samples and reacting to these with acoustic clicks. The same process is followed as the performer moves to Sensor 2, now introducing a sustained Db1. A brief focus on this pitch sees the performer move to Sensor 1 to end the scene.

Krieger treats Scene 3 as a central section in the work, extending the scene out over a six-minute period. Here the layered samples provide a continuous shifting drone over which the performer layers further pitches and multiphonics and melds his output into the sample layers utilising his "acoustic electronics" technique. Krieger's control and ability to play at very low dynamic levels allows him to imperceptibly add his acoustic layers to the sample layers, the listener often unable to distinguish between the acoustic or the electronic sound source. In this performance this audience experience is further enhanced by the sampled stereo output being spread throughout the room via a 7.1 surround speaker system.

Scene 4 structurally acts as a bridge, linking the "Straights" samples of Scene 3 to the "Noise" samples of Scene 5. Here Krieger simply continues the layering of the previous scene, augmenting his single pitch and multiphonic layerings with breath noise.

In Scene 5 Krieger structures his performance around the sample mapping, dividing the scene into five parts. The first and third parts are focused on noise with the performer's movements centred on the trigger points of Sensors 1 and 4 (see Figure 9). The second and fourth parts respectively utilise the "Clicks" samples assigned to Sensors 2 and 3. In the fifth part Krieger locates the area that is not assigned any trigger points (space between Sensors 1 and 4) where he is not detected by Sensors 2 or 3, and uses the ensuing silence (with the exception of the sensor clicks) to focus on slap-tonguing, this an introduction to the same timbre used in Scene 6.

With the "Clicks" samples of Scene 5, the PLaY+SPaCE system is set in a mode that allows trigger points to be locked into the system's timing. In this mode the performer may remain stationary at a trigger point and the system will recognise the performer's position at each ultrasonic pass, once every 512ms. This mode is utilised in the scene on Sensors 2 and 4, allowing continual retriggering of the "Clicks" samples assigned to these sensors. Krieger utilises the rhythmic repetition of the samples here to set up a chattering rhythmic background, over which he applies acoustic single note and multiphonic trills and tremolos, further key clicks and rapid *sotto voce* passages.

The final scene of the work is recapitulatory in nature, and following the increased rhythmic activity of Scene 5, the performer uses the slow moving and sustained layers from the "Straights" samples to ease the work to its conclusion. Acoustically, Krieger recapitulates on the numerous timbres he has exploited throughout the work, both contrasting and blending with the layered "Straights" and the "Sensor" samples. Interspersed with his blends and contrasts, the performer reacts to the triggered slaps. Relative to the delayed "Clicks" samples of Scene 5, the "Slaps" samples are also treated with delay, the performer often imitating the samples with his own acoustic slaps and imitating the delay through repetition and decrescendo.

In the final moments of the work Krieger moves to the sensor interface, located behind Sensor 1, and powering it down, stops the audible clicks of the system to conclude the work.

CONCLUSION

With the exception of basic demonstration pieces, *Sensience* represents one of the most elementary works yet designed for the PLaY+SPaCE system. The work's linear and rigid form, its focus on reactivity as opposed to open interactivity and its lack of any inherent sound generation or synthesis relegate the design of the work to this position. Yet all of these apparent deficiencies cannot be considered as shortcomings, each having been adopted (or rejected) in accord with the desired aesthetics and sonic outcomes required of the work.

As Manning (2004) points out, "evaluation of electroacoustic works should be based in the first instance on the perceived results and not in terms of the technical means by which they have been achieved". In this light, the level of reactivity or interactivity of a work becomes irrelevant: the computer music practitioner needs develop a technical repertoire suited to a range of music/performance settings, a repertoire that is utilised with a full knowledge of the capabilities and limitations of the computer music system used. Thence, selection of an interactively open or closed (reactive) design for a work may be driven by the aesthetic requirements of the work and its user.

As described above, aesthetic considerations, along with a limited timeframe for composition, were key factors in the development of the *Sensience* collaboration. Whilst a range of more complex levels of interaction may certainly have been explored, the choice of a simple reactive system provided multiple sonic environments suited to the performer's desire to systematically explore in real-time his own pre-recorded samples in conjunction with his live playing.

REFERENCES

Bongers, B. 2000. Physical Interfaces in the Electronic Arts. Interaction Theory and Interfacing Techniques for Real-time Performance. In Wanderly, M. & Battier M. (Eds.) *Trends in Gestural Control of Music.* Paris, IRCAM - Centre Pompidou.

Campbell, S. 2005. "PLaY+SPaCE: An Ultrasonic Gestural MIDI Controller." *Proceedings of the 2005 Australasian Computer Music Conference*: Australasian Computer Music Association: 43-49.

Campbell, S. 2003. "Bats, Max, Boids and Music." *Proceedings of the 2003 Apple University Consortium Conference*. Available from auc.uow.edu.au/conf/conf03/papers/AUC_DV2003_Campbell.pdf [Accessed 18th March, 2005].

Krieger, U. 2006, http://www.ulrich-krieger.de [Accessed 16/3/07].

Manning, P. 2004. *Electronic and Computer Music*. Oxford: Oxford University Press.

Manovich, L. 2001. *The Language of New Media.* Cambridge, Mass. MIT Press.

Winkler, T. 1998. *Composing Interactive Music, Techniques and Ideas Using Max.* Massachusetts. The MIT Press.

Poppi Narelle-Faye Doser
Electronic Music Unit,
Elder Conservatorium of Music
University of Adelaide
SOUTH AUSTRALIA 5005
poppi.doser@adelaide.edu.au

Synaesthesia and the Affect of Programming in SuperCollider

: : Forming a relationship with code in order to realise the body as an immersive environment : :

Abstract

> *[Humans] and things exchange properties and replace one another; this is what gives technological projects their full savour.*
> (Latour, 2004:34)

Whilst essentially unique, all systems of language share the ability to express connections, relationship between things constituting the driving force behind the evolution of ideas and as a result of this, existence in the physical. At present it is the language of computational technology that is most often utilised by sound and multimedia artists as a means to creatively realise concrete representations of subterranean sensory experiences. An example of this is the collaborative digital music adventures that take place in real time and remotely utilising the Open SoundControl (OSC) Protocol (Wright, 2006). Such an activity exists potentially to redefine current attitudes surrounding the inter-connection and interdependency of life.

The ability to creatively map scientific data and subjective detail in previously unimagined ways, to output high-level ideas based in a combination of the 'emotional and perceptible structural level'(Lockwood, 2003) has contributed to a deepening of thought regarding the relationship between the human being and their environment. Furthermore, it has given rise to a context wherein the brain and body (computational systems in their own right) and hence pleasure and affective learning no longer reside abstracted from one another. This is where my interest lies.

I am a student with a limited amount of preconceptions surrounding the abilities of programming given that I have not long been acquainted with digital music technology and compositional practices involving the use of computer code. It is most likely due to the premature nature of these relationships that I intuit computational technology's langue (its system of communication) as the most relevant means for explorative and interpretive analysis of the natural unfolding of daily habitual practices, i.e. activity and its movement, an event and its structure, within modern society. Henceforth, influenced by generative research methods, I have chosen to "immerse" myself into the art of object-oriented audio synthesis programming in order that I may realise and in turn musically "codify" an artistic expression of that which currently exists on the periphery of the neurologists' quantifiable logic: the synaesthetic experience during [and the affect of] the practice of coding.

James McCartney, author of the seminal object-oriented audio-synthesis programming language SuperCollider, regards 'synthesis' as a complex problem demanding a language that is highly expressive in order to deal with such complexity. (McCartney, 1999) I in turn perceive the text-based lexicon of McCartney's SuperCollider and the affect of programming in the object-oriented style as offering me a viable and effective means through which to undertake my research and explore the underlying hypothesis that initially inspired my university Honours project: the body as an immersive environment.

Within a many-layered, multi-dimensional present resides the antidote to the infancy of [our] immediate thoughts; the potential for a collective and/or individual experience that transcends traditional representations of thought, the lexical, the narrative of 'expectation denied and expectation fulfilled.'(Lockwood, 2004) Delving into this place
- - via code, I anticipate the physical act of programming as a way to quantify the senses and, during my future research I expect it to be an original line of enquiry into the feminine experience. At present I aim to realise the creation of a synaesthetic work indicative of the affect of programming.

References

Lockwood, Annea. "Annea Lockwood." *What is Sound Art?* 2003. http://emfinstitute.emf.org/articles/aldrich03/lockwood.html (30 August 2006).

McCartney, James. 2000. "SuperCollider 2.0: Why SuperCollider 2.0?" *01 Why SuperCollider*. 2006. http://sound.jp/dspss2004/materials (3 April 2007).

Nordschow, Randy, ed. "The Glass Concert: Annea Lockwood Beside the Hudson River." *New Music Box*. January 2004. http://www.newmusicbox.org/article.nmbx?id=2364 (20 September 2006).

Wajcman, Judy. 2004. "Technoscience Reconfigured." In *TechnoFeminism*. Cambridge, UK: Polity Press.

‘Open SoundControl is a new protocol for communication among computers, sound synthesisers, and other multimedia devices that are optimised for modern networking technology.’ In, Matthew Wright and Adrian Freed. 2004. “Open SoundControl: A New Protocol for Communicating with Sound Synthesisers.” *OpenSound Control Home Page*. http://www.cnmat.berkeley.edu/ICMC97/papers-html/OpenSoundControl.html (27 April 2006).

P[a]ra[pra]xis: Poetry in Motion

Josh Dubrau, Mark Havryliv
Faculty of Arts & Social Sciences
University of New South Wales,
Sonic Arts Research Research Network
University of Wollongong
Australia
jdubrau@uow.edu.au
mhavryliv@gmail.com

Abstract

P[a]ra[pra]xis is an ongoing collaborative project incorporating a two-piece software package which explores human relations to language and linguistics through dynamic text and sound production. Incorporating the linguistic theories of Ferdinand de Saussure, and the psychoanalytical work of Sigmund Freud on parapraxes, or "Freudian slips", our software utilises player response to automatically-generated changes in a narrative of their own writing to create music.

"P[a]ra[pra]xis Collection Editor" is a Java application which manages the relationships between words and a series of possible substitutions based on manipulation of linguistic slips and "free association". Another Java application, "Realtime P[a]ra[pra]xis", handles the real-time implementation of the above lingual substitutions according to a configurable rule-set.

A rule-set is a set of Boolean values characterised by the relationships between the original word and the substituted word; harmonically and rhythmically, music is controlled by bit-patterns derived from an individual rule, changing as rules are progressively applied during a performance.

References

Barthes, R. 1957. *Mythologies*. Seuil, Paris
Burt, W. 2006 *Poems of Rewi Alley*
Op De Coul, M. 2007 (http://www.xs4all.nl/~huygensf/scala/)
Freud, S. 1973 *The Psychopathology of Everyday Life*. The Hogarth Press and the Institute of Psycho-Analysis, London
Jenkins, G. 2006. "A Radical Midrash: a Reading With[in]" *Rhizome Magazine, Issue 1, 2006* (University of Wollongong, Australia)
De Saussure, F. 1983. *Course in General Linguistics* Duckworth, London

Polyphonic Listening:
Real-time accompaniment of polyphonic audio.

Toby Gifford & Andrew R. Brown
Queensland University of Technology
2 George Street
Brisbane, 4000
Australia
t.gifford@student.qut.edu.au
a.brown@qut.edu.au

Abstract

This paper outlines a technique for generative musical accompaniment of a polyphonic audio stream. The process involves the real-time extraction of salient harmonic features and the generation of relevant musical accompaniment. We outline a new system for polyphonic pitch tracking of an audio signal which draws upon and extends previous pitch tracking techniques. We demonstrate how this machine listening system can be used as the basis for a generative music improvisation system with the potential to jam with a live ensemble without prior training.

Introduction

The musical collaboration between people and machines has a long history (Levenson 1994). Technologies to support this collaboration have often built upon signal processing solutions designed from an engineering perspective, most recently those dedicated to music information retrieval. The approach adopted in this paper draws on similar techniques but approaches them with a sensibility derived from music perception studies concerned with salient, or musically significant, features. Our broad objective is to enable automated accompaniment of a polyphonic audio input, and the process outlined here is a step toward that goal focusing on the identification of the harmonic context of the audio source.

Among the important steps in achieving our goal of an autonomous computational improviser is the extraction of musically salient features from polyphonic audio input, and generation of an appropriately complementary musical response. Our previous work (Gifford & Brown 2006) focused on the latter problem and this paper reports on our work toward a solution to the former.

Broadly speaking musical features can be grouped into those that relate to pitch, rhythm, and timbre. In polyphonic material beat tracking can become complex but in the case of metrical music the task is not overly confused by additional instruments. However, the identification of specific note boundaries is quite involved because, while informed by metrical concerns, it benefits from the coordinated analysis of pitch and timbre, and their continuity or change over time. It is to this task of note extraction from polyphonic audio information that we focus in this paper. We will first describe our approach to an integrated analysis of these features with a particular focus on polyphonic pitch tracking. Following this we will provide an overview of how we demonstrate the effectiveness of the analysis through a simple real-time accompaniment process. Outlining this process is important not only to demonstrate the veracity of our claims that the salient pitch features have adequately been identified, but also because the design of the pitch tracking necessarily takes into account the generative musical application to which it is being put when deciding amongst the necessary trade-offs and making choices with regard to degrees of filtering or data abstraction.

Background

The perception of pitch alone is multifaceted and intrinsically tied up with timbral spectra, dynamic and phase (Pierce 1999). Whilst computational analysis of monophonic material has been possible for some time the ability to distinguish between multiple sources, sometimes called the 'cocktail party effect' remains elusive because of physical effects resulting from interference and overlay of spectral features between parts, including resolving displacements. This is confused further when attempting stream segmentation (extraction of parts) because of ambiguities in the apparent motion of parts (Shepard 1999).

However, the salient features of musical material as required for generative musical accompaniment need not be as accurately described as those required for transcription or music information retrieval. The pitch features identified as most important in a range of studies include pitch class set, pitch range or contour, and changes in these over time. These lead to structural features including tonality, harmonic change, textual density, grouping and proximity, that can be used for generative purposes (Sloboda 1988, Temperly 2001, Cope & Hofstadter 2001).

Having identified some of the salient pitch features from the raw physical description in the audio signal, we are in a position to use these to generate a simple accompaniment based on the rules of western music theory and those outlined by musicologists such as Temperly (2001).

Generative accompaniment

Generative accompaniment systems improvise a musical part to compliment incoming data. In our system this data is the data generated by a polyphonic pitch tracking algorithm. Generative accompaniment is different to some prepared-accompaniment systems such as *SmartMusic* and *In The Chair* that track an acoustic performance and play a synchronised prepared accompaniment.

GenJam, by Al Biles (1994, 2002) is a prominent generative accompaniment system. It uses a pitch to MIDI converter to track a live performer and a combination of prepared, recombinatorial and genetic algorithm processes to create an accompaniment. Our system differs from *GenJam* in that it tracks a polyphonic audio stream and does not rely on a database for the construction of accompaniment material. Our system is similar to many computer-assisted compositions systems except that it operates in real-time and is therefore better described as an improvisation system, or an interactive music system along the lines of Robert Rowe's *Cypher* (1993) or those created by musicians including Todd Winkler (1998) or Roger Dean (2003). By comparison our current generative accompaniment systems is quite rudimentary, but it does serve to demonstrate effective musical extrapolation from data produced by our polyphonic audio tracking process; a task not attempted by the previously mentioned systems.

Beliefs and prediction

A key metaphor that drives our approach at both the signal processing and generative accompaniment stages is the maintenance of expectations or beliefs about future events. In the absence of evidence to the contrary these beliefs are upheld or states are maintained. This approach provides some inertia to the system giving it greater stability and can also improve efficiency of algorithms when searches are focused or abandoned early as a result. This approach is informed by, but not rigorously consistent with, theories in cognitive linguistics (Jackendoff 1992, 2002) and computational neuroscience (Hawkins & Blakeslee 2004) .

Pitch Tracking

We wish to be able to extract pitch and timing information sufficient to recognise the key and rhythm from an audio stream in real-time.

There is a large body of work in monophonic pitch tracking, particularly from the speech analysis community. Much less attention has been devoted to polyphonic pitch tracking, which is generally regarded as a difficult and unsolved problem (Hainsworth 2004; Collins 2006 pp. 60-61).

Desired Features

(a) Tonality analysis – we want to be able to identify the current 'key'

(b) Timing information from pitch changes – we wish to be able to identify the times at which a new note is played, even in the case where the note changes are the result of a legato movement – for which no attack is present – so as to enable tempo, metre and rhythm to be inferred.

(c) Cope with low frequencies – much of the important harmonic information is provided by bass instruments and many pitch tracking algorithms require large window sizes to cope adequately with the lowest notes on the bass guitar.

(d) Bias towards less False Positives – with a focus on salient features it was more important to have correct information, even at the expense of less complete data.

Onset Detection

A large variety of methods for the detection of note onsets have been discussed in the literature, a survey may be found in (Collins 2006). Typically onset detection algorithms are analysed in terms of their performance on different types of sounds. Bello (2004) considers the broad classes of Pitched Non-Percussive (PNP), Pitched Percussive (PP), Non-Pitched Percussive (NPP) and Complex Mixtures (CMIX) sounds. Broadly speaking energy based techniques are reasonably successful for Percussive sounds but fail spectacularly for Pitched Non-Percussive. Bello reports on a technique based on spectral phase information that is relatively successful in the PNP class.

The system that we present in this paper concentrates on reporting onsets for the PNP class of sounds. An example of the types of onsets that we are considering would be a wind instrument playing a legato passage. Intuitively, in order to detect note boundaries in this class of sounds, one must have an accurate estimate of the frequency of the sound through time, so that an onset may be reported when the frequency changes substantially.

The reason that we concentrate on this class is that, in accord with the intuition expressed above, accurate onset timing information in the PNP class of sounds is an added bonus from accurate frequency estimation, which is the second goal of this system. Indeed Bello's system essentially estimates the instantaneous frequencies contained in the audio signal (although he is not explicitly interested in these estimates). The system that we present here operates on a similar principal to that of Bello, producing accurate frequency information and timing information for the PNP class of sounds.

We envisage that an improvising computational agent would run a number of onset detection algorithms in parallel, of which this could be one, so as to deal most appropriately with the variety of onsets that may be encountered in a complex audio stream.

Time Resolution

If the timing information is to be useful for beat and metre induction tasks, it is our view that a high degree of time resolution for onset events is desirable. As we are proposing to utilise frequency information to report on note boundaries, a practical consideration arises regarding the size of the analysis window that we use, namely that for low frequencies the analysis window may be too short to obtain an accurate estimate of the frequency. In signal processing there is generally a trade-off between time resolution and frequency resolution (Puckette 1998).

We are aiming at a time resolution of 1024 samples at a sample rate of 44100 Hz, corresponding to approximately 23*ms*. This was chosen as a reasonable goal as it is as below the generally accepted minimum perceptible time (Goebl 2001) and is a commonly available buffer size for audio hardware.

We wish to be able to determine the frequency content of pitched musical material to within a semitone across the range of frequencies commonly present in (western popular) music, say 60*Hz* to 16000*Hz*. A fundamental problem that we face is that a window size of 1024 samples represents around 1.5 cycles of a 60*Hz* frequency component. The low number of cycles makes it difficult to accurately estimate the frequency of low frequency components with the desired time resolution.

Harmonic Context Description

Because our goal is to isolate salient pitch material that will enable us to infer the current harmonic activity such as key or chord progression, we are not so concerned about accurately detecting all notes being produced. Our plan is as follows: Identify the spectral peaks in the analysis window – these are viewed as the *constituent frequencies*. Associate harmonically related frequencies together as being part of one *pitch*. This yields a number of distinct *pitches* present in the analysis window – this is the information that will be fed to the generative improvisation engine. The nature of overtone series for physically produced sounds suggests that an appropriate method for associating harmonically related frequencies is to start with the lowest frequency and associate with it any frequencies which are an integer multiple. To this end it is important to have an accurate estimate of the fundamental frequency as errors in this estimate become compounded with each multiple.

Harmonic Product Spectrum

A number of pitch tracking techniques attack this problem by using information from harmonics of the fundamental to refine the pitch estimate. The harmonics, being at a higher frequency, have more cycles within the analysis window and so can be more accurately estimated. Then, assuming that the sound source is harmonic, the frequency of the fundamental can be estimated as a common divisor of the frequencies of the harmonics. One such technique that operates along these lines is the Harmonic Product Spectrum (Noll 1969), which can be quite effective for pitch tracking of monophonic harmonic sources. Another technique which utilises information from a larger portion of the spectrum to infer the frequency of the fundamental is Goto's predominant-F0 estimation (2004).

The Chicken or the Egg?

However, approaches that use information from the whole spectrum to estimate the frequency of the fundamental introduce an undesirable circularity to the analysis – whilst trying to ascertain whether or not a given frequency is a harmonic of a given fundamental frequency, the process relies upon refining the estimate of the fundamental based on the assumption that it *is* harmonically related to the given frequency. For monophonic audio sources with a known spectrum this does not pose a great difficulty, but in the case of polyphonic audio composed of an unknown number of varied sources this can be problematic.

Independent Fundamental Estimation

In this paper we describe a novel frequency estimation technique that avoids such circularity by obtaining an accurate estimate of the fundamental frequency independently of the rest of the spectrum. Having done this, the estimate of the fundamental frequency may be further refined by whole-spectrum techniques such as above, but applying our new technique first facilitates the accurate assessment of which of the constituent frequencies are indeed harmonically related to each other *before* using the harmonic relations to refine the frequency estimates.

Frequency Estimation Techniques.

There are a range of techniques commonly used to estimate the fundamental frequency of a signal. We will explore the most prominent of them and highlight ways in which they may not meet our requirements.

Time Domain:

1. Zero Crossings. A simple method for frequency estimation is to count the number of times that the signal crosses zero in a given period. For low frequency signals, where the number of cycles is a small multiple of the analysis period, the discrete nature of the zero crossing events leads to large variations in the frequency estimate depending on the phase of the signal.

2. Autocorrelation. The autocorrelation of a signal is the correlation of the signal to a time-lag of itself. A periodic signal should be most highly correlated with itself at a time-lag equal to its period. Autocorrelation frequency estimation techniques utilise this by calculating the autocorrelation as a function of time-lag (called the autocorrelation function) and searching for a maxima of this function. In practice this technique has a large variance for low frequency signals and so is unsuitable for our purposes.

Fourier Domain:

Fourier techniques generally take as a starting point the spectrum of the windowed signal, usually obtained via a Fast Fourier Transform (FFT) algorithm. The most straightforward application of the spectrum is to plot its power versus bin number and identify the bins at which the power has a 'significant peak'. The centre frequency of the bin is then identified as a *constituent frequency* in the signal. A problematic issue with this approach is that the resolution of the frequency estimates is determined by the size of the bins. In a standard FFT the bins are equally sized at SF / N were SF is the sampling frequency and N is the number of samples in the window. In our case the sampling frequency is 44100*Hz* and the number of samples is 1024, yielding a bin size of 43*Hz*. This means that frequency estimates obtained in this manner are accurate to within 43*Hz*, which is ample for high frequency components but insufficient for low frequency components.

A number of techniques exist for increasing the resolution of the FFT:

1. Zero Padding. Before taking the FFT the signal is padded with zeros – i.e. the signal vector is concatenated with a vector consisting of zeros producing a longer vector upon which the FFT is performed. The zeroes do not effect the frequency composition of the signal, however the resolution of the frequency estimates are increased due to the larger number of bins (Smith 2003).

2. Parabolic Interpolation. The frequency of a component that is identified by a peak in the spectrum is interpolated by fitting a quadratic function to the value of the power spectrum at the peak bin and the bins either side of the peak bin. The refined frequency estimate is given by the location of the maxima of the fitted curve (Smith 2003). Other more elaborate interpolation schemes exist also (Milivojevic 2006).

3. Constant Q transform. Rather than measuring the power at equally spaced frequency intervals as the FFT does, the Constant Q transform measures power at exponentially spaced frequency intervals (Brown 1992). This means that the frequency resolution in percentage terms is equal across the spectrum.

4. Chirp Z transform. This transform utilises equally spaced frequency intervals but concentrated in a frequency band of interest, rather than across the whole spectrum (Rabiner 1972).

All of the methods above do a good job of interpolating the frequency spectrum and result in a higher resolution frequency estimate. However, analysis of a low frequency sine wave using these methods reveals that resolution is not the only issue. Indeed all of these methods yield a similar result, and have a high variance when estimating a single sinusoidal component of known frequency. In particular, for a 60*Hz* sinusoid over a 1024 sample analysis window, the variance in the frequency estimate for all of these techniques exceeds a semitone; consequently these techniques are insufficient for our purposes. The fundamental issue is that for a low frequency signal the peak of the power spectrum is simply not an accurate estimator of the frequency.

Phase Based Techniques

The Fourier spectrum contains more information than the just the power spectrum; additionally it contains phase information. The efficacy of exploiting this phase information for the purposes of frequency estimation was famously described by Flanagan and Golden (1966) in their description of the Phase Vocoder. The essential idea is that the instantaneous frequencies of the signal are equal to the time derivatives of the phases. The values of the instantaneous frequencies plotted against the Fourier bin number is known as the Instantaneous Frequency Distribution (IFD) (Charpentier 1986).

A number of techniques have been proposed for the calculation of the IFD. The original Phase Vocoder simply advances the signal by one sample, computes another FFT, and approximates the phase derivatives by the difference in phase divided by the sample period. This is however computationally expensive as it involves computing an FFT for every sample in the signal.

Charpentier (1986) proposed utilising symmetry properties of the Fourier Transform to approximate the FFT of a window advanced by one sample from the FFT of the original window. This way one needs only calculate one FFT per analysis window, and obtains very similar results to the Phase Vocoder technique.

Puckette (1998) expands on Charpentier's work and compares the accuracy of this technique to a similar technique but where the signal is advanced by H (the hop size) samples instead of one. The accuracy of the estimates of the IFD increase with larger hop size, but there is a trade-off. The author identifies two issues:

(a) The frequency estimate from this technique is essentially using information from a window of size N + H where N is the size of the analysis window, so the time resolution of this technique decreases with larger hop size.

(b) The estimate becomes increasingly vulnerable to mistakes in the phase-unwrapping. The measured change in phase corresponds to some integer number of cycles plus a measured fraction of a cycle. The number of whole cycles that the phase has gone through is unknown, except by virtue of some other independent estimation of the frequency.

The technique that we present in this paper, which extends this method, addresses these two issues.

To use the IFD to determine the *constituent frequencies* of a signal, one computes the spectrum, picks the peaks of the power spectrum, and then reports the values of the IFD at the bins corresponding to the peaks of the power spectrum. A further refinement to this class of techniques has been proposed by Kahawara (1999). The idea is that when plotting instantaneous frequency against frequency, where there is a true frequency component in the signal the instantaneous frequency should equal the frequency. In other words the mapping frequency -> instantaneous frequency should have a fixed point at every true frequency in the signal. Much as the discrete Fourier spectrum can be interpolated, the phase spectrum can be interpolated. Then searching for fixed points on the interpolated IFD yields refined frequency estimates.

Masataka Goto (2004) has utilised fixed point analysis of the IFD in conjuction with a Bayesian belief framework in his predominant-F0 estimation. A number of other authors have utilised Bayesian techniques in the context of Polyphonic transcription (Cemgil 2004; Hainsworth 2004)

Gifford-Brown Technique.

The technique that we propose is essentially a hybrid of the above techniques, tailored to suit our specific needs. It is a two stage estimation technique similar to those of Charpentier (1986) and Puckette (1998), which addresses the shortcomings of these techniques by utilising a combination of IFD fixed point analysis, belief propagation, and MQ analysis (McAuley & Quatieri 1986). The algorithm is as follows:

We analyse the input signal in non-overlapping windows of 1024 samples. Internally, we hold a number of **beliefs**, consisting of a frequency, amplitude and phase value for a component that we currently believe to be sounding. The beliefs are stored in a fixed number of 'tracks' using the terminology of MQ – analysis. Each window we iterate the following procedure:

1. Perform an FFT on the (rectangularly) windowed signal.
2. Pick the significant peaks of the power spectrum
3. The frequencies of the spectral peaks are estimated using Charpentier's technique.
4. These estimates are refined using Fixed Point Analysis.
5. The refined estimates are pair-wise matched with our current **beliefs.**
6. Any peak that is more than a quarter-tone away from the closest belief is considered to be a new component. If the refined estimate of the peak is 'sensible', then we put this estimate into a free track (if any are available) and consider it to be a *tentative* new belief. We record the frequency estimate as the frequency of this belief, the power of the spectrum at this peak as the amplitude of this belief, and we calculate the phase of this belief by performing a single component Fourier transform of the window centred on the estimated frequency. We also record a *tentative* track-birth, and if there was a component previously in this track we record a track-death.
7. Any peak that is within a quarter-tone of the closest belief is considered to be a candidate for continuation of the belief. We can now use the stored phase information for this belief to perform an enhanced frequency estimate using the N-hop technique of Puckette (1998). We can do this extremely efficiently by utilising the frequency value of our belief: we calculate just a single Fourier spectral coefficient centred at our believed frequency. We also calculate the number of whole cycles that this component should have gone through in one window based on our believed frequency. Adding to the whole number of cycles the partial phase advance (measured as the difference between this phase of the single calculated coefficient, and the stored phase of this belief) yields a total phase advance through the window which can be converted to a refined frequency estimate.
8. If the refined estimate from step 7. is 'sensible' (in this case by sensible we mean within a quarter-tone of our belief) then we consider this to be a *definite* continuation of the believed frequency. If the belief was previously *tentative* then we retrospectively mark it as *definite*, and retrospectively mark the track-birth as *definite*. We then use an average of the old belief and the new estimate of the frequency as the frequency for our belief. We use the phase value already calculated for the phase and the power of the peak as the amplitude.
9. Finally we go through our old beliefs, and for any belief that has not been continued we record a track-death, and fill the track with a new belief as needed.

The use of a two-stage estimation, firstly utilising the 1-Hop technique of Charpentier, and secondly the N-Hop technique of Puckette, circumvents the issues raised regarding the N-Hop technique. Firstly by encapsulating the information about the previous analysis window in a set of believed frequencies and phases, we increase the time resolution of the N-Hop technique to a single analysis window, at least in so far as detecting when a steady component terminates. Whilst it will still take two windows to get an accurate frequency estimate from the N-Hop technique once a new component starts, we can in the meantime use the 1-Hop estimate of the frequency during this first window. So for the first window of a new pitch our estimate is not as good as for any following windows – however we view this as acceptable especially since during the attack phase of a new note the pitch may be unstable in any case. Furthermore, the timing information regarding the beginning of a new component has a time resolution of one window, which is the desired result. Once the frequency stabilises the frequency estimate for the first window may be retrospectively adjusted if desired.

Secondly by maintaining a belief of the sounding frequency (or in the case of a new component by having a first-stage estimate of the frequency) we can alleviate the phase-unwrapping errors inherent in the N-Hop technique.

Filtering

The output of the above algorithm is a fixed number of tracks containing frequency beliefs (along with corresponding amplitude and phase), and a series of timing events for track-births and trackdeaths. The next step in the algorithm is to filter these frequencies, and associate them into notes. Here we are not trying to reproduce the notes in the audio stream as they may have been physically produced, but rather perform a sensible data-reduction that groups frequencies into salient units.

In each analysis window we aggregate frequencies that are harmonically related. Recall that this was one motivation behind the lengths we have gone to above to obtain an accurate frequency estimation for the lowest frequencies in the signal. So roughly speaking, we loop through all of our beliefs for a given window: starting with the lowest frequency and associating with it any frequencies that are within a semitone of being a harmonic of that frequency. Such a collection we will call a pitch, and associate with it the frequency of the lowest component. Then we take the remaining beliefs and iterate this procedure, yielding a number of believed pitches each window. In actuality we modify this procedure somewhat, since as described the algorithm would aggregate the fundamentals of a bass line playing a C2 with a melodic line simultaneously playing a C4 for example. Whilst it is not our intention to segregate the audio stream into distinct parts, we make some concession towards this by requiring that harmonics have no more than half the energy of the fundamental to which they are aggregated.

Having aggregated frequencies vertically into pitches, we then aggregate pitches horizontally (in time) into notes. Notes are formed by examination of the track birth and death timing information. A note is considered to start/finish at the time of the birth/death of the track containing the fundamental. The note structure consists of a start time, an end time, a pitch envelope and an amplitude envelope. The information about the notes is known in real-time with a latency of two analysis windows, however the accuracy of the resolution of the timing information for the start/end of notes is just one analysis window.

Once the notes have been formed we perform a real-time merge of any two notes whose frequencies are within a quarter-tone and where one note finishes in the same or previous window as which the second note starts. This step eliminates a great deal of erroneous discontinuities caused by noise. On the other hand it means that this algorithm will not pick up rhythmic information from repeated pitched notes. However, since the notes are of the same pitch, for a human listener to gather rhythmic information from them they must be marked in some other way, for example with attacks or articulatory information, and these markings may be picked up by an independent onset detection algorithm such as an energy based detection system, that operates in parallel with this algorithm. Note that this information is available with a latency of two analysis windows.

Finally we eliminate from consideration notes that are too short, in our case we have chosen three analysis windows as the minimum length for a note to be validated. Consequently the algorithm as a whole has a latency of three analysis windows, though the timing information (for feeding into a beat-tracking algorithm for example) has a resolution of one analysis window.

Extracting salient features

The results from the audio analysis are in the form of 'notes' with a frequency, amplitude, and duration. When listening to this output reproduced it is clear to hear that the transcriptions are not accurate. Often pitch slides and note ghosting are misconstrued as additional notes and jumps to related frequencies a fifth or octave displaced are not uncommon. These can be filtered by ignoring short notes or by increasing the belief threshold to provide additional stability. For the purposes of auto accompaniment, these 'notes' are considered like incoming MIDI messages, their frequencies quantized to an equally tempered scale.

The stream is generated in real-time and therefore timing and metrical placement information can be significant, however, in the trials reported here we focused on harmonic organisation. Integrating beat tracking and other temporal analysis will be a topic for further research.

Improvising accompaniment

Our intention for this accompaniment was twofold: firstly to demonstrate that the polyphonic listening process was sufficiently accurate and complete that a reasonable harmonic accompaniment could be generated from the data. Secondly, to explore methods of tracking changes in the reported salient harmonic features over time.

Data gathering

Accumulating the pitch material as it arrives is a trivial process, however there are some important aspects about the data stream that need to accounted for when generating an accompaniment. These include:

- The data is only loosely ordered
- Stream segmentation is absent in the data.
- The data may not be clean.

The loose ordering refers to the fact that while the assessment of 'note' candidates occur within a window because of the varying length of overlapping polyphonic notes the precise order of events cannot be guaranteed. Therefore the ordering of events can only be assumed as approximate.

The data is not segmented with relation to musical parts in any way, additionally at times significant overtones may appear as discrete notes. These significant overtones are usually harmonically related and consequently their salient value to the harmonic nature of the material remains relevant. In this work the input data is accumulated as a sliding windowed cluster.

Given that the pitch analysis process is not entirely accurate it is inevitable, despite filtering efforts at the analysis stage, that the data will not be clean and may contain erroneous pitches. The analysis process is sufficient that these are uncommon and so the improvisational process utilises a probability-based system that amplifies the statistical significance to further minimize these errors.

Data analysis

Our accompaniment process begins by accumulating the pitches derived from the audio analysis into a windowed histogram of pitch classes. This makes clear in a simple way the harmonic tendencies of the audio input. As incoming pitches are added their pitch class weighting is increased and the whole set is normalized reducing all other weights and acting as a form of memory loss for the system. A simple accompaniment can be generated by probabilistic selection from this histogram with reasonable effectiveness, but lacks musical structure and direction. The normalised histogram will track changes in harmonic content more or less quickly depending upon the weighting assigned to incoming notes, the higher the weighting the more rapidly the histogram reflects changes in the input data. The histogram provides a weighted pitch class set as a basis for speculation about current key and/or harmonic progression that can inform the generative accompaniment processes.

The pitch range of the incoming data is simultaneously tracked by finding the maximum and minimum pitch from the last few data. The buffer size of previous pitches can be adjusted depending upon the input material. We have found that for music where higher and lower instruments play continually, that a small buffer size (less than 10 notes) is sufficient. Knowing the range of instruments can, in simple cases, allow the accompaniment to occupy complementary pitch ranges, and otherwise acts as a basis for more sophisticated arrangement decisions based on implied instrumental functions, textural density, and larger scale structural changes.

Generating accompaniment

In demonstrating our polyphonic tracking system we have used two part audio files containing and bass and melodic parts. The generated material provides harmonic material in the form of chordal and arpeggiated accompaniment.

Pitch material for these parts is derived from a combination of direct selection from the pitch histogram within the dynamically tracked range, and from chordal and key estimations statistically derived from the pitch histogram. Compositional considerations include being more conservative about pitch choice on prominent beats and allowing passing notes to be less constrained by either the pitch class set or harmonic estimations. Textural density and accompaniment range is varied according to the distance between bass and melody lines.

At present rhythmic, and dynamic aspects of the accompaniment are unrelated to the tracking data, but we plan to extend our research to include estimation of salient aspects of these musical features in the future providing appropriate data that can influence these aspects of the generated material.

Practical results

The resulting improvised accompaniment appeared to the authors' ears to demonstrate a reasonable degree of awareness of the harmonic context of the audio stream, though a number of 'outside' notes were also generated. An inspection of the notes produced by the listening algorithm suggested that some of the notes perceived were erroneous, particularly in the lower frequencies. The erroneous pitches tended to be short lived however, so a secondary filtering process in the improvisation algorithm yielded a much more 'inside' accompaniment.

A parameter of the improvisation algorithm is the rate at which new pitches are introduced into the pitch-set histogram, and the rate at which old pitches are forgotten. There is a balance to be struck between responsiveness and stability. If new pitches are allowed to dominate the pitch-set too quickly the system is readily thrown from the tonal centre by an outside note. This problem is exacerbated by having noisy input data. On the other hand if old pitches are maintained for too long then the system does not readily adapt to harmonic changes. In practice we found that by altering this parameter we could obtain an improvisation that sounded acceptably in key, and responded to harmonic changes within a bar.

Conclusion

We have outlined a process for the real-time extraction of salient features from a polyphonic audio stream, and discussed the application of this system to a generative accompaniment engine. This process extends and improves upon a variety of pitch tracking techniques, and has been demonstrated to extract the harmonic context from a two part musical excerpt.

The process also yields timing information, which may be supplied to a beat induction algorithm. Our intention is extend the generative accompaniment engine to also produce generative rhythmic output.

We speculate that this process may be refined by virtue of a feedback loop with a beat-induction system. By giving greater weight to candidate notes that begin on a beat erroneous pitches may be filtered more effectively.

References

Bello, J. P., Duxbury, C., Davies, M., & Sandler, M. (2004). On the Use of Phase and Energy for Musical Onset Detection in the Complex Domain. *IEEE Signal Processing Letters, 11*(6), 553 - 556.

Biles, J. A. (1994). GenJam: A genetic algorithm for generating jazz solos. *International Computer Music Conference*, San Francisco.

Biles, J. A. (2002). GenJam: Evolution of a jazz improvisor. In P. J. Bentley & D. W. Corne (Eds.), *Creative Evolutionary Systems* (pp. 165-188). San Francisco: Morgan Kaufmann.

Brown, J., & Puckette, M. (1992). An efficient algorithm for the calculatio of a constant Q transform. *Journal of the Acoustical Society of America, 92*, 1394-1402.

Brown, J., & Puckette, M. (1993). A high-resolution fundamental frequency determination based on phase changes of the Fourier Transform. *Journal of the Acoustical Society of America, 94*(2), 662-667.

Cemgil, A. T. (2004). *Bayesian Music Transcription.* Radboud University of Nijmegen.

Charpentier, F. J. (1986). Pitch Detection Using the Short-Term Phase Spectrum. *International Conference on Acoustics, Speech and Signal Processing*, New York.

Collins, N. (2006). *Towards Autonomous Agents for Live Computer Music: Realtime Machine Listening and Interactive Music Systems.* Cambridge, Cambridge.

Cope, D. a. H., D. R. (2001). *Virtual Music: Computer Synthesis of Musical Style*. Cambridge, Mass: MIT Press.

Dean, R. (2003). *Hyperimprovisation: Computer-Interactive Sound Improvisation*: Middleton: A-R Editions.

Flanagan, J. L., & Golden, R. M. (1966). Phase Vocoder. *Bell Systems Tech Journal, 45*, 1493-1509.

Gifford, T., & Brown, A. (2006). The Ambidrum: Automated Rhythmic Improvisation. *Australasian Computer Music Conference*, Adelaide.

Goebl, W., & Parncutt, R. (2001). Perception of onset asynchronies: Acoustic piano versus synthesized complex versus pure tones. *Meeting of the Society for Music Perception and Cognition*, Kingston, Canada

Goto, M. (2004). A real-time music-scene-description system: predominant-F0 estimation for detecting melody and bass lines in real-world audio signals. *Speech Communications, 43*, 311-329.

Hainsworth, S. (2004). *Techniques for the Automated Analysis of Musical Audio.* University of Cambridge, Cambridge.

Hawkins, J., & Blakeslee, S. (2004). *On Intelligence*. New York: Times Books.

In the Chair. (2006) In the Chair Pty. Ltd. Adelaide. http://www.inthechair.com/

Jackendoff, R. (1992). *Languages of the Mind: essays on mental representation*. Cambridge, MASS: MIT Press.

Jackendoff, R. (2002). *Foundations of Language: brain, meaning, grammar, evolution*. Oxford: Oxford University Press.

Kahawara, H., Katayose, H., de Cheveigne, A., & Patterson, R. D. (1999). Fixed point analysis of frequency to instantaneous frequency mapping for accurate estimation of F0 and periodicity. *European Conference on Speech Communication and Technology.*

Levenson, T. (1994). *Measure for Measure: a Musical History of Science*. New York: Touchstone.

McAuley, R. J., & Quatieri, T. F. (1986). Speech Analysis/Synthesis Based on a Sinusoidal Representation. *IEEE Transactions on Acoustics, Speech and Signal Processing, 34*(4), 774-754.

Milivojevic, Z., Mirkovic, M., & Milivojevic, S. (2006). An Estimate of Fundamental Frequency Using PCC Interpolation - Comparative Analysis. *Information Technology and Control, 35*(2), 131-136.

Noll, M. (1969). Pitch Determination of Human Speech by the Harmonic Product Spectrum, the Harmonic Sum Spectrum, and a Maximum Likelyhood Estimate. *Symposium on Computer Processing in Communications, Polytechnic Institute of Brooklyn.*

Pierce, J. R. (1999). Introduction to Pitch Perception. In P. R. Cook (Ed.), *Music, Cognition, and Computerized Sound: An Introduction to Psychoacoustics* (pp. 57-70). Cambridge, MASS: MIT Press.

Puckette, M., & Brown, J. (1998). Accuracy of Frequency Estimates Using the Phase Vocoder. *IEEE Transactions on Speech and Audio Processing, 6*(2), 166-176.

Rabiner, L. R., Schafer, R. W., & Rader, C. M. (1972). The Chirp z-Transform. In L. R. Rabiner & C. M. Rader (Eds.), *Digital Signal Processing* (pp. 322-328): IEEE Press.

Rowe, R. (1993). *Interactive Music Systems: Machine Listening and Composing*. Cambridge, MASS: MIT Press.

Scheirer, E. (1999). Towards Music Understanding Without Separation: Segmenting Music with Correlogram Comodulation. *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, New Paliz, New York.

Shepard, R. (1999). Stream Segregation and Ambiguity in Audition. In P. R. Cook (Ed.), *Music, Cognition, and Computerized Sound: An Introduction to Psychoacoustics* (pp. 117-127). Cambridge, MASS: MIT Press.

Sloboda, J. A. (1988). *Generative Processes in Music: The Psychology of Performance, Improvisation and Composition*. Oxford: Clarendon Press.

Smart Music. (2005) MakeMusic Inc. Eden Prairie, MN. http://www.smartmusic.com/

Smith, J. O. (2003). *Mathematics of the Discrete Fourier Transform (DFT), with Music and Audio Applications*: W3K Publishing.

Temperley, D. (2001). *The Cognition of Basical Musical Structures*. Cambridge, MASS: MIT Press.

Winkler, T. (1998). *Composing Interactive Music*. Cambrige, MASS: MIT Press.

Luke Harrald
Elder Conservatorium of Music
Level 9, Schulz Building
The University of Adelaide, 5005
AUSTRALIA
luke.harrald@adelaide.edu.au

Collaborative Music Making with Live Algorithms

Abstract

This paper discusses developments of ENSEMBLE, an interactive improvisation environment based on the Iterated Prisoner's Dilemma.

The main emphasis of this paper is on the interactive version of ENSEMBLE, and its development for the work 'fr@gm3nT' [fragment], a collaboration between the author and saxophonist Derek Pascoe. Some of the lessons learned from non real-time, generative versions of ENSEMBLE are also discussed, along with the implications of the approach for algorithmic composition and live interactive computer music performance.

Introduction

Although many impressive generative composition systems have been developed in recent times, relatively few allow for interactivity with musicians in live performance. The emerging field of 'live algorithms' addresses this situation through combining non-linear generative composition techniques with live electronics and a sense of 'strong interactivity' exemplified through the practices of 'free' improvisation (Blackwell and Young, 2005). Previous examples of live algorithms include George Lewis' 'Voyager' system (Lewis, 2000), Al Bile's 'GenJam' (Biles, 2002) and Tim Blackwell's 'Swarm Music' (Blackwell and Bentley, 2002).

From the outset of its development ENSEMBLE has aimed to model the social dynamics of music performance, drawing inspiration from the use of performance indeterminacy pioneered in the works of composers such as Christian Wolff, Cornelius Cardew and John Zorn (Harrald, 2005). Using a modified version of the Iterated Prisoner's Dilemma (Axelrod, 1984) ENSEMBLE aims to model a group of improvising performers whose actions are constrained by sets of simple rules.

George Lewis suggests that the emergence of structure in improvised music occurs in much the same way as structure emerges in our every day lives. We interact with our environment, navigating through time, place and situation, both creating and discovering form (Lewis, 2004). As the Iterated Prisoner's Dilemma has proven its ability to model a diverse range of social situations without the need to address the details (Axelrod, 1997), it is not such a leap to suggest that it may prove useful in modelling improvised music.

Background

The Prisoner's Dilemma and the Arts.

While the Prisoner's Dilemma has captured the imagination of philosophers through the rather bleak outlook it presents about basic human nature, it has been somewhat of a rarity in the arts. In a one-off situation, it suggests that whenever there is uncertainty on what your opponent is about to do, then the only rational option is non-cooperation. This idea led several of the US government's advisors, including members of the RAND Corporation and British pacifist Bertrand Russell, to advocate a pre-emptive nuclear strike against the Soviet Union in the late 1940's (Poundstone, 1992).

Aside from the some of the classic films dealing with the madness of the nuclear arms race, for example Stanley Kubrick's Dr Strangelove (1964) (whose main character would appear to be a caricatured blend of several of RAND's key figures), art dealing more directly with the Prisoner's Dilemma is fairly scarce. A notable musical exception was put forward in the early 90's by Nick Didkovsky. Based on Douglass Hofstadter's 'Luring Lottery' (Hofstadter, 1983) Didkovsky's work explores resource sharing through allowing performers to compete for control of musical events via a network of Commodore Amiga 1000s . (Didkovsky, 1992).

Another exception is Bohemian Productions' innovative theatre work 'A Prisoner's Dilemma'. Featuring a number of interactive scenes, the actors portrayed various Prisoner's Dilemma scenarios under the control of the audience. The outcome of each game altered the plot so that each performance was different (Bohemian Productions, 2007).

ENSEMBLE

Through the ENSEMBLE project, several software applications have been developed in Cycling 74's MaxMSP environment (Cycling 74, 2007) based around a common Iterated Prisoner's Dilemma (IPD) engine. The IPD engine is an agent model consisting of eight agents. The agents interact with one another through a series of rounds, according to strategies that are predetermined prior to the first round. There are only two choices: cooperate or defect. The agents' environment is made up of their interactions and they communicate solely through the sequence of their own behaviour. The Iterated Prisoner's Dilemma is implemented as a competitive tournament whereby the agents interact in randomly selected pairs and are rewarded points depending on the outcomes of their interactions. The key musical concept here is that through cooperation, the agents reinforce previously introduced musical materials, while defection results in a random selection of new materials. Through the competitive nature of the model, the other members of the group may reinforce each individual agent's musical initiatives or they may be ignored, mirroring the musical dilemma facing real life improvisers. A comprehensive overview of ENSEMBLE can be found in Harrald, 2005.

ENSEMBLE is a modular system and the addition of various modules has allowed for the simple development of applications to suit a range of performance situations and compositions. Its development has been conducted in a number of stages as various modules have been created, beginning solely as a demonstration of the Iterated Prisoner's Dilemma Game, morphing into a non real-time generative composition system (IPD Score Generator, figure 2), followed by real-time generative systems used in installation works, and finally the interactive system through the incorporation of a fuzzy logic pitch tracker allowing a performer to interact live with the agent ensemble.

It has been suggested that a way to fast track progress in the research of live algorithms would be to link existing units dealing with analysis, synthesis and generative algorithms, each of which is individually the subject of much current research (Blackwell, 2007). Certainly the modular development process of ENSEMBLE strongly supports this possibility.

Lessons from non-real time

IPD Score Generator

IPD Score Generator is a generative composition application that, as its name suggests, generates MIDI files that can be easily imported into scoring applications. As working in the MIDI domain and in non real-time was much less CPU intensive, several important developments were made with the score generator, allowing the agent's in the IPD engine to have control over a wide range of musical parameters.

This application was an important step in the development of ENSEMBLE, in that, unlike the real-time and interactive applications based around the IPD engine, *IPD Score Generator* allowed the generation of musical materials that are pinned down and can be more easily analysed and assessed. As the IPD model can be quite volatile, often the musical surface produced is very transitory in nature. Several works have been generated using the application, with instrumentation ranging from solo piano, to full orchestra.

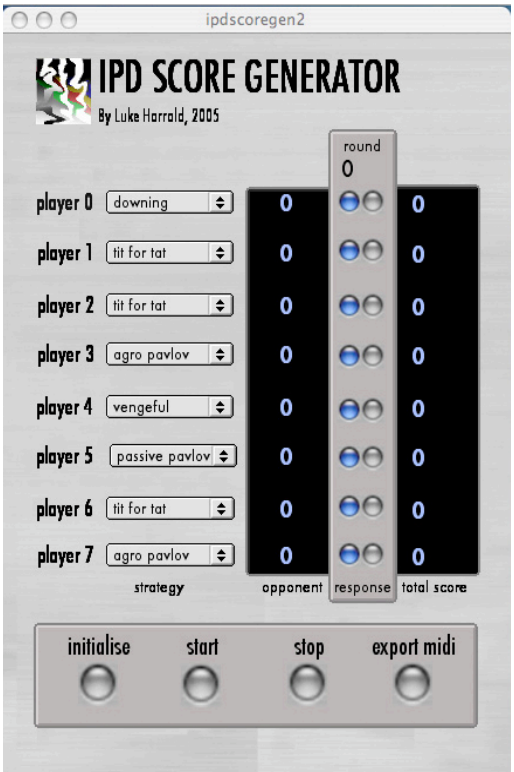


figure 1: IPD Score Generator (September, 2005) screenshot. On the left, drop down menus allow for user control of the agent's strategies (default is a random selection of strategies) while the panel on the right visualises the agent's interactions.

One of the most striking aspects of the musical works produced by IPD Score Generator is their immediate sense of phrase and directed motion, without any reference to traditional composition methods of thematic development or functional harmony. This sense of phrase occurs solely through the agent's interactions as they reinforce and abandon different musical materials. The score generator works at the meso- (or note) level of the work, and has no hierarchical operators to control other aspects of the form.

The Implications of Strategy

One of the key questions raised by IPD Score Generator is what combinations of strategies for interaction between agents can lead to the emergence of interesting musical structures? This is different in many ways to typical IPD research in that the main emphasis here is not on finding a 'best' strategy, but rather looking at the roles that different strategies can play in changing the dynamic musical state of the system. The current system implements seven rule sets (figure 2) that have been tweaked to suit the short memory of the agents; the agents only remember the preceding round.

Generally, in game theory, IPD strategies can be described as nice, or nasty, responsive or unresponsive. Nice strategies will never be the first to defect, while nasty strategies see cooperation as an opportunity to exploit their opponents in the next round. Similarly, responsive strategies will react to the actions of their opponent, while unresponsive strategies ignore their opponent's responses. A strategy's level of forgiveness refers to how quickly they will return to cooperation on their opponent's resumption of cooperation (Poundstone, 1992).

In a musical context, bearing in mind that in ENSEMBLE, cooperation reinforces existing musical materials, while defection results in a random selection of new materials, the rule sets can also be defined according to the roles they tend to play within the virtual ensemble. For example, nice rules can be thought of as 'passive', maintaining a state of equilibrium within the initial musical state, while nasty rules can be thought of as 'agitators' which push the musical state forward by introducing new materials. 'Responsive' rules tend to reinforce this push towards new states instigated by the agitators, while 'forgiving' rules can be considered 'dampeners', coaxing the responsive rules back towards cooperation and the reinforcement of the newly introduced musical materials.

Combinations of these musical behaviours are what give the music produced by ENSEMBLE its sense of phrase, and musical flow.

Name	Strategy	Attributes	Musical Role in the ensemble
TIT FOR TAT	Cooperate in the first round, mimic the opponent's response from the previous round in all subsequent rounds.	Nice Responsive Forgiving- once others cooperate	**Passive**- will maintain a stable cooperative musical surface. Once change begins to occur, it's responsiveness helps introduce new musical materials and consolidate these through cooperating with opponents once others begin to cooperate.
RANDOM	a random or irrational selection.	Unresponsive	Can play an important role as either an **Agitator** for the musical surface (in concert with nice, responsive rules) or a **Dampener** to pull the ensemble back towards more cooperative situations if there is a high level of defection.
VENGEFUL	co-operate until defected against , then defect for the next 5 rounds regardless of opponent's response.	Responsive Nice relatively Unforgiving	**Passive** until defected against, then major **Agitator**- spreads defection quickly through the ensemble, as the 5 rounds of retaliation effect multiple opponents. Can create a very chaotic musical surface punctuated by short cooperative periods.
COPYCAT	do whatever the player with the highest score did in the previous round.	Unresponsive tends to follow 1 round behind others	**Stabiliser**- tends to lag a round behind the other players. As such will allow sounds periods of mass cooperation and defection to linger a little longer than they otherwise would have.
PAVLOV	(traditional) win stay the same, lose change. counts a cooperate/ cooperate response as a loss (ie. opportunity to exploit, so will defect in the next round).	Responsive Nasty Forgiveness in order to exploit	Works well as an **Agitator**, cooperation is an opportunity to exploit! Also takes on the **Dampening** role will pull the musical surface back towards more cooperative states (before exploiting again).
PAVLOV	(passive) as above, but counts the cooperate cooperate result as a win and will not defect until defected against. Unlike TIT FOR TAT, a defect/ defect result will cause cooperation.	Responsive Nice, Forgiving- will hold out the olive branch	**Passive + Dampening-** tends towards cooperative states, and promotes them through holding out the olive branch in periods of high defection. Will retaliate on the first defection though, and continue to do so until it gets a defect/ defect result.
DOWNING	do what the most players in the previous round did.	Unresponsive to specific players, but responds to the overall state	**Stabiliser**- holds the musical state towards either mass cooperation or defection depending on the majority. The downing strategy can play an important role in pushing players towards one state or the other. Use with care- too many Downing's cause the system to get stuck!

Figure 2. Table of currently implemented rule sets and their roles as musical agitators, pushing the musical surface towards new states, consolidators, pulling the musical surface towards previous states and stabilisers, holding the ensemble's behaviour towards either cooperation or defection en masse.

fr@gm3nT [fragment]

Incorporating the performer's actions

In order to incorporate the performer's actions into interactive versions of ENSEMBLE, one of the agents was removed from the system, and the performer effectively 'wired into' the agent's place in the tournament. As such, the performer collaborates with (or competes against) an ensemble of seven agents.

The interactive versions of ENSEMBLE draw inspiration from the notions of performance indeterminacy pioneered by the New York School (see Cage, 1961), in that they aim to push the performer outside their comfort zone to create new musical experiences rather than drawing on the performer's previous knowledge. Although the improvising ensemble is a mode of music making that the system draws inspiration from, it was never intended as a replacement for a human ensemble, but rather as something that offers an experience similar in some ways, but with an emphasis on opening up new possibilities for collaborative music making. To this end, initial systems focused heavily on game play, and in particular 'gaming'. A Graphic User Interface was developed, effectively creating a 'video game '(figure 3), where the performer took on the agents in a competitive musical IPD tournament.

While this graphically oriented system generated some interesting musical results, and certainly is quite a departure from the normal experience of improvising, once work began with Derek Pascoe on the piece *fr@gm3nT* it became clear that when working with experienced improvisers, the interface was not really necessary, or particularly desirable. Pascoe found that he could beat the agent's scores much more often if he abandoned trying to play the game through the graphic interface, and concentrated on listening to the musical output of the system, responding to what he heard. This also made for a far more cohesive musical result. As improvisers generally have highly trained listening skills, and are very attuned to the actions of the other improvisers in a group situation, this was hardly surprising, but certainly an interesting observation.

One of the challenges of the interactive versions of ENSEMBLE has been the incorporation of pitch recognition which is based around Tristan Jehan's 'pitch~' object (Jehan, 2001). The biggest hurdle

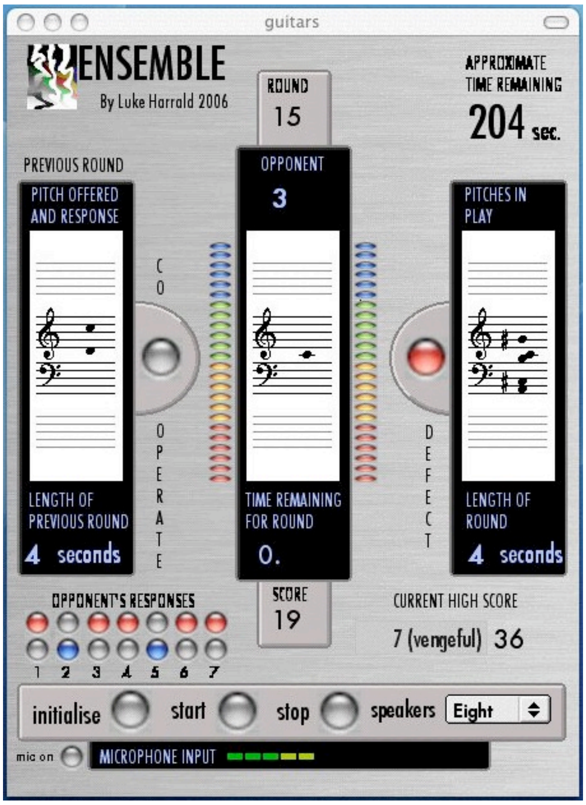


Figure 3: initial version of the interactive ENSEMBLE 'video game'. The performer takes on the ensemble of agents in a five minute battle to see who can get the highest score; generates a new composition at the same time. (May, 2006).

was getting the raw data from the pitch tracker into a useful format that could be understood by the agents. This was compounded a little for *fr@gm3nT* as the work incorporates saxophone multi-phonics and extended techniques. The solution lay in the development of several fuzzy logic operators that allow different aspects of the sound to be tracked, and categorised so that rather than precisely tracking individual sonic events, snapshots of the performer's actions are taken, and then a higher level 'type' of musical material is approximated. Although it may seem counter intuitive, this actually led to a far greater accuracy in recognising different sonic events, and also simplified the system considerably.

Fr@gm3nT makes use of a palette of 28 saxophone samples that are split up into eight types of sound, in some ways reminiscent of Cage's Gamut technique (Pritchett, 1993). In each round of the game, the performer's actions are compared to an offer from an agent. This determines whether the agents see the performer as cooperating with or defecting against their musical initiatives, and shapes their responses in subsequent rounds (figure 5).

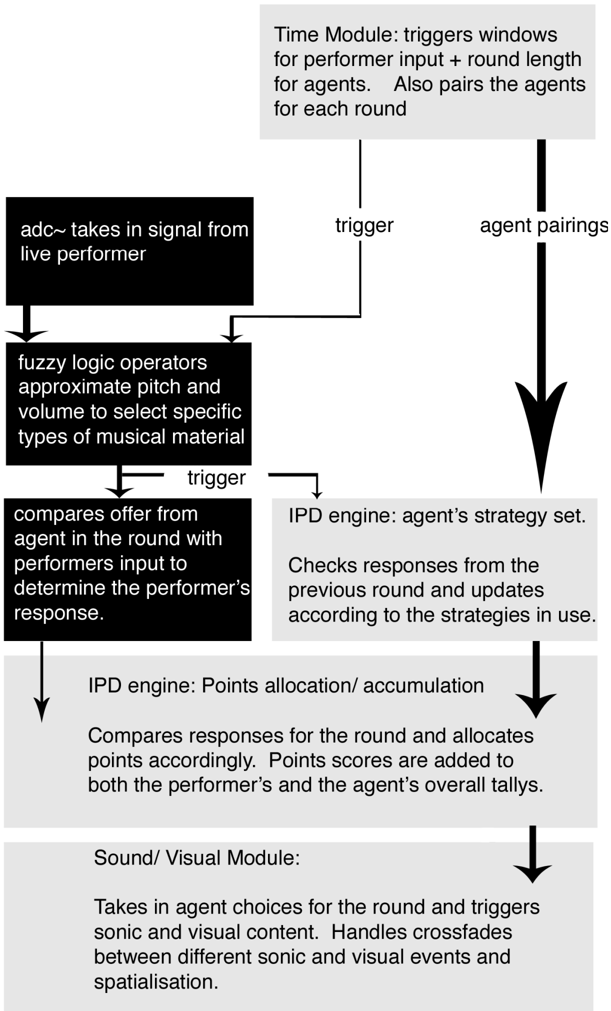


figure 4: modular structure of the ENSEMBLE application as used in fr@gm3nT.

Useful strategies for live performance

The current strategies used in ENSEMBLE are deterministic in that once the game begins the agents are unable to modify their behaviour to attempt to improve their performance in the IPD tournament. This was considered important to allow for a level of clarity in the interactions between the live performer and the agents, allowing the performer over time to become familiar with how different agents behave and how different combinations of strategy affect the global dynamics of the ensemble. Working with Pascoe in this area has been very interesting, as he generally uses a strategic approach to his own improvisations, and has a keen sense of strategy when interacting with fellow (human) improvisers.

The form apparent in the works generated by IPD Score Generator hinges on combinations of strategies that promote global behaviours oscillating between high levels of cooperation and high levels of defection. During periods of high levels of cooperation, the agent's musical choices converge towards a single sonic event, while high levels of defection lead to randomness and a rapid introduction of new musical materials. The speed that the system oscillates between these two states affects the volatility of the music produced.

While these formal considerations are certainly very useful in the live situation, as the performer is effectively also an agent within the system it is fruitful to consider the ensemble's behaviour with regard to the input from the live performer. In this sense, high levels of cooperation within the virtual ensemble are likely to lead to the reinforcement of the performer's musical initiatives (providing the performer has achieved a relatively high score in the tournament themselves), while defection will move the ensemble's output in other musical directions.

Modifying the system to allow the agents to change their strategies, or indeed evolve new strategies, to attempt to improve their performance in the IPD tournament as the game is played have been considered, although Didkovsky's work with the 'Luring Lottery' suggests that modifying the agent's behaviour to improve their overall performance in the tournament may not have a desirable musical effect.

Didkovsky found through his system that in a group of human performers, several behaviours developed in rehearsal prior to the performers having a full understanding of his system that did not occur in the actual performance of the work. These included 'arms escalation', as performers defected against one another to try and gain control of the musical events; 'de-escalation', as they realised that this method did not actually enable anyone make significant changes to the musical events; 'peace', as performers basically gave each other an equal chance of control; and 'destabilisation', as once peace was established, performers would try to take control through defecting from time to time. Unfortunately, the performance was rather tame as once the performers understood how the system worked they tended to work cooperatively as a group without any of the 'social storminess' of the rehearsal. (Divkovsky, 1992). As ENSEMBLE relies on these kinds of retaliatory behaviours to generate form, changing the agent's strategies to improve their performance during the tournament would appear likely to lead to the production homogenous global behaviours, and static music.

The aesthetics of interaction

Much of the music generated through ENSEMBLE draws on the tension and release paradigm, albeit without any reference to traditional compositional devices such as thematic development or functional harmony. As the user is free to choose the agent's strategies, the system is certainly not limited to this approach. Behaviours could be chosen that lead to completely cooperative states, creating either a drone or silence, or equally, unresponsive rules could be chosen to create random works.

Conclusions and future work

ENSEMBLE has proven its value as an algorithmic composition system through various works composed with the IPD Score Generator application. These works demonstrate a sense of musical phrase and form solely through the agent's interactions as they reinforce and abandon different musical materials. This process echoes the way in which form emerges in 'free improvisation' where there are no pre-determined structures, suggesting that the Iterated Prisoner's Dilemma offers a vehicle for modelling the interactions between improvisers without the complexity of attempting to incorporate a performer's musical training or cultural background.

These ideas have transferred across to the interactive system, with the agents able to reinforce or work against the input of a live performer. While the system used in fr@gm3nT is far less advanced than the system used in IPD Score Generator, it is hoped that future interactive systems will lead to the incorporation of a wider range of musical parameters to create more complex modes of interaction.

Acknowledgements

Thanks go to Derek Pascoe for his collaboration on fr@gm3nT, and the valuable contribution he has made to the interactive version of ENSEMBLE through his feedback and advice.

Thanks also to Tim Blackwell and other members of the Live Algorithms for Music research network (www.livealgorithms.org) whose encouragement and interesting conversations have helped clarify several of my ideas.

Special thanks go to Stephen Whittington for his ongoing support and mentorship throughout this project.

References

Axelrod, R. 1984. "The Evolution of Cooperation". Basic Books, New York.

Axelrod, R. 1997. "The Complexity of Co-operation: Agent-Based Models of Competition and Collaboration". Princeton University Press, New Jersey.

Biles, J.A. 2002. "GenJam: Evolution of a Jazz Improviser" in Bentley, P.J, and Corne, D.W. (eds.): *Creative Evolutionary Systems*. Academic Press, San Diego. pp. 165- 187.

Blackwell, T. M. and P. J. Bentley. 2002. "Improvised music with Swarms". Congress on Evolutionary Computation. Piscataway, New Jersey. pp. 1462-1468.

Blackwell T. M. & Young, M. "What is a Live Algorithm" *Live Algorithms*. viewed 4/6/2007. <http://www.livealgorithms.org>

Blackwell, T. M. 2007. "Swarming and Music" in Miranda, E. R. & Biles, J.A. Evolutionary Computer Music. Springer-Verlag, London. pp. 194-217.

Bohemian Productions. 2007. "A Prisoners Dilemma". *A Prisoner's Dilemma.* viewed 4/6/2007. <http://www.aprisonersdilemma.com>

Cage, J. 1961. "Silence: Lectures and Writings by John Cage". Wesleyan University Press, Middletown, Conneticut.

Cycling 74. 1990-2007. "MaxMSP". software. last retrieved 4/6/2007. <http://www.cycling74.com>

Didkovsky, N. 1992. "Lottery: Toward a Unified Rational Strategy for Cooperative Music Making" in *Leonardo Music Journal*, Vol 2, No. 1, pp. 3-12.

"Dr Strangelove: or how I learned to stop worrying and love the Bomb". 1964. Motion Picture. Hawk Films. Directed by Stanley Kubrick. Starring Peter Sellers.

Harrald, L. 2005. "Fight or Flight: towards the modelling of emergent ensemble dynamics" in *Proceedings of the Australasian Computer Music Conference (ACMC05)*, ACMA, Fitzroy, Australia.

Hofstadter, Douglas R. 1983. "Metamagical Themas", in *Scientific American*, Vol 248, Issue 6, p. 14.

Jehan, T. 2001 "pitch~". MaxMSP External. Last retrieved 4/6/2007. <http://web.media.mit.edu/~tristan/maxmsp.html>

Lewis, G. 2000. "Too Many Notes: Computers, Complexity and Culture in Voyager" in *Leonardo Music Journal 10*. pp. 33-39

Lewis, G. 2004. "Improvised Music after 1950: Afrological and Eurological Perspectives". in Cox, C. and Warner, D. (eds.) Audio Culture: Readings in Modern Music. Continuum International Publishin Group, New York. pp. 272-284.

Poundstone, W. 1992. "Prisoner's Dilemma" Double Day, New York.

Pritchett, J. 1993. "The Music of John Cage". Press Syndicate of the University of Cambridge, New York. pp. 48-50.

Synthesising Touch: Haptic-rendered Practice Carillon

Mark Havryliv
Fazel Naghdy
Greg Schiemer
University of Wollongong
Australia
mhavryliv@gmail.com,
{fazel, schiemer}@uow.edu.au

Abstract

This paper describes the design and construction of a prototype haptic carillon baton, and mathematical modelling of the carillon mechanism.
Other research which haptically renders the grand piano mechanism inspires analysis of the kinematic constraints of the carillon mechanism. Analysis is used to construct a physical model using Simulink.
This is then implemented numerically in a Java application. A microcontroller is programmed to interface the prototype's motor and force sensor with a desktop Java application, allowing realtime simulation of the computational model in conjunction with the prototype.
A strategy for containing all physical model computations on an AVR Microcontroller is outlined; this is designed to allow stand-alone operation of the carillon, removing the need for any other external computing hardware.

Introduction

Haptically rendered instruments are designed to remove a major flaw in otherwise useful 'practise' instruments; namely, the absence of an authentic sense of touch, or 'feel', which accompanies any instrumental interaction.

'Feel' can be simply defined as force felt by a player at the point of contact with an instrument; a brass player feels the vibration of his lips, a violinist experiences vibration and resistance at the point where the bow is held, a pianist feels different levels of resistance at the key, etc.

It is possible to predict and recreate (at least, theoretically) forces felt by a player at the point of contact with an instrument by analysing its mechanical properties. This requires an understanding of how the mechanical components of a particular instrument interact, or a *kinematic* analysis, prior to considering the effects of user input.

This type of analysis is already being used to develop physically modelled synthesis algorithms; the interactions between physical components that contribute to sound production are expressed as equations that model an instrument's response to an excitation from a player. A physical model for the synthesis of a violin, for example, will consider the interaction of the bow against the strings, the width of the bow, the damping and resonance of the string, the transfer of energy through the bridge, and the resonance of the soundboard.

A kinematic analysis with a view to haptically rendering an instrument, however, looks at the interactions between physical components that contribute to the force felt by the user. The kinematic constraints of a brass instrument include, for example, the width and depth of the mouthpiece, the length of the tube, and the type of metal used.

In haptically rendered instruments such as the vBow (Nichols 2002) and the Touchback Keyboard (Gillespie 1994, 1996), physical models compute the behaviour of their respective traditional instrument's mechanical systems under different excitations, or gestures, performed by a player. Using sensors (force, position sensors etc.) to monitor a player's gestures, the computational models are able to determine what a player might expect to feel in response, and then actuate this response using a motor.

A thorough taxonomy of new instruments which are either originally conceived, or emulate traditional instruments can be found in the author's previous work (Havryliv et al. 2006).

Mechanics of Grand Piano & Carillon

It is tempting to draw parallels between the conception of a haptic-rendered piano and a haptic-rendered carillon.

The traditional instruments they emulate share a one-dimensional input mechanism (piano: key; carillon: baton) arranged across the instrument in a similar way – i.e. white notes in a bottom row, black notes in the top. Unlike a pianist, however, a carillonneur usually strokes the baton from above with a closed, vertical fist.

In both cases, this one-dimensional input mechanism offers limited scope for controlling timbre: a player may only effect timbre or intensity by controlling the velocity of a striking mechanism (piano: hammer; carillon: clapper) against a vibrating, sound-producing surface (string; bell).

Impact velocity is determined by a player's *gesture*, which is best described as the displacement of the input mechanism over time. This term, gesture, wholly encapsulates the mechanical relationship between a player and an instrument; and it is during the execution of a gesture that a player relies on haptic feedback from an instrument.

It is also in the execution of gesture that differences between the mechanics of a carillon and those of a piano are revealed, and the extent to which pianists and carillonneurs rely on different types of haptic feedback becomes crucial.

Piano

As observed by Gillespie (Gillespie 1996) and others (Oboe et al. 2002), the grand piano mechanism

(Fig. 1) provides a considerable haptic challenge, and as such mechanical and haptic engineers have studied it extensively.

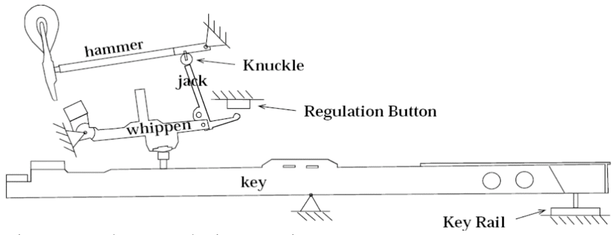


Figure 1. The grand piano action.

The mechanical behaviour of the piano mechanism is characterised by three, discrete phases of different kinematic motions when the key is being pressed downward:

> i) Acceleration – the key is pressed downward but the jack has not yet risen to contact the regulation button;
> ii) Let-off – starts when there is contact between the jack and regulation button, ends when there is no longer compression between the jack and hammer at the knuckle; and
> iii) Free-flight – the hammer is in free-flight towards the string, and the key, whippen and jack remain in motion until the key returns to the key rail.

A primary haptic challenge in this mechanism is rendering the trigger-like feel of the let-off phase. This makes the piano mechanism a system of discrete phases, and the transition between phases of the utmost importance. It is especially crucial for the pianist to have a haptic sense of this let-off transition, as they have no control over the movement of the hammer after this point.

> Our particular interest … lies in the fact that the key and the pianist's finger are completely decoupled from the hammer during its brief period of interaction with the string. The hammer flies free of the jack which initially propels it some 2.5 milliseconds before striking the string. The pianist has no means at his disposal for controlling the tone or the evolution of tone after the hammer has left the jack, except through the damper. Thus all parameters of the tone must be set up by the pianist before tone onset. (Gillespie 1996:9)

Figure 2 displays the motion of a hammer during a gesture which did not impart enough velocity for the hammer to hit the string.

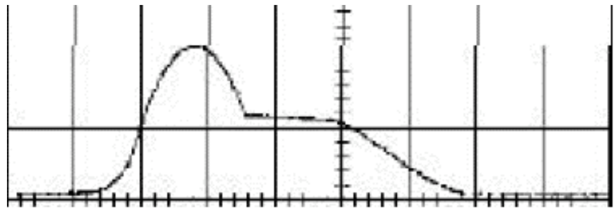


Figure 2. The horizontal line 1/3 of the way up this graph indicates the let-off threshold, i.e. the point at which the hammer flies towards the string; the parabolic trajectory after the let-off demonstrates that the hammer did not impact with the string, but rose and fell according to the laws of simple motion. (Image from Oboe et al. 2002:5)

A corollary to the importance of rendering these transition phases is the relative unimportance of comprehensively rendering the forces felt by a player during non-impact phases. This is evidenced by the common use of position rather than force sensors in haptic piano designs to determine a player's input; the forces felt by a player are closely aligned with the key's position – the mass of the piano mechanism contributes negligibly to the sensation of inertia mid-flight.

This is entirely different to the behaviour of the carillon, in which the combined mass of the mechanism is most certainly a factor that determines the motion of a baton and the force feedback felt by a player.

Carillon

The carillon mechanism (Fig. 3), whilst being a mechanically complex construction, is a simpler kinematic arrangement than that of the piano.

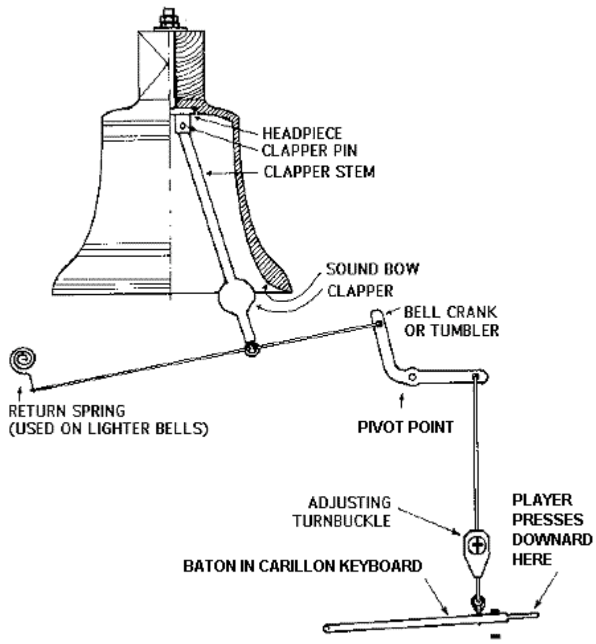


Figure 3. The carillon mechanism. This diagram does not show the complex mechanical arrangements required in a real carillon, but rather a kinematic view of the mechanical interactions which constitute the forces felt by a player.

The above figure identifies the essential elements of the carillon action. A player presses down at the end of the baton, which pulls on one end of a crank, the other end being connected to the clapper. As the crank rotates on a pivot point, the clapper is pulled towards the bell. (Note that the clapper is pulled upwards at an angle, circa 30-45 degrees.) An appropriate kinematic model for this mechanism is the frictionless pulley on an inclined plane of classical Newtonian physics. This is shown in Figure 4 where θ is the angle at which the clapper is pulled up against the inner bell wall, mg sin θ is the negative force applied to the clapper by gravity, Fp is the positive force applied by a player to the clapper and Fnet (i.e. Fp – (mg sin θ)) is the total force acting on the clapper. If Fnet is positive (i.e. the player exerts more force than gravity acting with the clapper's mass does), the clapper will be drawn up against the bell, and if Fnet is negative, the clapper will move away from the bell.

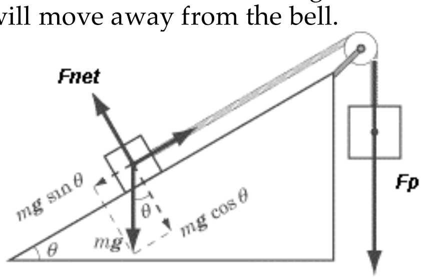


Figure 4. A frictionless inclined plane with pulley, the a simple and expedient model for analysing the forces at work in a carillon mechanism.

Here, it is important to note that we are not primarily concerned with the force felt by a player at the point of contact with the baton, but rather the net force acting on the clapper; this net force is used to determine the displacement of the clapper, and therefore the baton. This simplifies calculations when determining the force applied to the clapper upon collision with the bell.

Physics of the Carillon Mechanism

Continuing the analogy with a frictionless pulley, the small box on the incline represents the clapper, the circular pulley at the top right of the incline is the bell crank, and the larger box hanging down from the pulley is the baton. Note that in the case of the carillon mechanism, a stopper prevents the baton from moving continually upwards and the bell wall prevents the clapper from doing the same (imagine the pulley as an un-passable barrier).

The angle θ is the angle at which the clapper is pulled up against the wall, and is used to determine exactly how much gravitational force is applied to the clapper. In most carillons, this angle changes according to the position of the clapper; the effect of this is addressed further on.

In addition to these basic forces, the carillon mechanism builds in a number of sophisticated mechanical advantages for the player, starting with the baton which allows the player a torque-given advantage over the mass of the clapper of about 1/3rd; this is covered later in this paper.

Forces at work in the Carillon

There are four main forces in the carillon mechanism: the force of the clapper, the force applied by a player (both represented in Figure 4), the force applied at the impact between the clapper and the bell, and the force applied by the stopper to the baton as it returns to its resting position. These are broken into negative and positive forces:
Negative – clapper force and bell impact force;
Positive – player force, and baton stopper force.

Here, negative forces pull the clapper downwards and pull the baton upwards, whilst positive forces pull the clapper upwards and push the baton downwards. It is useful to note that the respective positions of the clapper and the baton are inversely coupled. i.e. when the clapper has covered 2/3rds of the distance upwards towards the bell, the baton is 2/3rds of the distance to its bottom stopper.

Clapper and apparatus force: F_C

A constant force is applied to the clapper, this force is the product of its mass, shape and the angle at which it is pulled towards the bell. The angle of the clapper effects the extent to which gravity pulls it downward. The simple equation:

$$F_c = clapperMass * (gravity * \sin\theta) \quad (1)$$

gives the appropriate force value. (The right hand side of this equation is labelled in Figure 4 as mg sin θ.)

Additional to the clapper, there are mechanical features which add to the weight felt by the player pressing the baton. These are primarily the masses associated with the transmission mechanism which converts the baton's motion to force applied to the clapper. However they are not analysed at this point as they differ dramatically between different carillons, and even different keys on the same carillon.

Further, the computational model does not need to distinguish between this added mass and the mass of the clapper. Because the angle at which the clapper moves towards the bell changes with position, the capacity to handle non-linearity in the force applied by the clapper has been built into the model.

The testing procedure we have developed considers change in both the clapper's angle and other transient forces as a kind of 'black box' – it is enough to understand their effect on the motion of the baton without a detailed understanding of their construction.

Impact between clapper and bell wall: F_I

Upon impact with the bell, a momentary, yet extremely large force is applied to the clapper; this force propels the clapper downwards, away from the bell. A detailed study of the dynamics of bell clapper impact at the National Carillon, Canberra (Fletcher et al. 2002) provides an excellent reference for modelling this impact, as well as providing useful data relating to bell and clapper masses.

The amount of time the clapper is in contact with the bell during impact determines the force of the collision, as well as the character of the resulting sound. Goldsmith (1960) extends Hertzian impact theory and develops an equation for determining the contact time during an impact between a sphere (clapper) and a very massive plate (bell wall).

The contact time (τ_H) is a function of the clapper's mass (m), the clapper's radius (R), the clapper's elasticity, the velocity of the clapper at impact (V):

$$\tau_H \approx 4.5\left[\frac{(\delta_1+\delta_2)^2 m^2}{RV}\right]^{1/5} \tag{2}$$

where

$$\delta_1 = \frac{1-\mu_1^2}{\pi E_1}; \qquad \delta_2 = \frac{1-\mu_2^2}{\pi E_2} \tag{3}$$

and μ_1, μ_2 are the Poisson's ratios (compression on impact) and E_1, E_2 the Young's moduli (elasticity) of the clapper and bell, respectively.[1]

After calculating the contact time, we are able to use this rather extravagant equation to determine the force curve through the impact time:

$$F \approx F_{\max} \sin(\pi t/\tau_H) \tag{4}$$

where

$$F_{\max} \approx 0.44 \frac{m^{3/5} R^{1/2} V^{6/5}}{(\delta_1+\delta_2)} \tag{5}$$

However, experimental results demonstrate that a far simpler impulse equation is just as effective in determining the force exerted by the impact:

$$\mathbf{F_I = (m(v_f - v_i)) / \tau_H} \tag{6}$$

where v_f is the velocity of the clapper after impact, and v_i is the velocity of the clapper prior to impact.

[1] In the National Carillon, the Poisson ratios are 0.37 & 0.29 for the bell and clapper respectively (the bell being bronze and the clapper, iron). The Young's moduli are 124 & 196.

The difference between the two equations results in a negligible change to the haptic response, and can be classed as unnecessarily complex, especially when compared to the gains in computational efficiency when using the simpler equation.

Debounce; baton stopper force: F_B

Like the impact force between the clapper and the bell, the force resulting from impact between the baton returning to its détente position and the baton stopper is similarly momentary. However, unlike the clapper bell impact force, no comprehensive study has accurately mapped the behaviour between the baton and the baton stopper. The point of contact between the baton and the baton stopper is a variably thick piece of felt, which has made such a collision difficult to model.

Further, the tension in the link between the key and the clapper mechanics may not be calibrated properly, and there will still be residual slackness in the link after the baton has hit the stopper, causing back-and-forth jitter before the baton bounces off. (Clavier calibration is relatively straightforward, and is performed using the adjusting turnbuckle, shown in Figure 3. These may slip even during a performance, and carillonneurs often perform minor adjustments while playing.)

Experiential and impartial observations, however, indicate behaviour very similar to that of a bouncing ball, and this has been implemented in our model – with a parameter to alter the degree of compression present in the felt piece, as one can vary the elasticity of a bouncing ball. Figures 5a and 5b demonstrate the effects of changing this parameter.

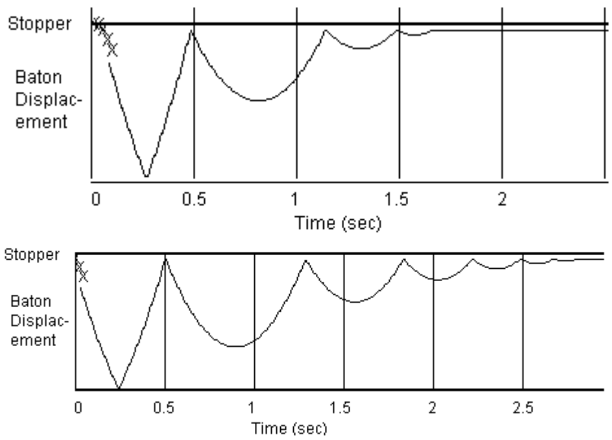


Figure 5: A graph showing the motion of a baton as the result of one gesture. The 'X' points mark the period during which the player is applying force. a) Assumes a thick, dull felt at the baton stopper, absorbing energy and not propelling the baton downwards with as much force as the lighter felt assumed in b). The force applied by a player is the same for both graphs.

Player applied force: F_P

A player may exert a variable force at any time through the performance of a gesture, and this force excites the rest of the system.

Straightforward, linear gestures

If the system is at rest, i.e. the baton is not in motion, a straight forward gesture that the application of a reasonably constant force for the duration of the baton's trajectory will result in a motion like that shown in Figure 6. Figures 6 and 7 show the movement of two clappers, 5kg and 9kg respectively, and are the inverse of baton motion shown elsewhere in this paper.

The following measurements are observed after executing our designed physical model in Matlab, further implementation of this physical model is discussed in a later section.

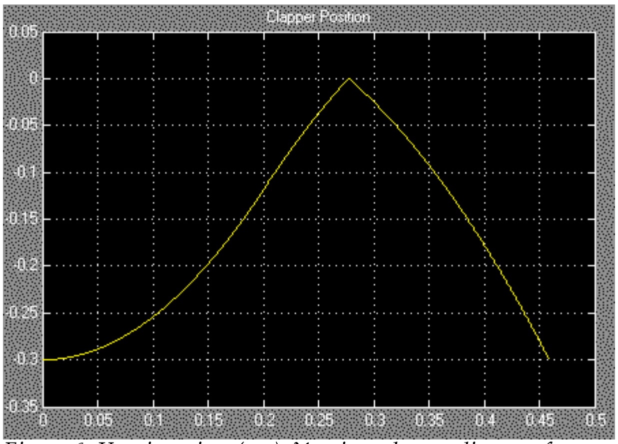


Figure 6. X-axis = time (sec); Y-axis = clapper distance from bell (metres). The trajectory of a 5kg clapper after the application of 80 N of force to the baton for approximately half the baton's full travel distance (until -0.15 on the y-axis).

Depending on the weight of the clapper, applying the same force will result in a different clapper motion. Figure 6 shows a 5kg clapper's motion after a player-exerted force of 80N; Figure 7 shows the motion for a 9kg clapper when the same force is applied for the same period:

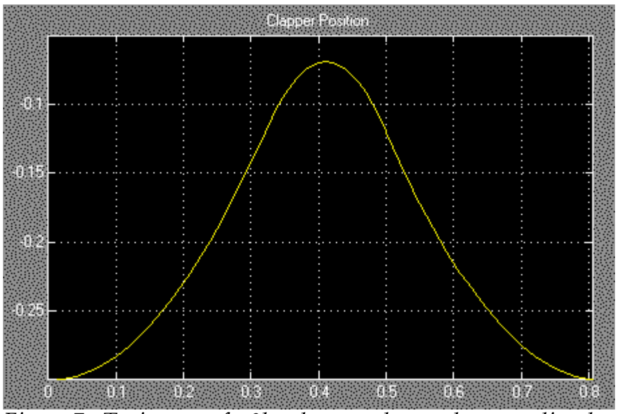


Figure 7. Trajectory of a 9kg clapper when a player applies the same force as that shown in Figure 6.

It is clear from Figures 6 and 7 that differently weighted clappers require different levels of force from a player. In the case of the above example, sufficient force to strike the bell gently and return to the détente position within 0.45 secs in the case of a 5kg clapper was only just enough to carry the 9kg clapper over 2/3rds the distance to the bell, and take 0.8 secs to fully return.

In this execution of simple gestures, the haptic response of the carillon resembles the piano, and indeed, a haptic solution could present itself in the form of a simple resistance mechanism, even springs! It is when more complex gestures, more complex and continuous excitations are employed by a player that the carillon exhibits force-feedback that requires the response capabilities allowed by a realtime physical model.

Complex, continuous gestures

Normal performance practise for carillonneurs involves developing a feel for the mechanical response of the instrument as they play it. It is not only important to develop an attenuation for the initial resistance caused by the mass of a clapper and its associated mechanism. Techniques like tremolo require a performer to interact with continuing forces in the carillon, such as the momentum a clapper builds up as it returns from an impact with the bell. A computational model based on a kinematic analysis makes this reasonably straightforward: Consider Figure 8, in which repeated gestures of the same magnitude and duration are performed by a player and note the difference in response from the system:

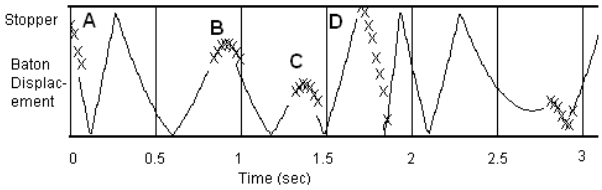


Figure 8. A graph showing the position of a baton through time as a player executes a series of gestures. The 'X' marks show when force is applied by the player to a baton; each gesture is exactly the same application of force over the same period of time.

Figure 8 shows show a series of complex continuous gestures recorded over 3 seconds. Gestures are denoted by 'X' marks. The first gesture is a straightforward attack of considerable force that propels the baton from the stopper with enough impact force to hit the bell again. The gesture at B catches the baton as it rises after the clapper's collision with the bell. This impact poses a counter-force that propels the baton upwards even as the player applies force to it. The same happens with the gesture at C.

The gesture at D illustrates the role of momentum as the baton moves downward from its impact with the stopper, the player adds to the downward force. Note the visual concentration of 'X' marks, which are less dense as the baton travels further during a period of time. The first and fourth gesture occur at the same point in the baton's trajectory. However the fourth gesture has the advantage of force imparted by the collision between the stopper and the baton. Consequently, it propels the baton significantly further and quicker than the first gesture, which was applied when the baton was stationary.

Implementation of a Carillon model

Simulink, Matlab's GUI solver for its ODE (Ordinary Differential Equation) engine, was used to prototype the physical model derived from the kinematic analysis established in previous sections. Whilst useful in testing the model's performance there are foreseeable budgetary constraints in controlling as many as 53 motor/sensor control systems – as required for the National Carillon. Simulink creates effective microcontroller code from GUI models, however particular supported microcontrollers are themselves costly.

The authors' proclivity towards low-level implementations also played a part in deciding to use a low-cost AVR[1] microcontroller to control the servomotor attached to the prototype (Schiemer et al. 2004). The cost advantages of realising this using open source software also played a part in deciding to develop the realtime physical model engine in Java.

Java and AVR microcontrollers

A desktop Java application was developed which solved the physical model in realtime according to external excitations, i.e. player applied force, test software applied force etc. Rather than using standard ODE solver software already available to Java, it was with a view to implementing the haptic engine for each baton on a dedicated 8-bit microcontroller that the Java engine was written to solve equations numerically, at steps of 1/1000th of a second. This is also the non-realtime rate at which Simulink solved the original implementation.

Furthermore, it seemed sensible to aim towards a microcontroller solution given that a microcontroller will in any case be required to read the force-sensor data from the prototype, as well as controlling the motor.

Java engine

The speed at which realtime physical modelling can be performed is a frustrating constraint; as a solution to this, a Java engine was developed which reduced the physical model to a small number of discrete elements and processes. Shown in Figure 9, separate forces are summed together in order to determine the net force being applied to the clapper. Note the positive and negative forces respectively. This net force is integrated once to determine velocity, then integrated again to determine position. This position data is then sent to the microcontroller which is controlling a motor, the hardware arrangement is discussed in the next section.

[1] Atmel AVR Microcontrollers (http://www.atmel.com)

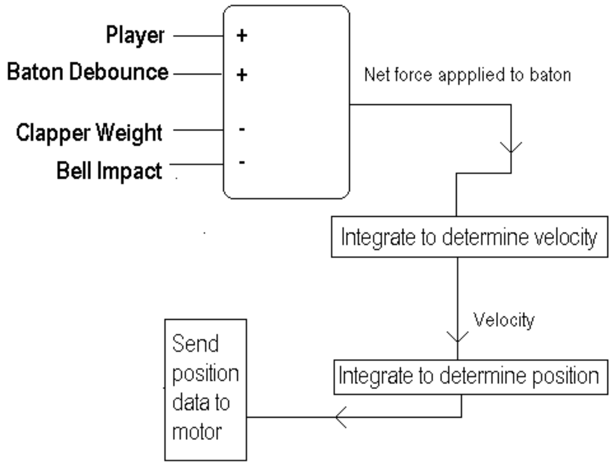


Figure 9. Simplified view of evaluations made by Java engine.

Integration is performed using the standard equations which convert static (non-changing) forces and motions using averaging (Eq. 5); however the very small step size at which these are performed allows *constantly changing* forces and motions to be integrated correctly:

$$\Delta p = v_i t + \tfrac{1}{2} a t^2 \tag{7}$$

where Δp is the change in position from the last time this equation was solved. The only data this method is required to store between realtime solutions is the previous velocity (v_i), the length of time since it was last solved (t) and the current acceleration (a), which is easily determined from the net force.

Whilst the realtime engine solves every 100 uSecs, it is only polled by the microcontroller every 1-2.5 mSecs; this enables it to solve in quick bursts before waiting a small period to allow time to pass and the microcontroller to request an updated motor position.

AVR and hardware

The AVR ATMega16 microcontroller is used to interface hardware components to the Java desktop physical model engine. Hardware components include a GWServo S04 BBM[2] servomotor and an Interlink[3] force sensor. A servo motor was chosen for its low power consumption, high torque, and because it accepts position, rather than velocity, commands which removes the need for a separate position sensor in the prototype – given that the microcontroller can be reasonably certain that the motor is in the position it was commanded to.

The servomotor accepts pulse width durations of between 1.164 to 2.055 mSecs to control its position over 180 degrees; the motor requires a delay of at least another millisecond before the next pulse

[2] 13kg/cm torque servo motor, available from Jaycar.

[3] http://www.interlinkelectronics.com/ FSR Part number 402

can begin. A fixed servo cycle of approximately 333 Hz (3 mSec cycle) is used.

Using the 16-bit interrupt-driven timer port to control the PWM allows the program code to use the entire servo cycle to perform serial transmission with the Java engine; this includes both transmitting the sensor value and receiving an updated motor position (Fig. 10).

Sensor data is transmitted to the desktop Java application at the beginning of a servo cycle, allowing the full cycle to be used for calculations and preparation of a new motor position command, in the form of a pulse width for the next cycle.

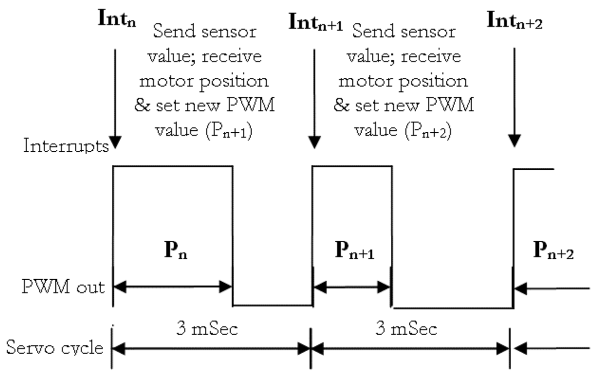


Figure 10. Timing diagram demonstrating integrated interrupt-driven PWM output and motor position calculations.

Mechanical Construction

The carillon allows the player a mechanical advantage over the clapper by employing the principles of torque and rotational force (Fig. 12). In the prototype haptic carillon (Fig. 11), the opposite effort has been made: to give the motor a mechanical advantage over the user.

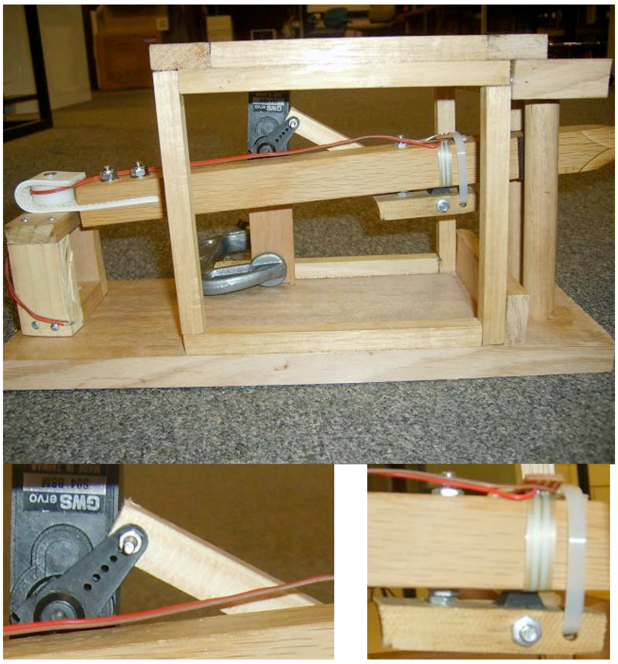


Figure 11. Prototype jig with motor and force sensor attached; b) close-up of the motor arm; c) close up of sensor attachment.

A video of this in action can be viewed at http://www.uow.edu.au/~mh675/baton2.wmv.

Although the baton appears to travel linearly, the mechanism is in fact a rotational one; this alters the amount of force the clapper applies to the bell than if the system were linear. The torque advantage is a function of the distance between the application of force and the pivot point, and the angle at which the force is applied; in the carillon's case the angle is the same for the clapper and the player, allowing the advantage to be expressed as L_P / L_C.

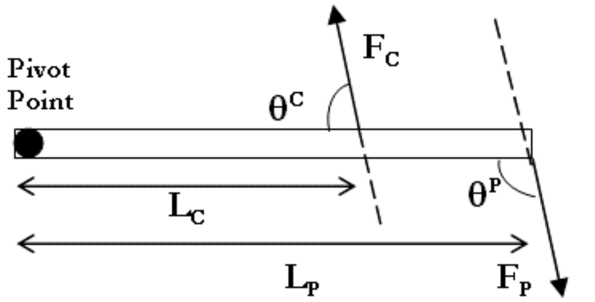


Figure 12. The player has the advantage of applying force over a greater distance to the pivot point than the clapper does; no further advantage is obtained by the angle at which forces are applied, as the angles θ^C and θ^P are complementary.

In the haptic carillon, the motor is coupled to the baton through a stiff wooden component which is free to rotate at both its connection to the motor arm and the baton. This configuration changes the angle at which the user applies force to the baton, and takes advantage of the torque law when angles are involved:

$$Torque = r * F * \sin\theta \qquad (8)$$

where r is the distance from the pivot to the point at which force is applied, and θ is the angle at which force is applied. This wooden link applies a force at an angle which reduces the torque the user can apply to the motor, whilst allowing the motor arm to rotate only a minimal distance to move the baton through its full trajectory.

Conclusion

This paper describes the design and construction of a prototype haptically rendered carillon baton. Using previously developed kinematic models of a grand piano action, our design has recognised the respective differences in performance technique which warrant a uniquely constructed model of a carillon action.

Also described is a method for the implementation of this model in discrete microcontrollers for cost-effectiveness, compactness and speed.

Developed physical models have been tested in Matlab, Java and realtime hardware implementations, with results in concord with behaviours observed at the National Carillon, Canberra.

Conclusion

This is an ongoing project supported by an Australian Research Council (APAI) Linkage in partnership with The National Capital Authority, ACT and Olympic Carillon International (Seattle).

References

Cusbert, KXT, J. (2006) In discussion with; June 2006.

Gillespie, B. (1996) PhD Dissertation " Haptic Display of Systems with Changing Kinematic Constraints: The Virtual Piano Action" Stanford University.

Gillespie, B. (1992) "Dynamical Modelling of the Grand Piano Action", *Proceedings of the International Computer Music Conference*, San Francisco, USA.

Havryliv, M., Schiemer, G. & Naghdy, F. (2006) "Haptic Carillon: Sensing and Control in musical instruments", *Proceedings of the Australasian Computer Music Conference*, Adelaide, Australia.

Goldsmith, W. (1960) *Impact*. Arnold, London.

Nichols, C. (2002) "The vBow: A Virtual Violin Bow Controller for Mapping Gesture to Synthesis with Haptic Feedback". Organised Sound Journal.

Oboe, R. & De Poli, G. (2002) "Multi-instrument virtual keyboard – The MIKEY project" *Proceedings of the 2002 Conference on New Instruments for Musical Expression (NIME-02)*, Dublin, Ireland

Schiemer, G. & Havryliv, M. (2004) "Wearable Firmware: The Singing Jacket" *Proceedings of the Australasian Computer Music Conference*, Wellington, New Zealand

Alistair Riddell
Centre for New Media Arts
The Australian National University
Canberra, ACT. 0200
Australia
alistair.riddell@anu.edu.au

Towards Interactive Gesture Control of Sound Patterns Using the Wiimote™

Abstract

This paper discusses research towards an interactive gesture based control strategy for sound event patterns or sequences. It posits a direction, which seeks to nuance such sound structures, more commonly thought of as fixed, with traits reminiscent of "live" performance. Such rigid formalism is common in many music sequences programs and while a great deal of effective music has been produced in this manner to date, the option of a range of control over sequenced sound introduces the possibility of an entirely new performance and aesthetic level.

Consider repetition in contemporary electronic music, which has become an essential and extensively used structural device. In Jazz, the practice of repetition is particularly sophisticated because it is directly under the control of the musicians who understand how to exercise variation and who can carefully modulate between direct statement and complex variation. One need only think of certain groove patterns in the hands of skilled players.

In seeking a similar relationship with computer-generated sound sequences, the research described here uses established real-time software, SuperCollider 3 (SC3) and a Nintendo Wiimote™ controller from the Games world as the physical interface. In using the Wimote, it is anticipated that the performer might have a more physical relation to the production and deployment of sound in sequences. The objective is to establish a dynamic/ theatrical relationship with complex sound structures through sensor data derived from human movement.

Introduction

Personal precedent for this research is founded in the work with *HyperSense Complex*[1] (Riddell 2005) around 2003–4. In the process of creating new performance approaches for this group, Simon Burton explored a strategy for working with sequences in an interactive manner. Although idiosyncratic, it constituted a response to developing performance approaches for this new interactive context.

Burton's idea was based upon cycling temporal placeholders or "buckets", like defined spaces on a sequential conveyor belt (initially silent buckets) [2]. These buckets could have sound event placed in them by the performer. (see Figure 1.)

When the performer triggered an event, it was placed in the nearest temporal slot in the cycle which up to that point the performer may not have either heard or appreciated was there. If the event coincided with slot one, the start of the cycle, it was given greater emphasis (attack characteristics and amplitude). This was most noticeable when the sound had naturally acute attack characteristics.

In simply bending a finger, the performer could place in those slots sound samples, which could then be manipulated in a variety of ways, such as by pitch or amplitude change or even simply muting by similar gestures with other fingers. Amplitude variation provided a very subtle but effective way to nuance a sequence. The performer could also remove events in the same way by bending a finger.

Further to this temporal activation of sound, Burton added the idea of branching, which was the subdivision of time below the initial cycle durations. So it became possible to create more complex rhythmic patterns for one performer.

```
Level 1 - 1     2     3     4     5    1,
Level 2 – 1  2  3  4  5  6  7  8  9 10 1,  etc.
```

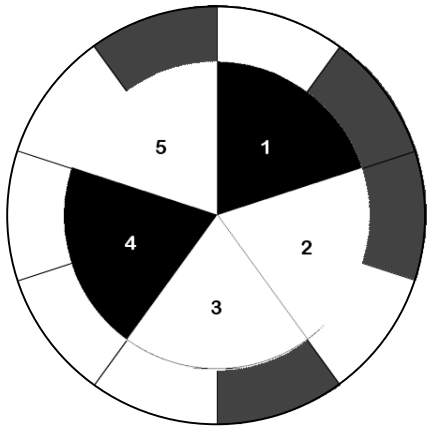


Figure 1. Diagram of Burton's "Silent Bucket" concept

The performer could thus add or subtract events from the sequences at two temporal levels. In retrospect, this composition process was challenging to engage. It required concentration on one's own patterns while considering the sum of the parts aesthetically. Activating and removing

[1] The performance trio *HyperSense Complex* used Flex sensors on the fingers to control sounds and their behaviour. *HyperSense* was a networked ensemble sharing the same audio processing space (see arrowtheory.com/hypersense/index.html).

[2] The work became known as "Drumming Tree".

events was frequently arbitrary. Precise management of events took considerable concentration in a context that could be externally influenced through effects imposed by one of the other performers.

Burton's approach was the result of thinking through a particular performance concept. Since there was no other option other than to use the sensors at our disposal and which in themselves, formed the central platform of our performance agenda, it was just assumed that the concept would engage that type of sensor and associated performance skill.

The idea of silent placeholders that could be filled with sound events was novel. It did provide some cognitive reassurance that a stable and predictable world existed into which one we could develop an improvisatory position. Even if we could not initially hear the temporary framework, it quickly became evident in the early stages of a performance.

If this approach had been for only one performer, the idea would have perhaps needed expansion or augmentation to include a greater range of parameters to expand the greater diversity in the compositional domain. This would likely have happened with time, practice and musical evolution, as is the case of all musical instrument development.

However, the initial implementation was realised through three performers acting simultaneously and in this respect, the relative simplicity of the system was obscured by a complexity of three performers aspiring to a unified sound. Three performers, each with differing sounds and effects, could create a compelling ensemble effect and a more engaging sound experience.

The network performance configuration, discussed elsewhere (Riddell 2005) and explore by artists like Phil Burke (Burke 2000) in recent years, remains an intriguing approach to the contemporary idea of the ensemble with many variations possible. While not explicitly discussed in this paper, it remains a logical extension to the practice described here.

So the previously described personal experience tangibly underpins this research into gesture and musical performance but not to the exclusion of other research the findings of which are presented at many conferences (CHI and NIME, etc). What is interesting is that much of the research seeks to explore the area through similarly idiosyncratic means, which it could be argued parallel the development of musical instruments in general. The body is a primary and logical source of expression as discussed in Strachan et al (Strachan 2007), who outline an approach to controlling a music player. Hashimoto et al (Hashimoto 2005), take a more conventional position, which considers that "in traditional musical instruments, the relationship between the action and the generated sound is determined by the physical structures of the instruments." There is some logic in that statement but Goto and Suzuki (Goto 2004) engage virtual musical instruments and sound synthesis in an attempt to explore gesture and performance beyond traditional instruments. These researchers, in part, circumscribed directions in gesture and musical performance currently in vogue.

Research Agenda

In practical terms, the research direction taken here differs significantly from that of the *Hyper-Sense Complex* project in both technology and anticipated creative outcome. The principle interest here is the extent to which patterns or sequences of sounds can be manipulated with arm based gesture control of a different physical character.

The research here does hold, as a fundamental position, the ideal of human instrumental performance but additionally recognizes that the mapping of such behaviour as that which lends traditional music a great sense of humanity, may not be possible or even desirable in respect to complex computer generated events. This research does, therefore, acknowledge an ambition to explore uncharted areas of interaction in the digital sound domain.

At the core of the approach is the use of sophisticated pattern generating systems and subsequent functionality associated with the inherent messaging system of the audio application, *SuperCollider 3*. It is recognized that other applications of a similar status (MAX, PD, Csound, etc.) could possibly be used as instead.

Technology

This research exploits the sophistication of Nintendo's Wiimote controller, which is extensively documented on the Internet (sites in reference section) and contains the following functionality:

Sensors
- Depth
- Tilt
- Position

Buttons
- 1 Trigger
- 6 Buttons
- 4 way tilt

Features
- Infrared receiver for depth and position
- Speaker for depth-of-sound
- Rumble motor
- 4 Led Display

Table 2. WiiMote Functions

The commercially available Wiimote functions using *BlueTooth*™ communication technology and possesses a unique physical design that has been researched by Nintendo. Admittedly, the intended use of the Wiimote is for computer games that require a more physical interaction but exactly what is possible at a subtle level of creative sound engagement is yet to be fully appreciated.

From the above list of functions and internal operations, it can be appreciated that the Wiimote has considerable potential as a sound performance device. This can be understood through the proliferation of such applications for the Mac, PC and Linux systems as *DarwiinRemote*, *Glovepie*, *Remote Buddy*, *WiiViewer* and *WiinRemote*.

This research is primarily interested in the Wiimote implementation for *SuperCollider* (SC3) developed by Pete Moss. However, it is worth noting that there exists a *Wiimote-api* which is "a library written in C to access the Wii Remote's abilities" (software reference). This offers a near complete set of functions pertaining to the Wiimote for those who wish to develop a different strategy for using the Wiimote.

Wiimote and SC3

The following discussion is based on Pete Moss' Wiimote implementation for SC3, which consists of the classes: `Wiimote.sc, WiimoteUgens.sc` and the plugin, `Wiimote.scx`. There is basic functionality for all the buttons and X, Y and Z movement, Rumble and LEDs. Help is also included in the distribution. Wiimote functionality is represented in the following Ugens:

WiimoteX	accelerometer Ugen
WiimoteY	accelerometer Ugen
WiimoteZ	accelerometer Ugen
WiiMotion	Motion Ugen
WiiTrig	Motion Trigger Ugen
Wiimote1	Button "1" Ugen
Wiimote2	Button "2" Ugen
WiimoteA	Button "A" Ugen
WimoteB	Button "B" Ugen
WiimoteMinus	Button "-" Ugen
WiimotePlus	Button "+" Ugen
WiimoteHome	Button home Ugen
WiimoteUp	Tilt up Ugen
WiimoteDown	Tilt down Ugen
WiimoteLeft	Tile Left Ugen
WiimoteRight	Tilt Right Ugen
WiiRumble	Rumble motor Ugen
WiiLED	LED Ugen

Table 3. Ugens available in Wiimote.sc

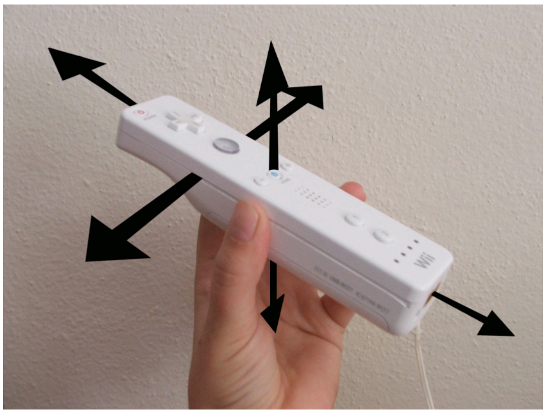


Figure 2. WiiMote Axes

So the Wiimote Ugens are controllers defined in the form:

```
   WiimoteX.kr()
   WiimoteY.kr()
   WiimoteZ.kr()
```

Ex. 1. X, Y and Z movement Ugens

Here control data received from the X and Y axis in the range 0 –1.0.

This nomenclature defines the basic operation of the key functions of the Wiimote. It is curious to reflect on the idea of using vibration as a means of experiencing a kind of haptic response to event triggering. This functionality is not represented in the following examples but would probably be easy enough to implement to, for example, delineate usable ranges of the X, Y and Z parameters.

It can be appreciated from the above description that the SC3 Ugens are control rate and so are used only inside SynthDefs, that is, on the server side of SC3. This functionality will not be discussed in detail here because it is not pertinent and also more or less self-evident from the examples for the Ugens.

Connectivity

After following the Wiimote initialization procedure control data is available in the SynthDefs. Wiimote data at this point is not useful for the pattern structures and will need to be processed through an OSCresponder, which facilitates client-side data in the following manner:

```
SynthDef(\wiiX, {
   SendTrig.kr(Impulse.kr(10),
   13,LinLin.kr(WiimoteX.kr(),
   -0.17,0.2,0.4,2.0));
}).load(s);
)
s.sendMsg("/s_new", "wiiX", node =
s.nextNodeID,1,1);

OSCresponder(s.addr,'/tr', {| t,r,msg|
// SendTrig data can be accessed here.

}).add;
```

Ex. 2. Client side "X" axis data

Now that we have data on the client side, it is possible to pass it to a pattern structure under certain conditions. In this particular instance, SendTrig is passing the control data from the "X" axis of the Wiimote to the OSCresponder at the rate of 10 messages per second. What the rate should be depends on the use of the data in the pattern structure. It would be expected that the pattern would be accessed in an OSCresponder as:

```
OSCresponder(s.addr,'/tr', { arg
time,responder,msg;
if ((msg[1] == node), {
   amp = msg[2];
   h = [0.125,Prand([0.1, 0.5, 1.0],
   2),0.5,0.125];
   j = Pseq(h, inf);
   k = Pseq((0..7) + 15.rand;
   Pdef(\x, Pbind(\dur, j, \amp, amp, \degree,
k, inf)));
});
}).add;
```

Ex. 3. A pattern accessed on the client side

Defining Pattern Control

Given the two control data access modes, continuous data or toggle state data, pattern structures should be able to use data from at least one. This leads to the question of what is the form of a pattern structure? Consider the following, which represents a basic structure defined through the class Pbind. This assumes an instrument called "someInstrument" which takes arguments defined by the names preceded by a "\".

```
r = Pbind(
   \instrument, 'someInstrument',
   \dur, Pseq([0.5,0.8,0.8,0.4,0.2],inf),
   \amp, 0,
//Prand([2.0,6.0,5.0,7.0,7.0,6.0,5.0,7.0],inf),
   \rate, Prand([0.6,Pseq([-1.5,0.7,-
   0.6,0.8,0.6,-1.5],Prand([3,4.0],1))],inf),
   \startpos, 109,
   \bufnum, 1,
   \panrate, Prand([0.9, 1.0, 0.5,0.3, 0.1,
   3.0, 2.0],inf),\release, 0.95
);
g = r.play; // This plays the above structure
```

Ex. 4. A pattern structure

The point to note here is that while this structure defines an initial state of the pattern it can be modified by making it a Stream and re-initializing it. This example shows the next stage in the creation of an interactive pattern:

```
//  g
g.stream = Pbind(
   \instrument, 'someInstrument',
   \amp,
   Prand([5.0,3.0,7.0,4.0,7.0,1.0,5.0,7.0],inf)
   ,
   \dur,
   Pseq([0.5,0.2,0.3,0.4,0.2,Pseq([0.15],6)],in
   f),
   \startpos, 107,
   \release, 0.95,
   \bufnum, 1,
   \panrate, Prand([0.9, 1.0, 0.5,0.3, 0.1,
   3.0, 2.0],inf),
   \rate, Prand([0.6, Pseq([-1.5,0.7,-0.6,-
   0.8,0.6,-1.5],Prand([3,4],1))],inf)
).asStream;
```

Ex. 5. A pattern structure as a Stream

Data for any of the second elements of the duple , e.g. `\release, 0.95` can be modified and the structure selected and executed in the conventional SC3 manner. Thus dynamically changing the structure.

At this point, however, there is no connectivity with the Wiimote nor is there provision for changing the content of the pattern structure "g".

Reflecting on the initial objective of this research, the primary purpose was to modify pattern repetition in such a way that retained certain characteristics while introducing variation of a particular type.

Basic functionality includes the ability to start, mute and terminate sequences as well as vary internal parameters.

The use of the Wiimote invariably leads to considerable data input because the fundamental operation is as a continuous controller. This is necessary for anything that is expected to operate in real-time. The question here is can such a continuous controller provide data to nuance patterns and still feel to the user like a real-time interactive device? This question will become more pertinent at the end of this paper.

The operational characteristics of the Wiimote suggest that discrete control data requires skilled use. Finely grained values are the outcome of careful control behaviour.

```
SynthDef(\wiiX, {
   SendTrig.kr(Impulse.kr(10),
   0,LinLin.kr(WiimoteX.kr(),-0.175,
   0.23,0.0,0.9));
}).load(s);

// Below here just to show initialization
s.sendMsg("/s_new", "wiiY", node =
s.nextNodeID,1,1);

OSCresponder(s.addr,'/tr', {|
time,responder,msg|
if ((msg[1] == node), {
   amp = msg[2];
   h = [0.125,Prand([0.1, 0.5, 1.0],
   2),0.5,0.125];
   j = Pseq(h, inf);
   k = Pseq((0..7) + 15.rand;
   Pdef(\x, Pbind(\dur, j, \amp, amp, \degree,
k, inf)));
});
}).add;
```

Ex. 6. Client side data used to control amplitude

The above example suggests that the amplitude of the pattern is being controlled not the individual sounds of the pattern. The difference, if that is the case is not significant and the more efficient way to do that is directly in the SynthDef.

Using variable controller data, not just toggle state data, can be handled as:

```
SynthDef(\wiiX, {
   SendTrig.kr(Impulse.kr(10),
   0,LinLin.kr(WiimoteX.kr(),-
   0.175,0.23,0.0,0.9)
   );
}).load(s);

// Below here just to show initialization
s.sendMsg("/s_new", "wiiY", node2 =
s.nextNodeID,1,1);

h = Array.fill(rand(13) + 1,{rrand (0.0125,
2.0).round(0.01)});
asize = h.size;
k = Pseq(((0..asize) * 50) + 300, inf).postln;

patfunc = {|parg, asize|
   if (parg > 0,{h[rand(asize)] = parg;});
   j = Pseq(h, inf);
};

patfunc.value(1,asize);
Pdef(\x, Pbind(\instrument, \theSynth, \dur, j,
\amp, 1.0, \ffreq, k));

OSCresponder(s.addr,'/tr',
{|time,responder,msg|
```

```
// Y axis
if ((msg[1] == node2), {
   if(running == 1, {
      patfunc.value(msg[3], asize);
      n = k.next;
      Pdef(\x).set(\dur, j);
      Pdef(\x).set(\ffreq, n);
   });
});
}).add;
```

Ex. 7. "Y" values inserted into a sequence

In the above code, the "Y" axis values are randomly inserted into the duration sequence. The length of the pattern is randomly determined at the start. The pattern gets updated 10/s but that could be slower and made variable by updating the SynthDef to change the Impulse trigger rate.

Buttons can be used for selective operations:

```
SynthDef(\wiiB, {
   SendTrig.kr(WiimoteB.kr(0.0),0,0.0);
}).load(s);

if ((msg[1] == node8), { // Button B
   h = Array.fill(rand(13) + 1,{rrand (0.0125,
   2.0).round(0.01)});
   asize = h.size;
   j = Pseq(h, inf);
   k = Pseq(((0..asize) * 50) + 300, inf);
   Pdef(\x, Pbind(\instrument, \theSynth, \dur,
   j, \amp, 1.0, \ffreq, k));
});
```

Ex. 8. A pattern changed through button activation

The above code allows button "B" to create a new pattern during performance.

The discussion here has focused on the operational state of the Wiimote and has not taken into account any skill level or experience. It is understood that this will have a considerable influence on the development of these operations.

Temporal Nuance

The tempo rate and amplitude changes afforded by movement in the X and Y axes exercise some degree of influence on a running sequence. This may not be particularly exciting but that could depend on the chosen temporal structure and sounds used. However, to further add to this chemistry of action, "Z" axis data could be used to create subtle changes in the timing of the structure.

The primary question is what are subtle changes to the timing of the structure? Let's initially assume that these changes are very small and that we want to alter the sequences based on how many elements there are in the sequence. Further subtlety might be found in dynamically changing the attack points of certain critical events within the structure. This assumes a recognizable metric structure like 4/4, etc. Returning to the build structure function (Ex. 9), something to generate regular metrical events could be constructed by:

```
h = Array.fill([2,4,8,16].choose,
{1/[2,4,8,16].choose});
```

Ex. 9. Generating a metrical structure with durations

The array "h" could look like one of these:

```
[0.5, 0.5]
[0.5, 0.5]

[0.25, 0.25, 0.0625, 0.0625]
[0.25, 0.25, 0.0625, 0.0625]

[0.0625, 0.125, 0.125, 0.0625, 0.5, 0.0625,
0.25, 0.25, 0.125, 0.25, 0.5, 0.5, 0.25, 0.5,
0.0625, 0.125]

[0.0625, 0.125, 0.125, 0.0625, 0.5, 0.0625,
0.25, 0.25, 0.125, 0.25, 0.5, 0.5, 0.25, 0.5,
0.0625, 0.125]
```

Ex. 10. Arrays generated by Ex. 10

These divisions of 1 provide the basis for a temporal structure that can be subtly modified to reduce the consistency of the values.

Taking this further, the author has reviewed another technique for sequence variation based on Jaffe's (Jaffe 1985) time map method of sinusoidal tempo oscillation and more recent work by Collins (Collins 2003).

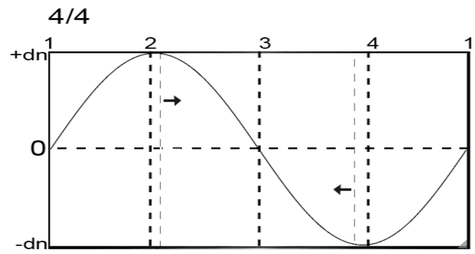


Figure 3. A simple sinusoidal mapping onto 4/4

The application of sinusoidal data to timing information within a defined temporal sequence seeks to achieve temporal displacement through dn augmentation, within that time frame but ensure correct overall temporal accuracy between sequences. In this case, a rhythmic value can be altered by the addition of a corresponding sine value.

```
var inc = 0;
d = 2pi/512; // Two PI and resolution
// create an empty array
a = Array.newClear(0);
512.do({|i|
// fill with Sine data
   a = a.add(sin(inc));
   inc = inc + d;
});
```

Ex. 11. Filling an array with Sine data

Having stored signal data in an array, it is just a question of using the data to compute certain points against the incoming data. This data is part of a process to converted deviation time, which can be added or subtracted to the event time to a performance out come. The event time is calculated in terms of the total sequence duration. Although this description is rather simplistic, it provides the foundation of a result for corrected deviations that maintains temporal integrity of the sequence duration.

Such temporal shifts in the rhythm domain seek to emulate not only the underlying stylistic and effect foundations of music such as Jazz but music of other world cultures where temporary displacement is an operational component in defining the power of the music. This technique and several others, which aspire to the same functionality, have been discussed in more technical detail in McGuiness (2004). However, they do not take into account the idea of influencing extant patterns by performance interaction.

Future Directions

The idea of a "pattern" has until this point been defined loosely as the repetition of a collection of events, possibly 3 or more, with the potential of degrees of variation that, for the most part, do not radically alter the pattern structure. The sequence is maintained and recognizable.

Now the function of the pattern will be reconsidered in terms of a sub-structure that is perhaps not audible as a singular repeatable entity, definitely exists but is not necessarily directly repeated. This concept has been understood for many years and has been manifest by various musicians, most notably in the improvisatory solos of Charlie Parker:

> "Still enriching the jazz vocabulary half a century after they came into being, the components of the musical legacy of Charlie Parker are his "personal repertory of melodic formulas that he used in the course of improvising." Owens points out that every noteworthy improviser puts together and reworks "a well-rehearsed bag of melodic tricks....His well-practiced melodic patterns are essential identifiers of his style." Earlier, Owens states that "...his 'spontaneous' performances were actually precomposed in part. This preparation was absolutely necessary, for no one can create fluent, coherent melodies in real time without [it]...."
>
> Of course, there's more than a constant re-combination of patterns and "licks" that makes Charlie Parker's solos so remarkable. Beyond these, and after acknowledging the use of musical "quotes" from sources popular to classical, factoring in his superb musicianship, his "sheer energy and self-confidence," Professor Owens himself reaches a thematic crescendo by detailing "...another factor that helps generate the sense of rightness in his music. Typically entire phrases, and even entire choruses and groupings of choruses, are goal-oriented; they arrive on a final note that lies at the end of a lengthy stepwise descent." This downward-scaling structure was something Parker alone contributed to jazz, perhaps the capstone of his great gift to the art form."(Simon).

The "precomposed" formalism is a strong theory founded in the plausible outcome of a lifetime immersed in an evolving Jazz idiom. The question pertinent to this research is, how might this apply to the Wiimote performance practice?

It will be recalled that an objective of this research was to imbue patterns with a sense of an organic form, structured but individual. The idea of repetition becomes crucial to the aesthetic architecture of the sound statement. Given Parker's talent, genius and musical context, expectations of a practice, using the Wiimote, achieving anywhere near as refined state are premature to say the least. However, the idea of a concatenation of patterns that have some kind of pre-configuration is a logical direction for a device with such a diverse control output.

So implicit in this approach is the establishment of patterns prior to performance. These could perhaps be rehearsed but the immediate question is, how might a performer access them?

In performance, the conscious selection and awareness of patterns in action, would prove to be too inhibiting in the pursuit of a coherent and aesthetically engaging continuity. The sound of the patterns should be the only guide to how they are deployed.

This suggested a rethinking of Simon Burton's idea of silent temporal spaces into which sounds are placed by the action of the performer. Suppose that all the patterns where simply place markers and that the performer's actions claimed them in a temporal space? The challenge lies in the substance of and access to, the performance data.

If the Wiimote data could be used to place sounds in patterns by approximation to the event space, then there may need to be system that selects the patterns themselves. The other problem with this is the sense of real-time engagement. This approach is contributory in nature not immediately interactive.

This direction suggests the possible use of L-Systems (McCormack 1996) in which a grammar could be use to define the construction and filling of patterns according to data received from the Wiimote. While this research has yet to be undertaken it does offer the possibility of formalizing pattern generation in the context of silent temporal units. The obvious problem is that the sense of response in the creation of new patterns or the modification of existing ones would possibly be compromised. To ameliorate this situation, the Wiimote could function on two levels. First, there would be the capacity to generate patterns and modify existing ones. Secondly, allow a more responsive interaction through direct manipulation of SynthDef Ugens. These could work in combination to produce an effective system for interaction. Although challenging to learn.

Conclusion

The Wiimote has, in its brief existence, attracted considerable interest, examination and re-deployment as a controller. It offers useful and interesting functionality in a small package—yet to be fully explored.

The research outlined here has suggested a few modes of engagement to influence sequence structures and their performance. As the structures become more complex, use of the device becomes something that inevitably requires training and practice, just as a traditional instrument does.

More importantly, the gestures become subtle or as expressed at the end of this paper, abstracted. If one were looking for a device where large-scale physical movement of the arms and body in general, resulted in a diversity of sound, then one would be disappointed. Working with patterns/sequences requires listening and responding, which is part of a skill that has to be acquired over time and creative outcomes. Grander theatrical gestures would seem to emerge from mastery of those refined subtle movements that generate real sound nuances.

Acknowledgements

Thanks to Pete Moss for the Wiimote SC3 classes and plugin , Alex Thorogood for SC3 consultation and Annette Marie for graphic elements.

References

Burke, Phil. 2000. *Jamin' on the Web-A new client/server Architecture for Multi-User Musical Performance.* http://www.transjam.com/info/transjam2000.pdf. (accessed 09-04-07)

Collins, Nick. 2003. "A Microtonal Tempo Canon Generation System After Nancarrow/Jaffe". www.cus.cam.ac.uk/~nc272/papers/pdfs/tempocanons.pdf. (accessed 08-4-07)

Goto. Suguru, and Tahahiko Suzuki. 2004. "The Case Study of Application of Advanced Gesture Interfacing and Mapping Interfacing, - Virtual Musical Instrument "Le SuperPlom" and Gesture Controller "BodySuit"". Proceeding of the Conference on *New Interfaces for Musical Expression (NIME04)*, Hamamatsu, Japan.

Hashimoto. Shuji, and Hideyuki Sawada. 2005. "A Grasping Device to Sense Hand Gesture for Expressive Sound Generation". *Journal of New Music Research.* Volume 34, Number 1/June.

HyperSense Complex. arrowtheory.com/hypersense/index.html (accessed 09-04-07)

Jaffe, David. 1985. "Ensemble Timing in Computer Music." *Computer Music Journal, 9:4, pp.38-48.*

McCormack, Jon. 1996. "Grammar Based Music Composition." In *Complex Systems 96: From Local Interactions to Global Phenomena,* R Stocker et at., Eds, ISO Press Amsterdam, pp. 320-336.

McGuiness, Andrew. 2004. *Microtiming Deviations in Groove".* MPhil. Thesis. The Australian National University.

Riddell, A. 2005. "HyperSense Complex: An Interactive Ensemble". Proceedings of the *ACMC'05* conference. QUT.

Simon. Richard. "Charlie Parker: The Young Man, the Music, the Memory." www.ufo-bass.com/bird2.htm. (accessed 08-04-07)

Strachan, Steven, Roderick Murray-Smith and Sile O'Modhrain. 2007. "BodySpace: Infering body pose for natural control of music player". Extended abstracts of *ACM SIG CHI* Conference, San Jose. www.dcs.gla.ac.uk/~rod/publications/StraMurOMo07.pdf (accessed 09-04-07)

Web References

www.wiili.org (accessed 08-04-07))

en.wikipedia.org/wiki/Wii_Remote (accessed 08-04-07)

www.wiili.org/index.php/Wiimote (accessed 08-04-07)

www.wiibrew.org/index.php?title=Wiimote (accessed 08-04-07)

Software

SuperCollider 3. www.audiosynth.com

Pete Moss. WiimoteLib. petemoss.org/SuperCollider/index.html (accessed 08-04-07)

Wiimote-api www.wiili.org/index.php/Wiimote (accessed 08-04-07)

Online Videos

Max Jitter Video Effect control (zoom..etc) www.youtube.com/watch?v=Q1OLhKTTDuM (accessed 08-04-07)

Wii Wiimote Ableton Live Controller www.youtube.com/watch?v=GYjKJoXU9vU (accessed 08-04-07)

Wii + Supercollider www.youtube.com/watch?v=z2Z1Pvt8Poc (accessed 08-04-07)

Pocket Gamelan: tuning microtonal applications in Pd using Scala

Greg Schiemer
Sonic Arts Research Network
Faculty of Creative Arts
University of Wollongong, 2522
Australia
schiemer@uow.edu.au

Manuel Op de Coul
Huygens Fokker Foundation
Muziekgebouw aan 't IJ
Piet Heinkade 5, NL-1019 BR Amsterdam
The Netherlands
coul@computer.org

Abstract

Microtonal tuning has been a characteristic common to many musical traditions yet despite a growing awareness of these traditions among many musicians today, a single system of tuning based on the twelve-note equal division of the octave continues to dominate development of multimedia applications. This paper describes a new software tool developed to export and document microtonal scales for use in computer music and multimedia composition. The tool was developed as a command script written by the first author using an editor, librarian, and analysis tool for musical tunings known as Scala, written by the second author. The tool called scaleplayer.cmd allows tuning to be exported from Scala to Pure Data, an environment for algorithmic composition where novel purpose-built performance interfaces can be prototyped easily. The tool allows composers to interact with thousands of historical and novel scales and to develop a user interface based on a new understanding of the tuning characteristics being explored. Pure Data has already been used to create performance interfaces for the Pocket Gamelan, a project that has allowed new microtonal tunings to be implemented and performed using Blue-tooth enabled mobile phone technology.

Introduction

The Pocket Gamelan project was launched by the first author in 2003 to address the challenge of composing music for a mobile computing environment. Central to this project was the development of an interactive musical performance interface that allowed non-expert performers to perform microtonal music using mobile phones. Several mobile performance scenarios were implemented as a way to explore the musical legacy of historical tuning systems as well as the tuning systems first explored by composer and theorist Harry Partch and later extended through the work of contemporary tuning theorist Erv Wilson (Schiemer and Havryliv, 2006).

Two new microtonal works for mobile phones have been created and performed at UK Microfest, 2005, NIME06, 2006 and Microfest, 2007. As part of the composition procedure, Pure Data (Pd) files containing microtonal data were created and documented using Scala. Another purpose-built tool, developed by Mark Havryliv, was then used to translate Pd files into j2me, a format suitable for java phones (Schiemer and Havryliv, 2005). Prior to performance realisation using multiple phones, performances were emulated and auditioned using Pd files running on a single desktop machine. The second author added enhancements to Scala that allowed tuning data to be exported and documented as text files readable as Csound and Pd files. These enhancements were added in version 1.7 to allow text files produced by Scala to be read as Csound files and further enhanced in version 2.2o to be read as Pd files.

Scala

Scala is cross platform freeware designed

> "for experimentation with musical tunings, such as just intonation scales, equal and historical temperaments, microtonal and macrotonal scales, and non-Western scales" (Op de Coul 2007).

Written in the programming language Ada, Scala has a graphical user interface as well as a command line interface to over 450 functions for scale analysis and manipulation. It has an extensive knowledge base that includes a scale archive containing more than 3400 scales. It recognises more than 1100 musical modes, more than 500 chords and supports more than 400 note naming systems. It offers flexible keyboard mapping, plays scale tones via the soundcard and exports tuning data to a variety of synthesizers with an internal tuning table. It can create MIDI files from a microtonal score, retune existing MIDI files or relay real-time MIDI messages. Its functionality is extensible through the use of command scripts and screen output can be captured to text files.

Scala Command files

Scala commands are sequenced using scripts known as command files. The command file type is .cmd. Screen output is first captured by opening a file, displaying an output string on the screen then closing the file. By observing syntax used in other programs such as Csound and Pd, command files may export tuning information to other programs.

Scala Scale file format

Scala uses an ASCII-based scale file format to store scales. The scale file type is .scl. This format has gradually been adopted as a de facto standard for microtonal scales. Exclamation marks precede comment lines. The first non comment line in a scale file contains a short description of the scale; the second line specifies the scale size; these are followed by a list of pitch values expressed in ratios or cents, as shown in Figure 1.

```
! 05-19.scl
!
5 out of 19-tET
 5
!
 252.63158
 505.26316
 757.89474
 1010.52632
 2/1
```

Figure 1. *Scala scale file format viewed as text file.*

Once a scale file is loaded into scale memory, other commands are used to operate on scale data. Other scale attributes such interval size, and where applicable, historical interval names, are displayed as shown in Figure 2, using SHOW or 'F6'.

```
5 out of 19-tET
  0:     1/1            0.000 unison, perfect prime
  1:   252.632 cents  252.632
  2:   505.263 cents  505.263
  3:   757.895 cents  757.895
  4:  1010.526 cents 1010.526
  5:         2/1         1200.000 octave
```

Figure 2. Scale file viewed using 'Show'

Pd-scale-player.cmd

Pd-scale-player.cmd is a Scala command file that generates tuned Pd files. It is included as part of the library of command files with releases from Scala V2.2o or thereafter.

A user first selects a scale from the scale archive and loads it into scale memory. Running pd-scale-player.cmd will generate a Pd file using tuning data in scale memory.

The command is invoked using Scala's GUI by selecting File, Shift+Alt+@ command file name. Alternatively, in the Scala command line interface, the user may type:

```
@ pd-scale-player.cmd
```

The first line in the scl file is used to document the Pd patch. The second is used as a parameter to generate objects in a Pd patch that match the size of any scale automatically. This allows the command to generate Pd files easily using scales of any size. To create a new scale in Pd, the user selects a new scale in Scala then re-runs pd-scale-player.cmd.

Echo commands

The ECHO command in Scala is used to display data captured in Pd files. Lexical functions were added to the ECHO command at the request of the first author so that external files could be created using data exported directly from scale memory. In Lexical functions were introduced in Scala v1.7 to export pitch information into Csound files. Some of these have also been used to create Pd files.

Lexical functions of ECHO are preceded by % and include in parentheses pitch parameters that convert values in scale memory into text output. Values in parentheses can be a pitch memory if preceded by a $; or they can be a scale degree if preceded by a %; or they can be literal, in which case they are not preceded by any token.

Lexical functions – v1.7

Lexical functions in Scala version 1.7 include:

%cents(pitch)
Gives the cents value.

%factor(pitch)
Gives the linear value.

%hertz(pitch)
Gives the frequency in Hertz relative to the base frequency.

%image(pitch)
Gives the ratio of a rational pitch or cents value of a floating pitch.

%midi(pitch)
Gives the fractional MIDI note number relative to the base frequency.

%name(pitch)
Gives the interval name of a rational pitch.

%octcps(pitch)
Gives the Csound/SAOL oct value relative to the base frequency.

%primes(pitch)
Gives the prime factorisation of a rational pitch.

Lexical functions – v2.2o

Additional lexical functions were introduced in version 2.2o to create Pd files:

%den(pitch)
Gives the denominator of a rational pitch.

%desc(scalenr.)
Gives description belonging to the scale memory.

%listfactor(scalenr.)
Gives the scale as a list of linear factors starting at degree 1. If necessary, start with 1.0 for degree 0.

%n(scalenr.)
Gives the number of notes in a scale memory. Adding and subtracting is also possible, like %n-1(0).

%num(pitch)
Gives the numerator of a rational pitch.

%scl(appendix)
Gives the name of the last scale file loaded, regardless to which scale memory. The appendix may be the name of a scale file or an empty string.

Generating Pd source code in Scala

When a Pd file is read using a standard text editor, Pd objects appear as a line beginning with hash (#). A Scala command file may be used to generate a Pd file if every # is preceded by ECHO followed by a space. This can be achieved using the search and replace function in any standard text editor.

Creating a Pd parent canvas

The first canvas, shows a Pd scale template SCALE.PD created using Scala's FILE and CLOSE command. All ECHO commands between FILE and CLOSE result in output that appears on the Pd parent canvas. The lexical function %desc is used to describe the scale. Additional Pd patches are placed on the parent canvas between the second ECHO and CLOSE.

```
FILE SCALE.PD
!
! Main Canvas
!
! Below this point observe
! PD syntax when using ECHO
!
ECHO #N canvas 24 5 750 350 10;
ECHO #X text 5 16 %desc(0);
!
! insert other Pd objects here
!
CLOSE
```

Example 1

In this example, the first ECHO #N creates a new canvas, in this case the parent canvas. The first four variables define the area. The 1st and 2nd are the vertical (24) and horizontal (5) coordinates at the top left; the 3rd and 4th are the vertical (750) and horizontal (350) coordinates at the bottom right. The 5th variable is the default font size (10) for the parent canvas and all embedded canvases.

The second ECHO command writes a comment in Pd describing the scale last loaded into scale memory 0. This is displayed at a position defined by horizontal (5) and vertical (16) coordinates. The scale description is supplied by the scale archive.

Pd – tuning data

As the primary focus of the project has been just intonation, tuning has been implemented using linear factors formed when numerators are divided by denominators. Each note in the scale is tuned by multiplying its linear factor by the base frequency. Though linear factors are normally associated with just intonation, Scala allows linear factors to be used to express non-just musical intervals.

Figure 3 shows the Pd sub-patch embedded in pd-scale-player.cmd which uses linear factors to tune oscillators. The oscillators are played by the note keyboard that is played by an algorithmically generated sequence. The interface used to play the file allows sequences to be transposed by octaves or using modes of the microtonal scale. The interface is shown in Figure 4.

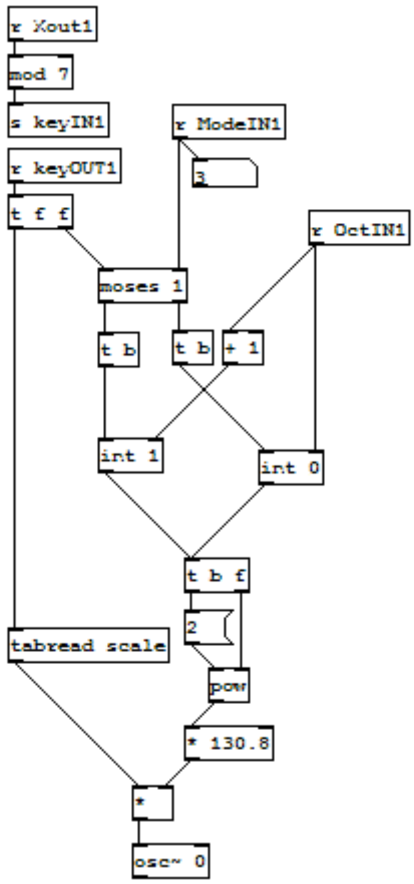


Figure 3. Oscillator in Pd tuned using linear factors

Example 2 shows the embedded object used to display the contents of a tuning array. The array is loaded with a scale and named using the lexical function %scl(). The first ECHO command forms the canvas of the array object at a point specified by horizontal (45) and vertical (68) coordinates. The canvas is then overlaid by the graph object in the last ECHO command.

```
! Canvas 2 - display scale
!
ECHO #X obj 45 68 cnv 15 254 99 empty
empty %scl() -40 -60 0 16 -212343 -
258699 0;
!
! Create array of linear factors
!
ECHO #N canvas 0 0 450 168 graph1 0;
ECHO #X array scale %n+1(0) float 1;
ECHO #A 0 1 %listfactor(0) \:;
ECHO #X coords 0 2 %n(0) 1 256 100 1;
!
ECHO #X restore 44 67 graph;
```

Example 2

When it is open, the canvas has a default hot spot size of 15, is 254 units wide, 99 units high. It has three symbolic properties, two of which are undefined (i.e. empty). The 3rd is the canvas label containing the name of the scale last loaded into scale memory from the scale archive. It is defined using the lexical function %scl().

The next two variables define the coordinates of the canvas label relative to the canvas, (-40 horizontal, -60 vertical); the 3rd variable selects the font (0 = Courier New); the 4th variable selects font size (16); the 5th and 6th variables define background and font colour (-212343 = grey, -258699 = red). The final variable 0 is unused on the parent canvas.

The second and third ECHO command creates the array for storing and displaying tuning data. The 2nd ECHO defines the visible geometry of the array. While the canvas is open, the 1st and 2nd variables are the vertical (0) and horizontal (0) coordinates at the top left; the 3rd and 4th are the vertical (450) and horizontal (168) coordinates at the bottom right. The 3rd ECHO creates an array of floating point numbers called 'scale'. Array size is defined by the lexical function %n+1(0) and refers to the size, plus one, of scale memory 0. By adding one, both the unison and octave are each aligned with opposite boundaries of the graph.

The 4th ECHO loads tuning values from scale memory 0 into the array using the lexical function %listfactor(0) which presents tuning data as a list of linear factors. The first factor is stored at location 0, the first location in the array. Because unison is not included in the list of linear factors created by Scala, 1, the value for the unison must be inserted before the list of factors. The list is terminated with backslash (\) followed by colon (:) to comply with message syntax used in Pd.

The 5th ECHO defines the range of values used to display tuning data within the geometry of the scale canvas. The first two variables are values at the horizontal (0) and vertical (2) coordinates in the upper left corner of the display; the next two are the values at the horizontal (%n(0)) and vertical (1) coordinates in the lower right corner of the display. The lexical function %n(0) divides the horizontal domain into a number of discrete points equal to the size of the scale in scale memory 0. The vertical domain covers the range of linear factors between the unison (1) and the octave (2) because all just ratios, by convention, are expressed within the range of one octave. The remaining three variables affect the size of the graph object. The 5th variable (256) determines the width of the graph, the 6th variable (100) determines its height and the 7th variable (1) determines whether the graph object is open or closed.

Pd – tuning documentation

Exporting tuning documentation from Scala to Pd presented a special challenge. This was largely due to the way Pd processes strings and represents numbers in floating point. It is a simple matter for Scala to create a text file that can be used to display scale specifications in other programs. It is not a trivial matter to display this clearly in Pd as text is interpreted in a way that does not always allow Pd to display a text file literally.

In Pd some words are reserved for use as part of a token. For example, a comment begins with the token '#X text'. This is followed by two variables describing the vertical and horizontal coordinates of the character string displayed. Finally, this is followed by the string itself.

Example 2 displays the specifications of any scale exported from Scala to Pd. These include not only the name of the scale, displayed using %scl(), but data about each pitch such as intervals formed relative to the unison, and the historical names of intervals where these exist. Intervals formed are expressed as tuning ratios or cents and displays using the lexical function %image (pitch) which automatically expresses just intervals as fractions and non-just intervals as cents. Scala also automatically identifies historical interval names and these are displayed using %name(pitch). In addition to these, new lexical functions were created to display numerators (%num()) and denominators (%den()).

The problem of documenting scale information in Pd has been addressed by using Scala scale memory as general purpose memory. This allowed us to create and store sets of vertical and horizontal coordinates for creating a coherent display of Pd comments. Working with Scala scale memory in this way was somewhat similar to working with general purpose registers when programming in assembler. Scala uses commands like COPY, MOVE, CLEAR, ADD, etc. which operate on scale memories. Unlike a general purpose register which is a single unit of stored data, a scale memory is a complex array of data on which many operations may be performed iteratively whenever a Scala command is performed.

Tuning documentation is displayed as rows of pitches arranged in four columns. Iteration is used to display the notes of the scale in successive rows. The HARMONIC command was used to produce a number sequence. Normally this command is used to create harmonic scales, but was used here as a way to number the lines in a text display of tuning information. The program code in example 3 is an embedded command file called 'ordinal.cmd'. When this is run later (in example 4), the HARMONIC command will produce a number sequence between 1 and the number of notes in any scale loaded into scale memory 8. The sequence created is stored in scale memory 0. The size of the scale in scale memory 8 represents the number of lines in the display.

```
! Generate PD Scale template
!
FILE ordinal.cmd
ECHO ! ordinal.cmd
ECHO ! This command is a template gen-
erated by PD-Scale-Player.cmd
ECHO !
ECHO HARMONIC 1 %n(8)
CLOSE
```

Example 3

Four display columns are aligned using horizontal coordinates that are entered using the INPUT/LINE command shown in example 4. The four variables 40 80 200 and 305 in the following line, are the coordinates of the four columns; these are stored by default in scale memory 0, and then saved in scale memory 5.

```
! Display coordinates for 4 columns
!
INPUT/LINE
40 80 200 305
COPY 0 5
CLEAR 0
```

Example 4

Example 5 runs 'ordinal.cmd', the command file generated in example 5. INSERT is used to increase scale size by 1. The extra number allows the unison to be listed as ordinal 0. This is a 'housekeeping' measure to display the first note correctly. Ordinals are then saved in scale memory 1.

It is then necessary to space each row so that characters are displayed in 10 point. To do this, the first line must start at 0, followed by the 2nd at 11, the 3rd at 21, the 4th at 33, and so forth. The ITERATE command is used to achieve this. The command modifies ordinals to create the sequence 0, 11, 22, 33, etc. This produces lines spaced 11 points apart. Y-coordinates are then copied into four separate scale memories, one for each column. This is necessary because the interative process used to generate each column is destructive.

```
! Generate ordinals
!
@ordinal.cmd
INSERT 1 1
COPY 0 7
SHOW/LINE 0
ECHO
!
COPY 0 1
CLEAR 0
!
! Define height of each row
! vertical spacing - 11 points
!
ITERATE/scale "INSERT 1 11" 1
ADD 1
ECHO Y-coordinates
SHOW/LINE 0
!
! Store column co-ordinates in
! scale memories 1 2 3 and 4
!
COPY 0 1
COPY 0 2
COPY 0 3
COPY 0 4
```

Example 5

Finally, ITERATE is used to display the salient features of any scale as comments in Pd using coordinates generated in Examples 4 and 5. Comments are aligned in four columns. In the 1st, note order is displayed; in the 2nd, the image (just ratio or cents); in the 3rd, linear factors; and in the 4th, historical interval names, where these exist.

All rows except the 1st are generated iteratively. This was necessary in order to represent the first scale degree as ordinal 0. All numbers in scale memories are represented as fractions; coordinates are stored as a numerator over the denominator 1. This makes it necessary to read only the numerator and write it for Pd to display. The lexical function %num(pitch) was used for this purpose. All lexical functions shown in Example 6, specify pitch using scale-degree followed by scale-memory.

Conclusion

Exporting tuning data from Scala to Pd became complicated by their different idiosyncrasies. Yet despite these complications, well documented Pd patches serve to identify the landmarks composers need for microtonal exploration. Command files in Scala also give control over the layout in Pd as text editing makes it possible to align Pd objects precisely. Moreover, Scala command files provided an extremely stable development environment for prototyping microtonal mobile phone applications using Pd. The techniques described have been applied for Csound source files as well as Pd files. It is possible also to generate microtonal source files for other computer music languages such as MaxMSP and Supercollider, or using Scala to create microtonally tuned musical applications in java.

Acknowledgements

The Pocket Gamelan project was funded by an ARC Discovery Project 2003 – 2005.

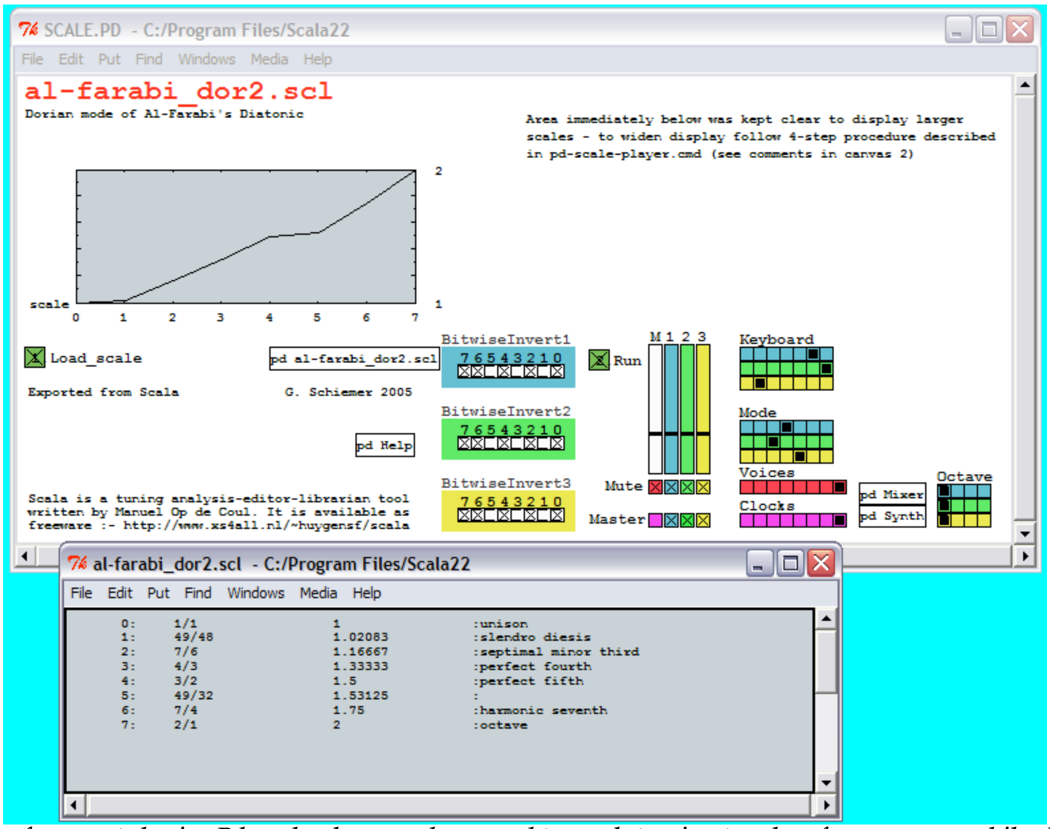


Figure 4. Interface created using Pd-scale-player.cmdwas used to emulate microtonal performance on mobile phone

```
! Display scale specifications
!
! First line of each column is read out separately
! All other lines of each are read out iteratively
!
! column 1 - Ordinals
!             x-coord 1  y-coord 1
ECHO #X text %num(%1%5) %num(%i%1) 0:;
ITERATE/SCALE "echo #X text %num(%1%5) %num(%i%1) %num(%i%7):;" 8
!
! column 2 - JI ratio (or cents)
!             x-coord 2  y-coord 1
ECHO #X text %num(%2%5) %num(%i%2) %image(%0%0);
ITERATE/SCALE "echo #X text %num(%2%5) %num(%i%2) %image(%i%0);" 8
!
! column 3 - Linear factors
!             x-coord 3  y-coord 1
ECHO #X text %num(%3%5) %num(%i%3) %factor(%0%0);
ITERATE/SCALE "echo #X text %num(%3%5) %num(%i%3) %factor(%i%0);" 8
!
! column 4 - Historical names
!             x-coord 4  y-coord 1
ECHO #X text %num(%4%5) %num(%i%4) :%name(%0%0);
ITERATE/SCALE "echo #X text %num(%4%5) %num(%i%4) :%name(%i%0);" 8
ECHO #X restore 190 200 pd %scl();

Example 6
```

References

Op de Coul, M. 2007 Scala Home Page http://www.xs4all.nl/~huygensf/scala/

Schiemer, G., and Havryliv, M. 2006 "Pocket Gamelan: tuneable trajectories for flying sources in *Mandala 3* and *Mandala 4*" Proceedings of the 2006 International Conference on New Interfaces for Musical Expression (NIME06), Paris

Schiemer, G., and Havryliv, M. 2005 "Pocket Gamelan: a Pure Data interface for mobile phones" Proceedings of the 2005 International Conference on New Interfaces for Musical Expression (NIME05), Vancouver, BC, Canada.

Chatter and Listening: A wireless multi-node sound sculpture

Alex Thorogood
The Australian National University
Centre for New Media Arts
Faculty of Arts
Peter Karmal Building #121
Childers St. Acton
ACT 0200
Australia
alex@greenmeat.net

Abstract

'Chatter and Listening' is a project that creates a number of re-locatable art objects with sonic attributes. These objects act as single entities; they can create decisions, observe their environment and operate alone. They also act in the narrative of a colony where each entity is guided in behaviour by the state of the colony and its surrounding environment. The project aims to create a multi-node spatial sound sculpture that operates wirelessly over a range of distances to create a distributed sonic art object.

'Chatter and Listening' uses technologies including embedded micro controller systems and wireless communications combined with other artistic practices to realise an intermedia artwork.

This paper describes the concepts, technologies and a composition philosophy behind the work. It also looks at the behavioural model of the Australian Magpie that is used in creating objects that communicate and generate sound over a wide spatial location.

Introduction

Sound sculpture is the combination of sonic and sculptural art practices. It draws stimuli from fields that include acoustic design, architecture, electronics, design and fabrication of physical objects, intermedia practices, music and sound design. The output of this practice can take many different forms ranging from auditory and physical embodiments to simulations inside a computer system. Sound sculpture re-defines itself with every work and re-codifies the space in which it resides.

Acoustic attributes that are inherent to sound sculpture can express narratives and abstractions that transmute the physical intersections of the object with the space it inhabits. This paper describes a sounding work from the concept up. (The concepts are important to establish as they define the direction that the implementations of the physical materialization, wireless communications and the sonic composition are to take.)
'Chatter and Listening' explores narratives in a colony of sounding objects by employing the behavioral model of a natural system. To date the majority of musical and sound compositions that utilize behavioural models of living dynamic systems have comprised of a sonification of single agent movements in space through time in the meta-framework of a colony. Common models employed include schools of fish and flocks of birds. These works produce a composition when a number of the agents interact through a set of defined behaviours, sonifying the school or flocks positioning in space through time (Blackwell 2002).

Other dynamic systems are used in music/sound creation. These include evolutionary algorithms that use a breeding and selection criteria to produce musical information (Todd 1999). A number of agents are encouraged to interact by breeding with other agents that are deemed fit by a set of programmed criteria. The breeding produces a number of offspring that carry traits of both parents by genetic cross over, and mutations are introduced to liven up the genetic pool.

The artificial ecosystem (Conrad 1985) is an environment where virtual organisms move about interacting with each other and the changing environment they inhabit. As in the real world this environment and its inhabitants change dynamically over time. For example, depending on the time of the day or generation behavioural changes can be seen; different generations express individual characteristics (traits) evolving from the parents; roles of behaviour change from predator to prey. The result of this implementation in a sonic artwork is a sonification of the organisms state parameters into an audible work that provides intrinsic results (Dorin 2006).

One undertaking of this project is to map a behavioural model of a *single* species of animal to provide a successful composition paradigm for a sound space. The exhibited behaviours will be analogous to the real world interactions that the animal displays. Stimulations of the inhabited environment, other individuals in the group and other species all provide influence to the behaviours. These behaviours are employed are implemented at a program level on embedded technologies which are implanted in the distributed sculptural nodes.

The interaction the colony nodes have with one another is two fold. Communications of an individual state and intentions (behaviours), and the state of the colony at large, is carried out on a 2.4GHz wireless broadband network. Sonic materializations of the individual and colony at time are actuated by the behaviours and communications. The sonic events can be received, depending on an individual's location, and realised as another communication expression.

Related works

Electronic Sonic

Early use of electronics to create sonic output in sound sculpture can be seen with Percy Grainger's 1952 work *'The kangaroo Pouch Free Music Machine'.* In this work were eight oscillators that were controlled by a largish sheet of brown paper with pitch control graphs in different colours (score) that was fed from one roller (the "Feeder") to another (the "Eater") driven by a motor. The sonic output form this machine materializes as a wild sweep of the oscillators gliding over the intervals set out on the graphs.

Sounding Objects

The form of an object can facilitate the perception of a life force. This can be seen with sculptural practices that aim to represent the embodiment of living forms, either in traditional or conceptual narratives as in the work of Patricia Piccinini and her project *'we are family'* 2003[1]. The essence of a living thing is inherently multi dimensional; form being one dimension, so projecting different or multiple dimensions on an object has the potential to give varying perceptions of a life force in that object.

'*Cyber Squeeks*', Ken Rinaldo 1994, imitates intelligence of machine life forms by creating a number of organic looking creatures that create sound based on input from various activations, for example viewer interaction. The response of the objects is a tuneless welter, signifying a life like response to stimulation.

Location location

When observing or interacting with a sounding object the viewer is engaged in a multi-sensory experience, that requires an amount of listening as well as viewing the physical embodiment. The innate physicality of sculpture allows it to be viewed from many angles and the overall observation on the object and space is sometimes an important contributor to the work.

A sound work (since the advent of tape fetishism) is viewed primarily in a two-dimensional sound field i.e. Left, Right. However in discussing the qualities of a sculptural object the listening of the sound field emanating takes on a multifarious nature, where the viewers position dictates the immediate perception of how and when it is heard redefining the work in space.

The work '*finitorium*' *1993,* by Chris Ulbrick explores the subjectivity of location in relation to an object and it's spatial situation. Sound is mapped in the installation space by positional information of the viewers at time that is relayed to a computer by sensors. The effect is of changing the spatial relationship of the voice and abstract audio elements in the vertical and horizontal plane, which is different every time because the number of people viewing the work and their distribution will always change.[2]

[1] http://patriciapiccinini.net/ cited 29/03/07

Concepts

In this section a number of key concepts will be defined that will relate to the dimensions of the project at an abstract level. The ideas that these concepts bring to the table will underpin the directions of the other components in this paper.

Natural Systems

Individual co-located agents of species of organism can be seen in a host of naturally occurring co-operative systems as colonies. A colony is a complex system of individuals interacting with each other and their environment (Booker, 2004). Colonies have a hierarchy with at least one order of individual higher than another and they appear to act co-operatively and purposefully.

Colony behaviour emerges from auditory cues not necessarily within our range of hearing, visual displays, contact between individuals and environmental pressures.

Cognitive perception

When functioning correctly individuals have an awareness of the representation of the environment that they exist within. This is a multimodal experience built from unimodal cortical responses including visual, auditory and somatosensory association. A perceptual/cognitive system extracts dimensional information from the stimulus array (Lawrence, 1981). If there is a change for some reason in the environment then those stimuli will be drawn to our attention and our perception of the multimodal locality will adapt. (Downar, 2002)

Core Understanding

Behaviours of individuals in a reasoning society arise from streams of information from the immediate environment, and perceptions in the state of the self in relation to others in the group. When same society individuals communicate there is an empathy that is constructed from historical memory and the perception of a future self in relation to that society.

[2] Some observations of the works in section 2 were derived from Ross Bandt's book and accompanying CD Sound Sculpture: Intersections In Sound & Sculpture In Australian Artworks

Behavior as Composition

On a theoretical level *'Chatter and Listening'* looks at using the behavioral model of living systems for electroacoustic music composition. This paper argues that musical and sounding works are based on different instigators of behaviours. An example of this is with a scored composition, where an instruction set telling a performer how to behave with their instrument produces the desired outcome of a composer. In sound designs events are more directly related to occurring behaviours of real world materializations, where the time and space of the sound event is placed according to some acknowledged auditory 'what' and 'where' subsystem (Kubovy 2000). An example of this is with a familiar bird vocalising in a tree. You are familiar with the sounds the creature makes and you know of its relative position from visuomotor responses ('where'). Your auditory system acknowledges the edges of the sound, where the frequencies and amplitudes begin and end 'what', defining an auditory object. When a number of these objects develop in structure and interact based on dynamic and purposeful behaviours the product displays new meanings.

Adaptive Behaviour

In the article Designing and Understanding Adaptive Group Behavior Maja (1995) introduces the concept of 'basis behaviors' for the building blocks of synthesized group behaviors. In *'Chatter and Listening'* basis behaviors are used to operate each nodes behaviors, thereby forming the composition. A set of basis behaviors has criteria that make for maximum efficiency based on relevance to obtaining goals or helping obtain desired goals and how each behavior relates to another. Each basis behavior cannot be achieved with any other of the other basis behavior in that set or reduced to them. Maja (1995) gave autonomous control to 20 mobile robots using basis behaviors. With the implemented basis behaviors adaptive agent control, social interaction and learning was achieved in the robots.

When designing basis behaviors a number of techniques emerge from the article for finding an effective and efficient set. One suggested approach used in *'Chatter and Listening'* includes: firstly finding a naturally occurring organism with a set of behaviors that may appeal to the wanted outcomes as a good starting point for finding a package of relevant and efficient base behaviors. This can be the case because the species to be looked at has refined the behaviors through many generations of evolution for maximum gain and efficiency. Also of use may be putting a list of behaviors into an enumerated table and evaluating them on the relevance they have to the desired goals and the relation they have to one another. From this process a number of quality base behaviors should emerge.

Behaviours of Natural Living Systems

Various species that operate in natural dynamic systems including wasps, ants, bees and birds have been observed and their behaviours mapped out. This mapping has given researchers new ways of approaching and efficiently managing modern complex systems. Although the individuals operate with an amount of simplicity the colony structure provides a distributed and complex system that is capable of definite tasks, which exceed the capabilities of a single individual.

The observation of birds in nature is a fascinating insight into the behaviours of living complex systems. As a higher-level approach *'Chatter and Listening'* includes the practice of natural system observations. Craig Reynolds observed the activities of fish and birds in his seminal work on flocking simulations (Reynolds C. 1987). Flocking, in a computer science term, is a simulation of group behavior primarily associated with the visual domain; examples include a flock of birds or a school of fish. It is a type of optimization in creating complex systems of sociological/biological behaviors. A number of sound based works have used these models based on dynamic natural systems as a type of musical score (Klein, 2002) (Spector 2003). The movements of the individuals or the flock in space are mapped to different musical pitches and when put into the context of the flock over a time it will produce a complex and evolving piece of music, a sonification of the flock.

Species Australian Magpie

The Australian Magpie is a medium sized black and white songbird, found across the Australian continent. There are several subspecies scattered across the country, but there are definite behavioural characteristics that they have in common. The species behavioural model has been chosen for the project because of the personal affinity and positive observations the author has had for the animal. Since European settlement in Australia the Magpie has shown to be highly adaptable and sociable, where they are able to sustain populations with the encroachment of human development, often interacting with people.

Flocks of Magpies tend to be no more than seven at one time, and these small groups will most often be seen using their skills to hunt at ground level, taking to the higher levels of the trees only for roosting and breeding. When feeding Magpies use a method of extraction foraging that involves location of prey by listening for the sound of vibrations made by potential food makes and not with visual cues (Simpson 1993). After extraction the Magpie will often hide and later retrieve food in a complex manner using a memory skill called caching.

Magpies have a high level of individuality and have some unique behaviour such as social play and cooperative behaviour with same and different species. Below (Table 1) is a 'basis behaviour' table (see beginning of section 5) that outlines the appeal the Magpie exhibits for behaviour set implementation.[1]

Behaviour name	Behaviour function	Skills
Flocking	Safety in numbers Social permutations	Entity Recognition.
Extraction Foraging	Food gathering	Listening. Location identification.
Caching	Resource control	Memory
Cohesion	Effective teamwork	Entity Recognition. Task Reasoning. Communication.
Agonistic	Nest protection	Aggressive response. Vocalisation.
Social Play	Learning to handle prey Social bonding Education in food, risks, predators, vocalisation	Pseudo aggressive response. Communication. Vocalisation.
Alarm calling	Flock protection	Entity Recognition. Vocalisation.
Courtship	Female calls to male for procreation	Sexual awareness. Vocalisation.

Table 1. *Basis behavior table of the Austalian Magpie*

Communications

Each of the sounding objects in *a 'Chatter and Listening'* colony makes decisions based on the behavior criteria, a number of activities then occur on the 2.4GHz data band and the audio level. A telemetry device contained within the vessel transmits and receives communication of machines therefore relaying their behavior. This data communication is the means that all the machines in the colony use to configure the dynamics of the system at any given time.

The sounding element of the work is the auditory embodiment of the behaviours programmed into the colony. Just as a natural living system that communicates with a sonic language has an impact to the external world the colony will convey itself at an audio rate.

Songbirds

Animals communicate in a variety of ways such as auditory communication, visual displays, touch and smell, and one of the most efficient types of communication for a bird species is through vocalisation (song). Bird song is able to be transmitted over long distances very quickly, then to be received and decoded by another of that family group or another species. This type of communication where an entity transmits and all other entities can hear is a broadcast communication. The song defines an intention and is purposeful. Two examples from the basis behaviour table that are different but both requiring vocalisation are alarm calling and social play. In alarm calling one magpie will view a threat and make a decision to raise an alarm start producing the definite vocal signature that alerts the other magpies of that family to take action. This is unlike in a social play situation where two or more magpies will start playfully squabbling with each other, in a physical sense. The vocalisations are to acknowledge and demonstrate that the current activity is not aggressive.

Wireless Sensor Networks

Development of technologies and ontology's in wireless sensor networks has created advancements in this field recently. These networks are deployed for various applications like environmental monitoring of microclimates, patient and drug monitoring in hospitals and environment control in offices (Akyildiz 2002).

A network of this type comprises of *n* number of wireless sensor nodes in an environment, not placed at any definite position, and because of this the means by which the connections the nodes have to relay any piece of information from *a* to *b* becomes important, possibly complex as nodes may drop in and out of a network for a number of reasons at any given time. The way in which the connections are configured for communication is called the network topology.

Wireless peer to peer sensor networks use a topology, generally a Mesh topology (see figure 1), that encourages low transmission distances and more efficient network communication by allowing nodes to only communicate with their nearest neighbour[2]. In this type of wireless network topology each node is considered the same as the other, and with these networks distributed over possibly large distances messages are relayed with an understanding that each node will route another's message. The Mesh allows distributed nodes to be

[1] Sets of the behaviour of the Australian Magpies in section 5.2 were derived from "Australian Magpie, Biological and Behaviour of an Unusual Songbird", Gisela Kaplan 2004

[2] A summary on Mesh topologies can be viewed at http://en.wikipedia.org/wiki/Wireless_mesh_network. cited 19/03/2007

connected along a number of different paths to route data, and because of this the communications link is self-configurable as nodes drop out or new ones join in. (Akyildiz 2005)

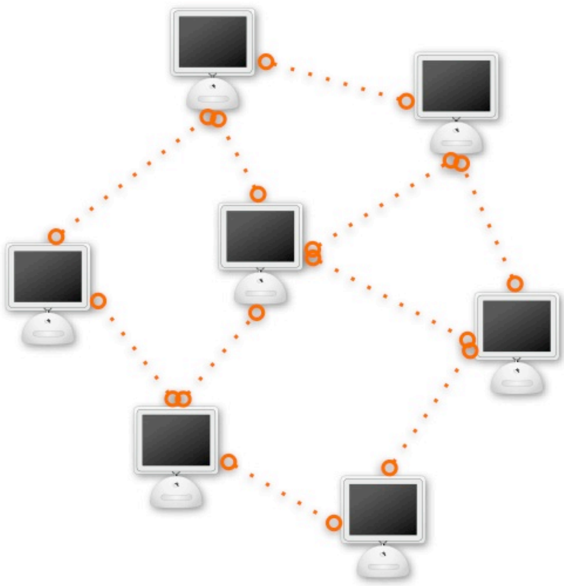


Figure 1. *Visualization of a Mesh peer-to-peer network.*

Hardware technologies

Each machine is based on a modular design, which allows reconfiguration for installations and less expense in replacement if only single module goes bust in the system. The modules that are at the core of the system are:

- micro controller unit
- wireless transceiver
- MP3 player/recorder
- Sensor array

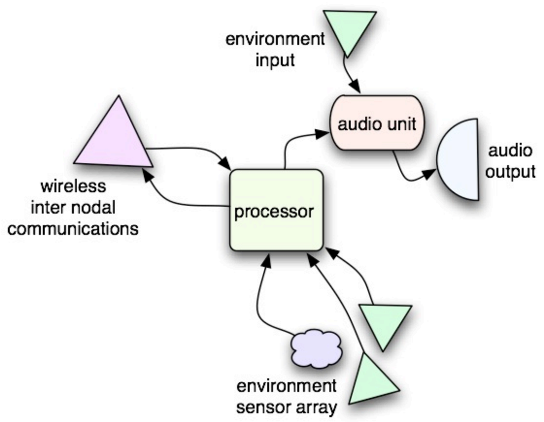


Figure 2. *Hardware technologies abstract diagram.*

The wireless communications requires a transceiver that has the ability to communicate state information of the unit and other units in real time. For this a MaxStream XBee Pro™ module that transmits at 2.4GHz with 60mW power [1]. This is a modern device that supports complex topologies, addressing and error handling. A sensor array relays the condition of the environment to the unit and colony on demand and consists of microphones to relay sound pressure levels, a light dependent resistor to gauge the light level of the environment and a temperature sensor.

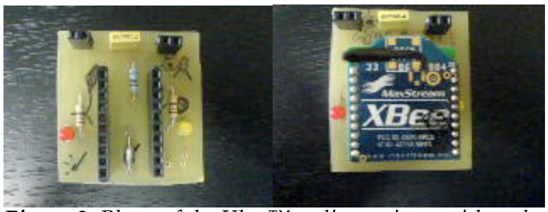


Figure 3. *Photo of the Xbee™ radio carriage, with and without the module installed.*

The audio unit defines the sonic attributes of the machine and is able to randomly access a database of sound designs to be played back. Recording functionality is also included to take segments of the sonic environment to play back at a latter time. A number of lower level MP3 units are available from different suppliers and have become more prevalent since the introduction of low cost encoder/decoder IC technologies. The lower level MP3 units that were researched did not fulfil all the necessary requirements, i.e. they were either too expensive or did not have all the desired features.

To overcome this an off the shelf MP3 player/recorder was chosen. It has the requirements and is also simple enough electromechanically to install a new switching mechanism, and in functionality to manipulate electronically by the micro controller.

Firstly the buttons are removed and hook-up wire attached to the necessary pads.

Figure 4. *Photo of the opened MP3 unit with hook up wire connected to the button pads.*

Then a transistor array is used as a switching block for all the controls.

[1] Information for the XBee Pro from MaxStream can be found at http://www.maxstream.net/products/xbee/xbee-pro-oem-rf-module-zigbee.php. cited 19/03/2007

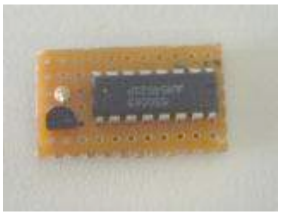


Figure 5. *Photo of the transistor array used as the new switching mechanism. The small transistor/resistor pair on the left is for the play/stop function as it runs on a different power rail.*

And finally all the wires and new circuit board are loaded into the back of the unit and the case returned.

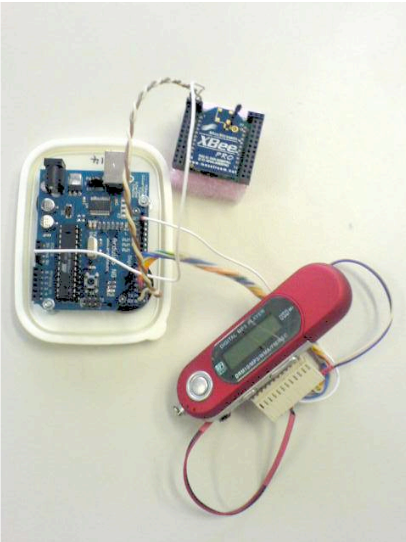


Figure 6. *Photo of the core components for the hardware system. Note the new connector block on the bottom of the MP3 player/recorder.*

Of course a speaker and embedded amplifier will also be required to externalise the audio. The micro controller performs all the calculations from the sensors and wireless input to make decisions on the state of the entity and what sonic attributes it will have at time.

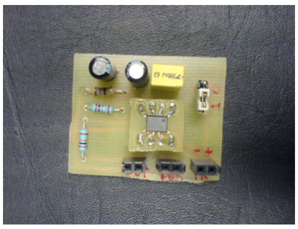


Figure 7. *Photo of the 3W LM4991MA amplifier circuit.*

Software abstractions

The implemented behavior systems and hardware technologies are at the service of a type of awareness the nodes will process. A node will have awareness of space through the environment sensors relaying information about the space they inhabit. An awareness of the group and their place within the group will exist by the data level information telemetries. This awareness is configured and realized by the software.

Creating self-awareness in communication nodes becomes possible when the data they are transferring as well as their own and other nodes capabilities are 'understood' (Wang 2003). The capabilities that each node possesses for purposes of this project are the same as any other node that will be communicated with. A decision unit from time to time will request a group check, which sends a request from the radio that all nodes in the group are inclined to respond. The responses are cross-referenced with an array of the last known members in the group; if there are any differences then the array is updated. This means if a node drops out or is no longer able to communicate then the topology reconfigures itself and capability of all nodes is known as good.

To make sense of the data received from the radio and environment sensors, an understanding unit distributes the data to the destination requesting it or passes it to the decision unit to process. For example let's say the temperature rises for some reason, referring to figure 8, the environment monitor notices a difference and compares it to a database of the past temperatures, finding the difference then requesting the understanding unit to do something with the data. The understanding unit takes the difference and (in this case) interprets it as hotter, understanding then requests the decision unit to do something. The decision unit will then request the state of the machine to update accordingly and radio the ID of itself and the new state to the rest of the colony, unless for some reason it decides not to.

In each behavioral state there is a set of sonic attributes that relate to that state. The state is set by the decision unit and is based on streams of information from the immediate environment and the state of the self in relation to other nodes in the colony. Separate from the set state is the state dynamics unit that takes input from all meaningful sources to create dynamics to the sonic motif. Requests are made to gather information on the current state of the machine, environmental data, the state of other colony members and request are made to other nodes in the colony to understand what their particular state dynamics may be. An example of this latter point could be that the state dynamics unit wants to create a harmonic relation within the colony after checking the state of other colony nodes and environmental data. A request would be sent out the radio to the state dynamics unit in other nodes in the colony to exhibit a part of a harmony. The nodes would then respond if they would comply or not. This unit communicates with the MP3 player/recorder and commands it's functions.

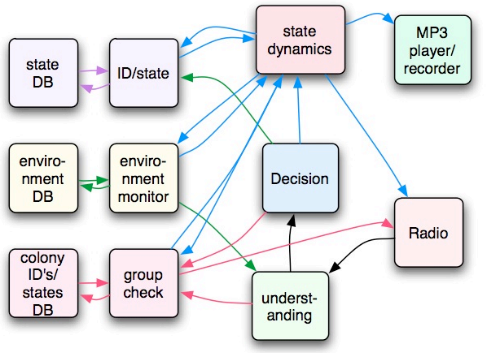


Figure 8. *Ontology of software abstraction examples.*

Conclusion and future work

Using a behavioural model for a composition paradigm embedded in a colony of sculptural forms leads to diverse applications for use in a distributed art object. This paper has mapped out a range of concepts and a theoretical realization of these through dynamic systems of behaviour. Furthermore it describes the Australian Magpie as a single behavioural model to adopt for the use in a sound sculpture work.

The means in which objects exhibiting the adopted behavioural model are able to communicate leads to exciting avenues in wireless communications topologies on the level of wireless sensor networks. An application of a wireless topology as well as the hardware and software ontologys of the theoretical components has been explained for the use in a distributed artwork.

The full realization of a body of work that adopts the processes outlined in this paper is to be carried out. The final work is to be produced for gallery exhibition and installation in space. A robust package is being developed that will cope with the handling and other stresses of these types of environments.

Once a package has been designed then it will be portable to other projects and other artists to use. The range as well as the type of communication and the reconfigurable nature of the technology means that the scale of the project that implements what is outlined in this paper has the ability to encompass an area the size of the corner of a room to networking over the area of a city.

Acknowledgments

I would like to thank Alistair Riddell for his time and support from the conception of this project, Jocelyn Smith in playing with the coloured cardboard state machines with me, and the Centre for New Media Arts at the Australian National University for allowing the time and resources to make this work possible.

References

Akyildiz, I. Su, W. Sankarasubramaniam, Y and Cayirci, E. 2002. “Wireless Sensor Networks: a survey”, *Computer Networks*, vol 38:4, pp 393-422.

Akyildiz, I. Wang, X and Wang, W. 2005. “Wireless Mesh Networks: a survey”. *Computer Networks*, volume 47:4, pp 445-487-422.

Bandt, Ros. 2001“Sound Sculpture: Intersections In Sound & Sculpture In Australian Artworks”. *Craftsman House.*

Blackwell, T. and Bentley, P. 2002. “Improvised music with swarms”. *Proceedings of the 2002 Congress on Evolutionary Computation 2002.* vol 2, p 1462-67.

Booker, Rebecca. 2004. “What Ants and Bees Tell Us about Social Network Theory”. *AAAS Dialogue on Science, Ethics, and Religion January 2004.*

Conrad, M. and Strizich, M. 1985. “A Computational Model of an Evolving Ecosystem”. *Biosystems 17.* p 245 - 258.

Dorin A. 2006. “The Sonic Artificial Ecosystem”. *Proceedings of the Australasian Computer Music Conference 2006.* p 32 – 37.

Downar, Jonathan. Crawley, Adrian P. Mikulis, David J. and Davis, Karen D. 2000 “A multimodal cortical network for the detection of changes in the sensory environment”. *Nature Neuroscience*. vol 3, p 277 – 283.

Kaplan, Gisela. 2004. “Australian Magpie, Biological and Behaviour of an Unusual Songbird”. *CSIRO publishing.*

Kubovy M. and Valkenberg, Van. *2000* “Auditory and visual objects”. *Cognition.* vol 80, p 97-126.

Lawrence M. Ward James. and Russell, A. 1981 “Cognitive Set and the Perception of Place”. *Environment and Behavior*. vol 13, p 610.

Maja, Mataric. 1995 “Designing and Understanding Adaptive Group Behavior”. *Adaptive Behavior.* vol 4, p 51.

Reynolds, Craig. 1987. “Flocks, Herds, and Schools: A Distributed Behavioural Model”. *Computer Graphics*, vol 21:4.

Simpson, K. Day, N. and Trusler, P. 1993 “Field Guide to the Birds of Australia ‘Simpson and Day’”. *Penguin for Lifetime Distributors*. New South Wales, Australia.

Spector, L. Klein J. Perry, C. and Feinstein, M. 2003 “Emergence of Collective Behavior in Evolving Populations of Flying Agents” *In Proceedings of the Genetic and Evolutionary Computation Conference*. Springer-Verlag. p. 61-73.

Todd, P.M. and Werner, G.M. 1999. “Frankensteinian Methods for Evolutionary Music Composition” *Musical networks : parallel distributed percetion and performance*. Cambridge, MA MIT Press. xv, 385.

Wang, J. Brady, D. Baclawski, K. Kokar, M. and L Lechowicz. 2003. “The Use of Ontologies for the Self-Awareness of the Communication Nodes”. *In Proceedings of the Software Defined Radio Technical Conference*.

Sebastian Tomczak
Masters Student, Electronic Music Unit,
University of Adelaide
North Terrace
Adelaide, 5000
Australia
sebastian.tomczak
@student.adelaide.edu.au

Handheld Console Comparisons: Lateral Consumer Machines as Musical Instruments

Abstract

The genre of micromusic, also known as chiptune, highlights the practice of utilising obsolete videogame consoles in the composition and production of music to an extent that is beyond a novelty or simply a passing fad. This is often realised via relatively old hardware that is not designed for the composer or musician, executing modern-day music performance or composition software. These videogame consoles have thus become "lateral machines" – devices that have outlived their original purpose and have been reincarnated as musical instruments and composition environments.

Prominent examples of hardware / software combinations used in chiptune composition and performance include packages such as the Midines *(Kann, 2005) and the* SynthCart. *The* Midines *is a hardware-based data interface used to control the Nintendo Entertainment System's 2A03 sound chip via MIDI. The SynthCart allows an Atari 2600 user to control the console's various unique sounds in real time via a set of keypad input controllers.*

The GameBoy music community is active and vibrant within the micromusic genre. GameBoy hardware is a topic that has been discussed a number of times. In particular, the question of which GameBoy model has 'the best' sound is a common theme. Abstract, ill-defined terms such as 'warmth', 'retro' and 'feel' are often used in the context of this discussion. Words that are less difficult to define such as 'noise' are also used without a large amount of empirical evidence.

The aim of this paper is to investigate, clarify and extend upon recent comparisons within the field regarding the sonic qualities of various models of the Nintendo GameBoy. A number of important points and areas of interest are discussed. Sonic qualities are compared and contrasted (by means of spectral and waveform analysis) in terms of their relationship to multiple units of the same model as well as the difference between various models. In addition, the amount and type of discrepancy in quality of different units of the same model (when compared between various models) is also a noteworthy one.

This line of inquiry is based upon the sampling of a large number of GameBoy units across all models including the original GameBoy, GameBoy Color, GameBoy Pocket, GameBoy Advance and GameBoy Advance SP. To some degree, a link between serial numbers and sound quality can also be made.

This comparison focuses on the point of a consumer-targeted machine such as the Nintendo GameBoy having been developed as a toy, with apparently looser parameters for the quality of production on a unit-by-unit basis than more professional musical equipment. By providing an in-depth analysis, the GameBoy music community can be well informed about the differences between the hardware models and the sounds that they create – an important aspect when considering the philosophy of authenticity within the micromusic genre.

References

Kann, Chris. Midines. Software and hardware. Published by Wayfar. http://www.wayfar.net/index.php. 2005. Accessed March 10, 2007.

Slocum, Paul. Synthcart. Software. Published by AtariAge. http://qotile.net/synth.html. Accessed March 10, 2007.

The Mood Juke Box project: classifying & retrieving music by its perceived emotional content

Alexandra L. Uitdenbogerd
RMIT University
GPO Box 2476V
Melbourne, 3001
Australia
sandrau@rmit.edu.au

Ian Kaminskyj
Monash University
Wellington Rd
Clayton, 3168
Australia
Ian.kaminskyj@eng.monash.edu.au

Abstract

Recently there has been a dramatic shift in how people obtain and listen to music. Songs are downloaded, played on PCs or transferred to portable MP3 players. This provides opportunities for developing novel and creative ways of locating or presenting music. One proposed method of finding music is to retrieve it by its perceived mood. This is of particular use to film and television directors who need to find suitable background music, but also has uses for the general population, such as for providing a romantic setting. This paper describes the background, aims, methodology and expected timeline of the Mood Juke Box project. Its principal objective is to develop robust techniques for mood-based retrieval of music. It is a collaborative research project between RMIT university and Monash university.

Introduction

Music is an important part of human culture, and with increasing on-line access to music, there will be greater need for technologies that improve our ability to find music in digital libraries.

In addition to the more usual methods of locating music via metadata such as title, composer or artist, other methods based on the *content* of the music, or other people's opinions of it, are being developed. Commercially, it is already possible to locate specific recordings via a sample recorded on a mobile phone (Wang 2003); the provision of recommendations via peer-based feedback exists in various forms (for example, user-rating-based recommendation of music pioneered by Shardanand and Maes, (1995)). A relatively new approach is to retrieve music by specifying the desired mood that is evoked by the music. For example, a film director may want to use music to evoke "agitation" during a scene, or someone may want to locate suitable music for a romantic evening. Currently this can only be achieved if music in the database has been labelled by hand by human annotators. Continuing the example, the Confutatis movement of Mozart's Requiem might be labelled "agitated".

There is little prior work on retrieving music by mood. The topic seems to have been first explored in our 2002 paper that surveyed psychology and sociology research related to music (Uitdenbogerd and van Schyndel 2002). The first researchers to build mood-based music classifiers published in 2003 (for example Liu et al. (2003)), and it has now become a regular topic within the music retrieval field. Liu et al. (2003 and 2006) described a music mood classifier that used seven audio timbre features, plus audio intensity and rhythm features to classify songs into one of four moods based on the two-dimensional (2D) model of emotions. As music often varies in mood as it progresses, the researchers also implemented a mood tracker. Experimental evaluation was based on musical experts' labelling of the mood of a collection of 250 works from the classical and romantic eras of classical music.

Feng et al. (2003) also used the 2D model for classifying music according to mood. They used a small set of features based on the tempo or pace of the music and articulation, or how the tune is played. Access to music was via representation of the pieces in a 2D mood space.

While the prior work in the field demonstrates mood classification techniques and a few user interface options for mood-based browsing, a thorough investigation remains to be achieved. In particular, the full integration of results from the field of music psychology has not yet occurred for mood-based music retrieval.

Project objectives

The aims of the Mood Juke Box project are to address the problem of music retrieval according to subjective attributes, such as the evoked mood of the music, and to discover how best to deduce the associated mood of a piece of a music through analysis of the audio signal itself. In order to address the above, the following more detailed aims will be examined.

Aim 1: *Determine the best way to represent mood classifications for retrieval of music by mood.*

There are two main ways that mood is quantified in music psychology experiments: a set of mood labels (Farnsworth 1958, Schubert 2003) (such as "cheerful", "sad", "calm", or "agitated"), and a 2D representation consisting of arousal and valence (Schubert 1999). The 2D approach as implemented in prior work (for example Liu et al. 2003 and 2006) limits the the mood categories to just four. However, there are more moods that are associated with music. If humans can distinguish them, then a mood classifier should also do so.

Some music varies in its emotion content as it progresses (Schubert 2004). It may be necessary to classify the start and end moods of music, for example, to allow a smooth mood progression for a playlist, that is, a list of musical works that are played in succession. A related problem is that listeners may wish to avoid certain moods in the music that is playing. For this to be possible, the detection of changes in mood in the music would be required.

Aim 2: *Determine the best way to evaluate fuzzy classification problems, such as the mood of music.*

Most problems that require classification, such as the presence or absence of cancer, tend to be well-defined, and rely on what is often called a "gold standard" or "ground truth". These denote a set of values that are assumed to be true for the problem, to allow systematic classifier training and evaluation. In the case of music mood, the choice of mood label for a given piece may differ depending on the listener. In the field of information retrieval, subjective assessments of the relevance of documents are used as the ground truth for evaluating document ranking algorithms (Sparck Jones 1995). A similar approach is applied to music retrieval based on melodic content, or generally perceived similarity to a given query piece (Downie 2006). However, in the retrieval case, binary or graded relevance judgements (that is, an item is classed as relevant or irrelevant to a query) are made between document and query pairs, whereas for classification, items in a collection belong to a class independent of a query. As with multiple relevance judgements for a query-document pair, multiple assessments of music according to mood label will lead to different assigned labels.

Aim 3: *Evaluate attributes of music that are needed to predict mood.*

Psychological studies suggest that tempo, tonality, distinctiveness of rhythm and pitch height are very important predictors of mood in music. In addition, however, other aspects of a musical arrangement, such as timbre and the strictness of the timing can also affect perceived mood (Livingstone and Brown 2005, Uitdenbogerd and van Schyndel 2002). Our aim is to evaluate these attributes.

Aim 4: *Determine the most appropriate interfaces for retrieving music according to mood.*

One possible approach is to visually cluster music based on the 2D mood model. However, as music collections can be large, a method for meaningfully managing the access to the collection is not immediately obvious.

Aim 5*: Determine whether music mood can be used to improve the accuracy of music recommender systems.*

User ratings have successfully been used to predict whether someone will like a new piece of music (Shardanand and Maes 1995). In 2002, we proposed that both demographic data and data from the music itself could be incorporated to improve music recommendations (Uitdenbogerd and Zobel 2002). Since then, we have shown that demographics on their own can work as well as typical collaborative filtering algorithms (algorithms that recommend new items based on user ratings) (Yapriady and Uitdenbogerd 2005). More recently, Yoshii et al. (2006) implemented a hybrid recommender system that uses audio features in addition to user ratings. These were low-level features rather than those associated with such perceived attributes as genre and mood.

Part of this project will involve comparing the effects of combining high-level and low-level features for music recommendation.

Aim 6: *Develop efficient techniques for mood-based music retrieval.*

The methods that are best in terms of *efficiency* for mood-based music retrieval will need to be determined after *effective* retrieval techniques have been discovered. An indexing technique is likely to be needed for efficient retrieval.

Methodology

For each of the aims outlined above, the following methodology will be applied in achieving these objectives.

Aim 1: Representing Mood

Experiments will be designed that compare the use of the 2D model and the emotion-label approach for music classification according to mood. These experiments will involve user studies as well as automated classification techniques. Furthermore, user studies will examine the continuously varying mood problem.

Aim 2: Fuzzy Classification

Few approaches have been proposed for effective evaluation of fuzzily defined classification problems in the field of music (Kaminskyj 2004, Mitri, Uitdenbogerd and Ciesielski 2004). There are, however, methods that have been applied in other domains such as computational linguistics (for example Grefenstette (1993)). These will be compared in the context of music mood classification, in addition to a method we have developed based on confusion matrices. The data collected from the user studies will be used for these experiments.

An alternative approach is the use of fuzzy classifiers (Yang, Liu and Chen 2006), in which the uncertainty of the mood classes is built into the system itself. Yang et al. evaluated their fuzzy system using standard techniques for classification evaluation.

Aim 3: Finding Mood Attributes in Audio

In order to establish the audio attributes that predict mood, several approaches will be used. First, we will verify the results from the psychology literature with data that has been labeled according to the features found to be important, such as tempo and tonality. Second, low level features and existing audio extraction techniques will be tested to see how well they model the important mood attributes. Third, existing compositional references on mood will be consulted to potentially find more attribute ideas. Fourth, new features will be developed that better model the attributes associated with mood that currently are not represented well enough by the existing features.

Classification Methodology

As shown in Figure 1, during the development, testing and subsequent operation of a pattern recognition system, three separate stages are required: data collection, data preprocessing, and classification.

Figure 1 Typical Pattern Recognition System

The first stage will involve collection of appropriate data to develop and test the classification system. Stage two relates principally to the extraction of features that will allow reliable and efficient operation of the classifier. Musical pieces need to be preprocessed so that features relevant to the classification method can be extracted.

The purpose of feature extraction is threefold: (a) to reduce the amount of data that is handled by the classifier, in comparison with the full recording, (b) to produce measurements which are invariant to extraneous variable variations (for example, the pitch range of the piece), and (c) to make the classification task as easy as possible for the classifier. Achieving (a) reduces the data storage requirements of the classifier and improves its speed of operation. Applying (b) allows improvement in classification reliability in the face of extraneous variable variations. Accomplishing (c) improves classification reliability.

Following stage two, the classifier makes use of the features extracted to arrive at a decision as to the most likely class that an unknown input sample belongs to. In this application, this relates to the most likely perceived mood of the musical piece. The amount of effort that is expended in selecting high quality features will directly determine the difficulty of the classification task. Where features have been found with very high discriminatory power for the classes being separated, the classification task becomes trivial and almost any classifier will do. If such powerful features cannot be located, the task undertaken by the classifier becomes much more difficult. For these situations, its performance becomes pivotal in determining the recognition capability of the overall classification system. Therefore, considerable effort will be directed towards devising a powerful combination of features so as to make the classification task as easy and reliable as possible.

Various classifiers will be tested, however it is possible that multiple classifiers may be more appropriate than a single one. These can be arranged in a number of different architectures. In addition, when combining the results of multiple classifiers, confusion matrices can be used with considerable success to determine the best overall decision, based on the results of the individual classifiers.

Aim 4: User Interfaces

To test interfaces for the Mood Juke Box, several different interfaces will be designed and built. User studies will then be run to determine their usability. These studies will require users to complete tasks, such as retrieve music for scenes in a film. Using standard usability methodology, user properties will be measured, including:

- learnability: the time taken to reach proficiency in the task,
- efficiency: the time taken to complete a task,
- memorability: user retention over time,
- error rate: the number of errors a user makes while carrying out a task, and
- satisfaction: the user's subjective opinion of the system

Aim 5: Recommender System

The use of mood data will be tested within a recommender system by comparing different formulations of evidence such as user ratings, demographic data and mood data, to determine the relative accuracy of recommendations to users. For example, linear sum with coefficients for scaling will be tested by learning optimal coefficient values on a training data set, and then performing evaluation on test data sets.

Aim 6: Efficiency

Once all the above research has been completed, several possible retrieval implementations will be compared for efficiency through the use of timing and space use experiments.

Timeline & Deliverables

We have commenced work on the project this year, and expect it to run until the end of 2010. During this time, several sub-phases will be defined with specific deliverables expected and managed. The breakdown of these phases is given below.

Phase 1: 2007

- Install signal processing development environment and other required software,
- Develop recommendation and user profile collection software,
- Develop software for mood-related feature extraction techniques,
- Define test bed and evaluation methodology, and
- Design core approach for Phase 2.

Phase 2: 2008

- Obtain ethics approval,
- Develop methodologies for evaluation of fuzzy classification systems,
- Define and develop a music categorisation system according to mood, and
- Collect mood information from listeners.

Phase 3: 2009-Mid 2010

- Present initial findings at conferences,
- Run experiments that compare user-preference based methods of recommendation to those incorporating mood,
- Develop techniques for presenting mood-based music selections, and
- Enhance machine listening based music categorisation system by incorporating mood.

Phase 4: Mid 2010 onwards

- Run experiments that compare user interfaces,
- Present results, and
- Prepare documentation and finalise project.

Conclusion

The increased availability of music on-line justifies the development of a variety of ways for users to interact with and retrieve it. The Mood Juke Box project aims to develop robust techniques for mood-based retrieval of music, and to verify these with a solid evaluation methodology. It is a collaborative research project between RMIT university and Monash University, which is currently in the prototype development stage.

References

Downie, J. S. 2006. "The music information retrieval evaluation exchange (MIREX)", *D-Lib Magazine*, 12(12).

Farnsworth, P. R. 1958. *The Social Psychology of Music*. Holt, Rinehart and Winston, New York.

Feng Y., Zhuang Y. and Pan. Y. 2003. "Music information retrieval by detecting mood via computational media aesthetics", *In J. Liu, editor, Proc. IEEE International Conference on Web Intelligence*, pp. 235-241, Washington, DC, USA, October. IEEE, IEEE Computer Society.

Grefenstette, G. 1993. "Evaluation techniques for automatic semantic extraction: Comparing syntactic and window-based approaches", *Technical report,* University of Pittsburgh.

Kaminskyj, I. 2004. *Automatic Recognition of Musical Instruments Using Isolated Monophonic Sounds.* PhD thesis, Electrical and Computer Systems Engineering Department, Monash University, Melbourne, Victoria, Australia.

Liu, D., Lu, L. and Zhang, H.-J. 2003. "Automatic mood detection from acoustic music data", *In H. H. Hoos and D. Bainbridge, editors, International Conference on Music Information Retrieval*, Vol. 4, pp. 81-88, Baltimore, MD, October.

Livingstone, S. R. and Brown. A. R. 2005. "Dynamic response: Real-time adaptation for music emotion", *In Y. Pisan, editor, Australasian Conference on Interactive Entertainment*, Sydney, Australia, November.

Liu, D., Lu, L. and Zhang, H.-J. 2006. "Automatic mood detection and tracking of music audio signals", *IEEE Transactions on Audio, Speech, and Language Processing*, 14(1):5-18, January.

Mitri, G. , Uitdenbogerd, A. L. and Ciesielski. V. 2004. "Automatic music classification problems", *Proc. Australasian Computer Science Conference*.

Schubert, E. 1999. "Measuring emotion continuously: Validity and reliability of the two dimensional emotion space", *Australian Journal of Psychology*, 51:154-165.

Schubert, E. 2003. "Update of Hevner's adjective checklist", *Perceptual and Motor Skills*, 96:1117-1122.

Schubert, E. 2004. "Modeling perceived emotion with continuous musical features", *Music Perception*, 21(4):561-585.

Shardanand, U. and Maes, P. 1995. "Social information filtering: Algorithms for automating word of mouth" *ACM Conference on Computer Human Interaction*, pp. 210-217. ACM Press.

Sparck Jones, K. 1995. "Reflections on TREC", *Information Processing and Management*, 31(3):291-314.

Uitdenbogerd, A. L. 2002. *Music Information Retrieval Technology*. PhD thesis, School of Computer Science and Information Technology, RMIT University, Melbourne, Victoria, Australia.

Uitdenbogerd, A. L. and van Schyndel, R. G. 2002. "A review of factors affecting music recommender success" *In M. Fingerhut, editor, Third International Conference on Music Information Retrieval*, pages 204-208, Paris, France, October.

Uitdenbogerd, A. L. and Zobel, J. 1998. "Manipulation of music for melody matching", *In B. Smith and W. Effelsberg, editors, Proc. ACM International Multimedia Conference*, pp. 235-240, Bristol, UK, September. ACM, ACM Press.

Uitdenbogerd, A. L. and Zobel, J. 2002. "Music ranking techniques evaluated", *In M. Oudshoorn, editor, Proc. Australasian Computer Science Conference*, Melbourne, Australia, January.

Wang, A. 2003. ""An industrial-strength audio search algorithm", *In H. H. Hoos and D. Bainbridge, editors, International Conference on Music Information Retrieval*, Vol. 4, pp. 81-88, Baltimore, MD, October.

Yang, Y.-H., Liu, C.-C. and Chen, H. H. 2006. "Music emotion classification: a fuzzy approach", *In Y. Rui, W. Klas, and K. Mayer-Patel, editors, Proc. ACM International Multimedia Conference,* pp. 81-84, Santa Barbara, CA, USA, October. ACM, ACM Press.

Yapriady, B. and Uitdenbogerd, A. L. 2005. "Combining demographic data with collaborative filtering for automatic music recommendation"

In R.Khosla, R.J.Howlett, and L.C.Jain, editors, Knowledge-Based and Intelligent Information and Engineering Systems, Vol. 9. KES, Springer, September.

Yoshii, K., Goto, M., Komatani, K., Ogata, T. and Okuno, H.G. 2006. "Hybrid collaborative and content-based music recommendation using probabilistic model with latent user preferences", *In Tindale A. Dannenberg R., Lemstrm K., editor, International Conference on Music Information Retrieval*, Vol. 7, October.

Guitar Playing Robot

Daniel Vogrig
davog1@student.monash.edu.au
Stephen Patterson
focus_17@optusnet.com.au
Ian Kaminskyj
Ian.kaminskyj@eng.monash.edu.au
Monash University
Wellington Rd
Clayton, 3168
Australia

Abstract

Robotic devices capable of playing musical instruments have been given much attention over the past decade, though devices capable of playing actual guitars have been met with limited success. The aim of this project was to construct a device capable of playing an actual unmodified acoustic guitar in a manner that most closely represented the way in which an actual human guitarist plays. An existing prototype, capable of playing only a single guitar string, was used as a starting point. Many methods of both plucking a string and actuating positions along the fret board of the guitar were researched. A solution was selected for achieving both of these functions and was consequently designed and constructed. The final prototype was actuated by numerous solenoids and a motor. Control was achieved via an Altera FPGA device and proved very successful at achieving the desired functionality. Song information can be input in the standard MIDI format and subsequently downloaded to the device and played at a reasonable tempo. The use of a very capable and flexible FPGA controlling device has provided ample room for expanding the current functionality of the robot in a wide range of respects.

Introduction

This project was conducted as a 2006 final year Electrical Engineering thesis project at the Electrical & Computer Systems Engineering department of Monash University. The objective of the project involved taking an existing guitar playing robot and upgrading the robot so as to be able to play over all six strings of an actual unmodified acoustic guitar.

Background

A previous robot was provided which came with severe limitations in its ability to play music; the main issue being that it was capable of playing only one string. The mechanical device used to actuate this one string made expansion to play additional strings impossible given the space constraints of an actual guitar. Thus, it was decided that a complete overhaul of the device would be required. Novel methods of plucking a string were researched and discussed.

Various devices have been created in the past in an attempt to robotically actuate stringed instruments, such as the guitar. The most notable of these are the:

- CrazyJ robot (Lawrence, Howard and Knueven 2000)
- LEMUR Guitarbot (Singer, Larke and Bianciardi 2003),
- Afasia: one man multimedia band (Jordan 2002), and
- the Taito guitar robot (Inglis 2004).

The Crazy J robot design is the most similar to what we were trying to achieve and it was the only guitar robot we discovered capable of playing an actual unmodified acoustic guitar. It consisted of a frame above an actual acoustic guitar where there were picks and fingertips that plucked and depressed the strings. This robot was very successful in its ability to play a variety of notes at a decent tempo. It was limited by the fact that the fingertips at the fretting end of the guitar were stationary, covering positions over only the first 4 frets. This limited the robot's playing note range. There were 24 separate fingertips which were spread out over 6 strings and 4 frets. Solenoids were located beside the guitar to actuate the fingertips via an intricate lever mechanism.

Plucking was achieved by using a pick on a thin vertical shaft. Bidirectional picking of the strings was achieved through the use of two solenoids per string.

Technical Details/Methodology

Our project was split up into several main components: plucking mechanism, fretting mechanism, movable carriage, electronics, onboard software and PC software.

The following section details the design process that was used to arrive at the final prototype.

Mechanics - Plucking Mechanism

As shown in Figure 1, the plucking mechanism on the existing guitar used a motor to spin a plectrum across a single string of the guitar. This device was very bulky and could not be used to fit within the confines of a 6 string playing mechanism.

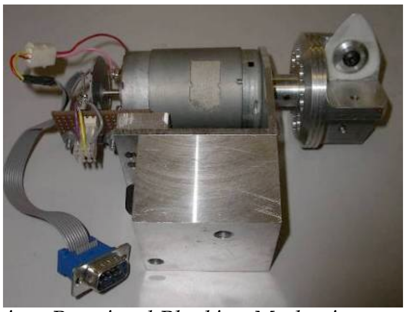


Figure 1.Previous Rotational Plucking Mechanism

Upon researching methods of exciting a string, it was decided that given the size constraint, the best method of excitation would be to use a harpsichord style plucking mechanism, similar to that used on the CrazyJ.

Actual harpsichord jacks were sourced to provide actuation of the string. A solenoid and lever mechanism was designed in order to move the plectrum up and down through the line of the string. Six identical devices were created such that each string could be actuated individually and at similar amplitudes of oscillation. The final prototype of the device is shown in Figure 2.

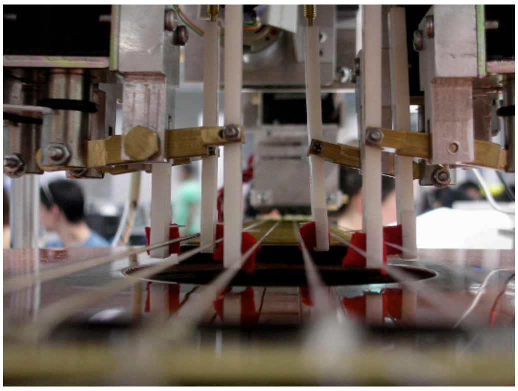


Figure 2. Harpsichord Plucking Mechanism

This device worked extremely well in achieving the desired functionality. Each string could be plucked individually at speeds up to approximately 9Hz.

The use of harpsichord jacks to excite the string produced sounds very similar to the sound created by a human guitarist's plucking action.

Another benefit of the new design was that on the return of the jack to the up position, the plectrum would not pluck the string; rather it would bend out of the way to bring a damper cloth into position. These damper cloths provided the capability of providing "note-off" functionality.

Mechanics - Fretting Mechanism

The most complicated part of a human guitarist's playing action is the work done to actuate differing positions on the fret board. Not only does a human guitarist depress one string at a time to play certain individual notes, but by depressing certain combinations of notes over the six strings of a guitar, the player can achieve a wide range of chords.

On researching what had previously been done in the area of mechanical actuation of the fret board, it was found that the majority of solutions involved a large static array of pneumatic or electromagnetic solenoids. This can be seen in such devices as the CrazyJ Guitar Playing Robot, the Taito Robot Guitar and Sergi Jorda's Afasia project.

This solution was well beyond our project budget due to the large amount of requisite solenoids and the serious mechanical challenges inherent in fitting that many solenoids within the limited available space.

A device capable of moving up and down the neck of the guitar with an array of "fingers" covering three frets was chosen to perform actuation of the fret board. This was done as it was achievable within the project budget and closely approximated what a human guitarist actually does while playing the guitar.

By modification of the available finger positions on the movable carriage, a range of chords could be achieved. An array of "fingers" was designed capable of playing all of the basic chords a guitarist would desire to play whilst minimising the number of solenoids required to achieve this. This array of solenoids is shown in Figure 3.

Figure 3. Minimalist Finger Array

In order to achieve actuation of these points, in the requisite space, small high-force solenoids were required. In order to minimise the power requirements of the device, magnetically latched solenoids were chosen with a magnetic holding force of 480g. This force was empirically determined as being adequate for holding down guitar strings.

These solenoids were arranged in a staggered way such that they all fitted onto the movable carriage. As shown in Figure 4, they were attached to levers on a number of combined pivot pins.

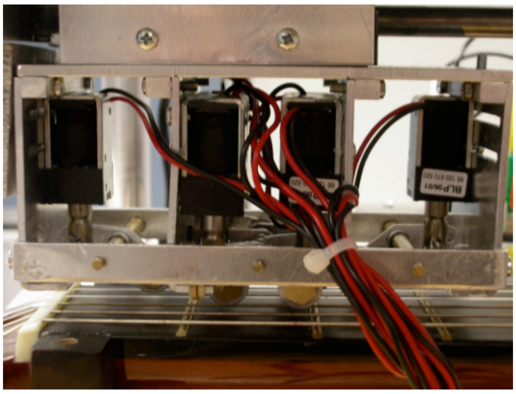


Figure 4. Fretting Mechanism

Mechanics - Movable Carriage

The fretting mechanism was required to move to different fret positions to play different notes and chords. This was achieved by attaching the carriage to a corkscrew thread salvaged from an old inkjet printer. As shown in Figure 5, a DC motor, salvaged from the same printer, drove the corkscrew thread.

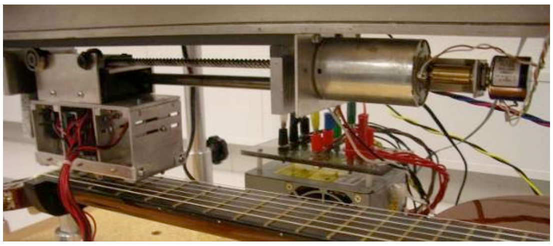


Figure 5. Movable Carriage Mechanism

Position feedback from the movable carriage was provided by a 10-turn potentiometer coupled to the drive shaft of the motor.

Electronics

The design of the electronics of the device had the following requirements:

- drive the magnetically latched fretting solenoids in either direction,
- drive the movable carriage motor in either direction at a variable speed,
- drive the plucking solenoids with a suitable amount of current, and
- provide feedback of the movable carriage position using sensor information.

These functions were separated into individual printed circuit boards (PCB) with a central breakout board from the Altera FPGA controller providing input/output (I/O) to these separate boards. For example, Figure 6 shows the solenoid driver PCB used to drive the fretting solenoids to hold down strings of the guitar.

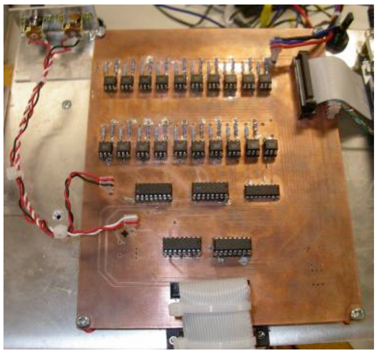


Figure 6. Fretting solenoid driver PCB

Figure 7 and Figure 8 show the circuitry used to drive the numerous solenoids of the robot.

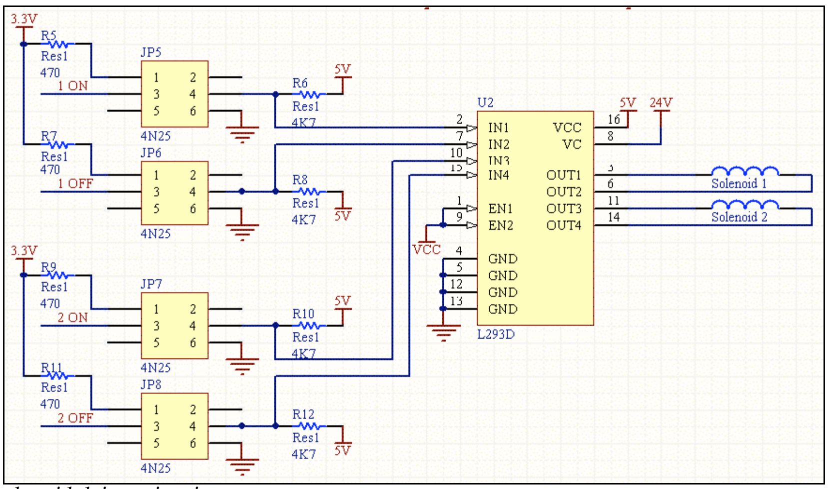


Figure 7. Fretting solenoid driver circuit

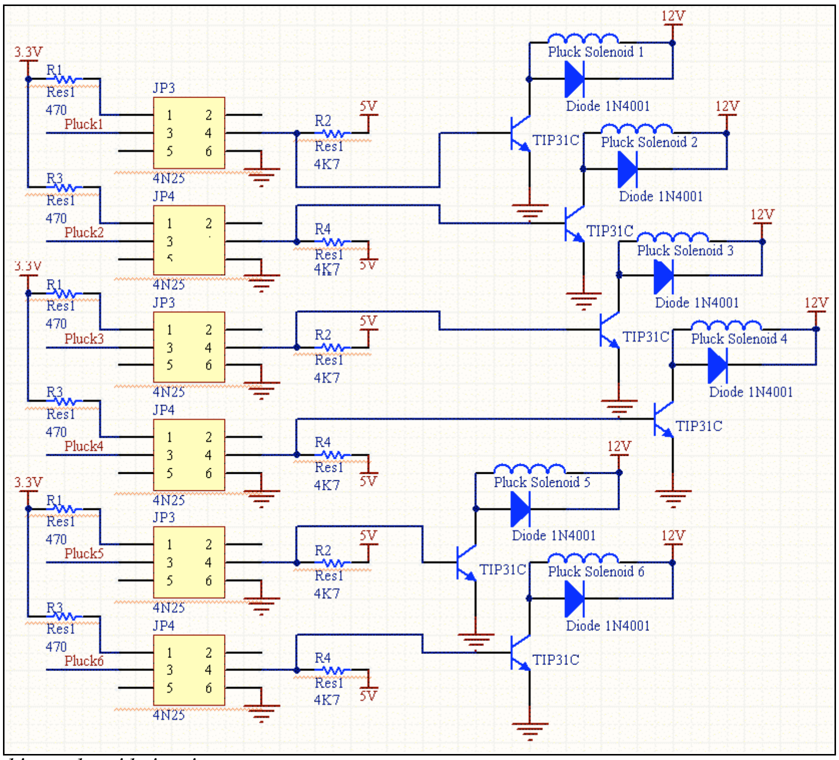


Figure 8. Plucking solenoid circuit

Fretting electronics

Fretting was achieved using a similar mechanism to that used for plucking except that it is done to a smaller scale and with magnetic latching pull-type solenoids. These solenoids require both a +24V supply to pull the solenoid in and a -24V supply to drop the solenoid out again. A small elastic band was used to assist the solenoid pin to drop back out again, when required.

As a bi-directional motor driver was used to control the direction of current, two digital outputs from the Altera microcontroller were used for each solenoid. All of these microcontroller outputs were opto-isolated by an optocoupler circuit. The full circuit of one fretting solenoid is shown in Figure 7.

Plucking electronics

Referring to Figure 8, each plucking solenoid was controlled by a single digital output signal from the Altera microcontroller. The 3.3 V output of the microcontroller was converted to a 5 V TTL level via an optocoupler. This voltage level was then subsequently converted to the 12 V needed by the solenoid by a motor driver IC.

Software

The software written to control the device consisted of two distinct parts:

Extraction Software - PC software for decoding type 0 and type 1 musical instrument digital interface (MIDI) files and writing their song data into a control information format easily utilised by the robot, and

Robot Control Software - Altera FPGA NIOS II microcontroller software used to run through PC extracted song data and send out the relevant commands to drive the robot in time with the song information.

PC extraction software

MIDI extraction software was developed for the project that extracts note information from a MIDI file and converts it into a format the robot is able to read. It does this by grouping the MIDI notes into chords, and then assigning a string to each note of the chord. A bridge position is then calculated depending on the array of available solenoids.

The program subsequently creates an output stream which, when read by the Altera microcontroller DE2 development board (Altera 2006), sends the necessary note on and off information as well as motor control information to play the guitar.

This output stream can be sent either to an external file or out the PC serial port to the Altera microcontroller memory.

Robot control microcontroller software

From the point of view of the robot, the initial header bytes of the song data obtained from the PC MIDI extraction software contains timing information for the song (pulses per quarter note and tempo) and the length of the song data.

Song information is then presented to the robot in a format consisting of a range of sequential chronological commands. The possible commands understood by the robot are:

- Note On: energises a finger solenoid or moves the finger carriage into position,
- Note Off: de-energises a finger solenoid,
- Pluck: plucks a string on the guitar,
- Dampen: dampens a string on the guitar, and
- Delay: waits for a prescribed amount of time.

Precise timing of these distinct tasks performed by the robot was programmed utilising a port of the MicroC real-time kernel on the NIOS soft-core microcontroller.

The device was programmed such that a central "control" task runs through the data stream byte-by-byte, posting messages into message-boxes as it proceeds. These message-boxes are then accessed by their corresponding relevant task so as to ensure the robot performs the desired function.

A collaboration graph of how this was achieved is shown in Figure 9.

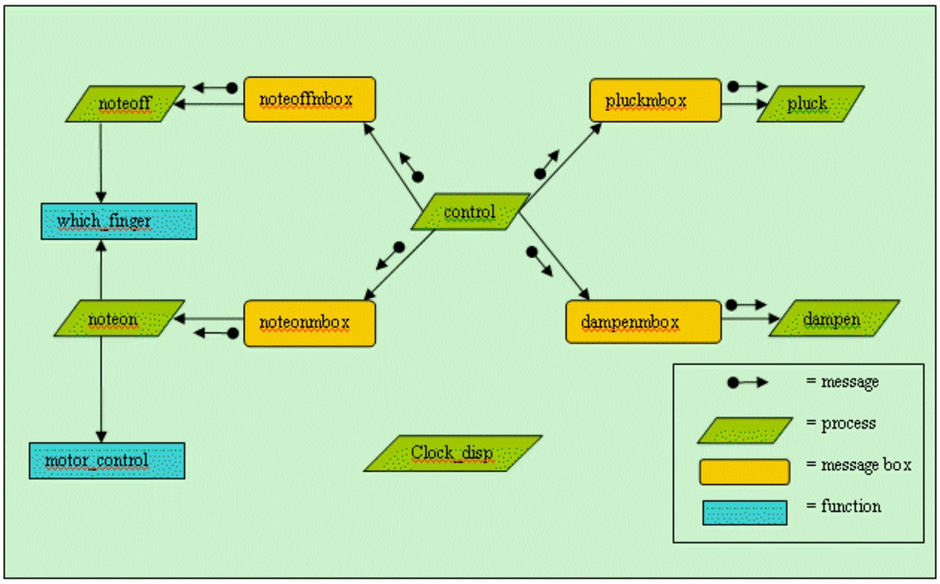


Figure 9. Real-time Collaboration Graph

Results

The plucking mechanism developed was found to perform well; strings were plucked in a timely fashion at a reasonable tempo.

For the fingering mechanism, it proved difficult to get all of the solenoids effectively holding down their respective strings. Once in place, however, acceptable operation was obtained.

Results of the position control system are shown below in Figure 10. The thin pink line indicates the measured position whereas the thick blue line represents the intended position. As shown in Figure 10, the position control was reasonably accurate. In the real world however, even the slightest error in the position can have a dramatic effect on the pitch of the It note played. The positioning control, therefore, was not perfect and could use further calibration. The final prototype is shown in Figure 11.

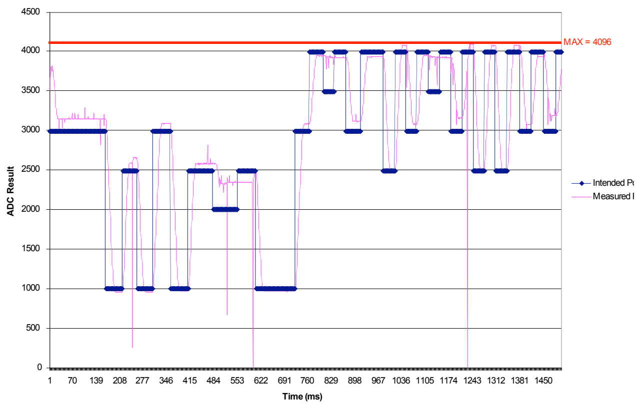


Figure 10. Position Control Results

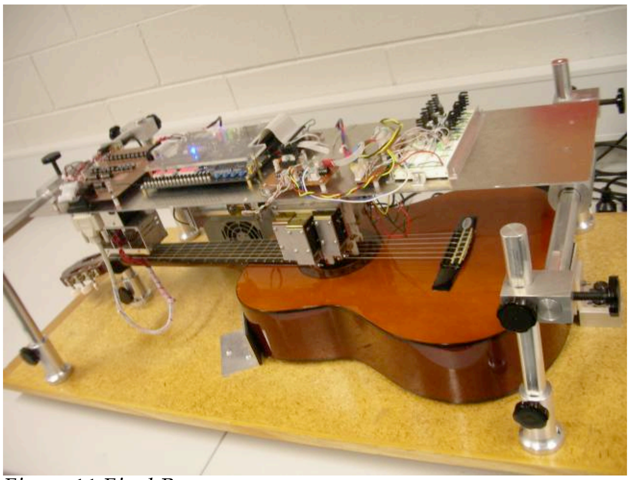


Figure 11 Final Prototype

Discussion

The robot has inherent limitations in its design that make it unable to play a wide range of music. These limits include (1) speed of movement and, (2) unplayable chords, and (3) unplayable notes.

The main speed limitation present in the device is the time it takes to get the movable fretting carriage into position. With the current position system, this can take anywhere from 100 - 200ms. This delay is offset against subsequent delays in the song by the MIDI extraction software, but it can still be too long to keep up with some song data.

Movement of the fretting carriage also generates noticeable mechanical noise which can interfere with the music being produced by the guitar. One way this could be overcome in future, however, is to use a suitable pickup inside the guitar, which would allow amplification of its tones. In this way, the desired guitar tones could be subsequently amplified to a level where the background fretting carriage mechanical noise becomes insignificant.

There are also a few problems with the setup of the robot as any minor adjustments to various heights or angles of the arrangement can cause the robot to fail to play correctly. This may require a good 30 minutes or more of adjustment to the robot in order for it to play properly and in tune. This can cause the robot to become frustrating to work with and is an area that still requires extensive investigation in an attempt to rectify it.

The fretting mechanism also needs to be expanded by the addition of more solenoids and improved by the use of a rubber material to hold down the strings.

The project could be improved in many areas, the main ones including:

Addition of more solenoids

The current project shows that it is possible to fit solenoids side by side in an array over all six strings. A future development could be to add more solenoids to the array so that every string and fret position is covered. This would allow more demanding songs to be played.

Improved fretting solenoid driver circuit

A major improvement to the solenoid driver circuit would be to modify it so that a different voltage is applied to allow the solenoid to drop out again. According to the data sheet, if a correct negative voltage is placed on it, there is a minimum force required to drop it out again.

Improved plucking mechanism

The current plucking mechanism severely limits the range of expression possible with the guitar robot. For example, playing at differing volume levels, i.e. with different dynamics, is currently not possible. The timbre of the plucking cannot also be varied. Different plucking mechanisms could be investigated which would better facilitate this.

Improved Manufacturing

The manufacturing process could be simplified and improved by remaking a lot of what was made. Parts could be made more accurately as there is now more information about the correct placement of all the parts and how they will work together.

Improvements to both the PC and Altera software

Currently, both programs work quite well depending on the type of input MIDI file but there are problems when some MIDI files are used. These problems are mainly due to the fret position found by the algorithm and also the relative sluggishness of the motor driver compared to what is sometimes required of it.

Conclusion

The resulting robot has achieved most of the project design objectives; being capable of (1) playing over all six strings of the guitar, and (2) playing a range of chords. Using a PC, the robot can accept standard MIDI files as input. These are downloaded to it and subsequently control its playing. Some simple songs have since been demonstrated on the robot, including Deep Purple's "Smoke On The Water", "Twinkle Twinkle Little Star" and AC/DC's "Back In Black". It differs from previous designs by the use of the movable fretting carriage as well as the use of a more flexible and capable FPGA controlling device.

With only minor improvements to its existing operation (eg. addition of more solenoids), a significantly broader range of chords could be achieved together with a more flexible arrangement for playing notes faster. In addition, given the use of a FPGA for this application, this has left the door open for significantly increasing the capabilities of the guitar robot. For example, given the large number of free I/O pins still available on this device, together with its significant processing capability, it would be quite feasible to control two or even three guitar robots with the one FPGA. Therefore, MIDI files comprising musical pieces meant for two or three guitars could be downloaded to the one FPGA, which would then control these different guitars to perform these multi-part pieces.

Real-time digital signal processing could also be performed by such a FPGA, either in hardware or software or a combination of the two. Therefore, it would be possible for the FPGA to perform a Fast Fourier transform on the sounds produced by the guitar, analyse the spectrum produced and detect, for example, the guitar going out of tune.

Acknowledgements

Many thanks must go out to the team at the Monash University Electrical Engineering workshop, notably Tony Brosinsky and Maurice Gay. They worked tirelessly to help us to complete this project.

Bibliography

Altera, 2006. "Altera DE2 User Manual" Retrieved October 12, 2006, from www.altera.com/education/univ/materials/boards/DE2_UserManual.pdf

Inglis, P. 2004, "The Taito Robot – guitar", Retrieved October 12, 2006, from http://www.thewholeguitarist.com/musos/taito.htm

Jordan, S 2002. *Afasia: the Ultimate Homeric One-man-multimedia band*. Retrieved October 12, 2006, from http://delivery.acm.org/10.1145/1090000/1085203/p1-jorda.pdf?key1=1085203&key2=1504959511&coll=&dl=ACM&CFID=15151515&CFTOKEN=6184618

Lawrence, J., Howard T. and Knueven, S. 2000. *CrazyJ Guitar Playing Robot*. Retrieved October 12, 2006, from http://www.me.gatech.edu/mechatronics_lab/Projects/Fall00/group3/contents.htm

Singer, E., Larke, K. and Bianciardi, D. 2003. "LEMUR GuitarBot: MIDI Robotic String Instrument", *In Conf. Proc. NIME-03*, Montreal, Canada, pp. 188-191.

Danielle Wilde
Monash University Art and Design
CSIRO Textile and Fibre Technology
Dept of Fine Art, Building D
PO Box 197 Caulfield East VIC 3145
AUSTRALIA
d@daniellewilde.com

hipDisk – an interactive sonic system inspired by core-body gesture.

Abstract

This paper describes the development and articulation of hipDisk, a musical interface that highlights, by making visible, the dynamic relationship between the wearer's hip and torso. The resulting interface effectively turns the body into an instrument by augmenting it with instrumental capabilities.

The hipDisk interface raises questions about the role of a direct exploration of, and response to, the physical affordances and capabilities of the body in the development of computer-mediated, interactive, sonic interfaces. The idea that such a focus can provide added value is discussed in relation to the hipDisk interface.

Keywords

HCI, full-body interaction, wearable interfaces, body dynamic, performance, soft and hard electronics, soft sensors, gesture-control, sound

Introduction

Possibly the most undignified musical instrument ever, *hipDisk* (see Figure 1) exploits changing relationships between torso and hip to actuate simple tones. Horizontal disk-shaped extensions of the body exaggerate, so make highly visible, the interdependent relationship of the hip and torso. Soft switches, strategically placed around the perimeter of each disk, allow the wearer to play a one-octave chromatic scale, and so play simple melodies, restricted only by core-strength and flexibility.

hipDisk was designed to inspire people to swing their hips and explore and extend the full range of movement available to them through a simultaneous, interdependent exploration of sound. In creating *hipDisk*, the objective was to move beyond limb- and digit-triggered switches and explore full-body movement for actuation. The resulting body-instrument interconnects choreography and composition in a fundamental way.

hipDisk sits within a broader research framework that explores the role of a poetic approach in the design of interactive interfaces, ranging from abstracted prosthetics through to invisible, virtual systems. This broader research targets not only sound output systems but also systems for the output of changes in colour, light and shape.

The paper begins with a brief survey of related research. This is followed by a detailed discussion of the conceptual and technical development of hipDisk. An evaluation of the instrument is then provided, along with an analysis of audience response. Finally, directions for further investigation are suggested.

Figure 1. *hipDisk*, demonstrated by the author.

Related Work

hipDisk is primarily related to the following research areas: human computer interface design (interaction design, HCI and CHI) – in particular wearable, gesture-controlled and sonic interface design. It is also informed by research into technical textiles and soft electronics.

The measurement of gestural input for the control of computer-mediated interactive interfaces has traditionally focussed on limb- and digit- triggered gesture. This is perhaps due to the complexity of multi-axis input. The softness and flexibility of the body's corporeal structure add to this challenge. Research into soft electronics and soft sensors is enhancing our ability to measure complex and fuller-body physical gestures. The work of Farringdon et al (1999) and, more recently, Gibbs and Asda (2005), and Helmer (2007) clearly shows this.

hipDisk uses custom-built soft sensors to measure core-body gesture, but rather than measuring body displacement directly on the body – as in the work cited above – the custom switches developed for *hipDisk* support the measurement of a mechanical event that takes place in a physical extension of the relevant body parts. By approaching the problem in this way, it was possible to sidestep many of the more complex issues associated with the measurement of full-body gesture. In this way the resulting interface is related to Kei Kagami's Head Holder (2006), a dress that explores dynamic structures by means of rods and strings that cause exaggerated movement of the garment by the motion of the model or wearer.

An ability to understand and measure changing body dynamics seems integral to an appropriate development of interactive gesture-triggered systems if they are to reflect the complexity of the body's use, and dynamic potential. Extending and exaggerating the body and specific dynamic relationships seems to enhance our ability to read changes in dynamic, so simplify technical requirements.

There exist numerous interfaces for gestural control of sonic output. Jordà provides a detailed compendium in his PhD thesis (2005). For more recent examples, Bardos et al's Bangarama (2005) provides a suprising, albeit obvious example of a gesturally-controlled sonic interface; and Helmer's Wearable Instrument Shirt (WIS) (2007) provides an example that incorporates custom-designed soft electronic sensing. Both Bangarama and WIS manipulate guitar samples, but in the case of WIS, the same interface can also be used to manipulate other instruments – i.e. the interface is not instrument specific.

Many sonic output devices manipulate samples, mapping physical gestures to sonic gesture. An early example is Waiswiz's Hands (1985), which allow both recording and manipulation of samples for output. Zigelbaum et al's Ringalings (2006) allows individual users to map specific gestures to triggered sound samples on a case-by-case basis before performing or experimenting. Goto's BodySuit (2006) is used to control percussion robots, translating or altering the gestures algorhythmically before sending them on to the robots. Direct manipulation of artefacts can also support the mapping of physical gesture to sonic gesture. Hewitt's eMic (2003) and Singer's Sonic Banana (2003) provide clear examples.

But *hipDisk* does not map physical gesture to sonic gesture. Nor is it a manipulable artefact.

The output of *hipDisk* is clearly related to the author's previous work Ange (2004), which allows the player to trigger sound samples and control volume, so mix up to twenty-four samples in realtime, simply by manipulating their volume. In the case of Ange, sonic complexity is achieved through simple means. But this is where any similarity between the interfaces ends – though output is clearly related and Ange is worn, gestural control of Ange is limited to finger or hand pressure, provided either by the wearer or someone else. Core-body gesture of the wearer does not affect output in any way.

hipDisk allows the wearer to trigger individual tones, and so build or play simple melodies. In this way it comes closer, perhaps, to traditional acoustic instruments such as the piano or the recorder. The input/output relationship is also simple and direct, the tones triggered through core-body gesture, allowing us to make a clear correlation to physical interfaces like Dance Dance Revolution (Smith, 2004).

hipDisk

In this section the context of the creation of *hipDisk* is discussed, as well as its impact on the development of the interface. A detailed technical overview is then provided.

Context

The first prototype of *hipDisk* was conceived and developed during Reskin, ANAT and Craft Australia's wearable technologies lab, which took place over a three-week period in January and February 2007.

Intensive residential labs provide a particular framework for the creation of new work. Reskin was no exception. The first two weeks of Reskin were focussed almost entirely on skill acquisition. There was a limited amount of ideation, or development of ideas, and it was only during the third, and final, week of the lab that the focus was on the development and construction of new works. In accord with this framework, *hipDisk* was conceived and created during an intense seven-day period.

From the outset, the outcomes of Reskin were intended for public display. This requirement had a clear impact on the development of work. For example, on the final day of the lab a gala event was held for public presentation of work and on the following day, the WearNow Symposium – an international symposium focussed on wearable technologies – was held at the National Museum of Australia. The symposium included a session devoted to Reskin outcomes, including live demonstration of the created works.

Though it was clear that such a short, albeit intense, period of work, which included skill acquisition, could only result in the creation of prototypes, these prototypes had to function in highly publicised and public contexts. Small and subtle work would, necessarily, be lost in these contexts. As discussed below, the *hipDisk* developed accordingly.

Conception

The original intention that led, finally, to the creation of *hipDisk*, was to explore ways of using conductive fabrics to measure or track physical movement, or changes in the body's dynamic, so that these changes could be used to actuate digitally-mediated events.

Experiments for the input system included weaving conductive thread into three-dimensional ridges or channels that could be placed against the sides of the torso to measure bend, and the creation of small wing-shapes, that extended out from the body in a similar manner.

Both of these experiments were concerned with the body's movement along a single axis. Whilst not problematic in and of itself, such a restriction seemed inappropriate for the context in which the work would be demonstrated. The concern was that a physically responsive system demonstrated by raising the arms and bending from side to side was not only limited, but, in this particular context, would seem somewhat ridiculous. As a result, the focus shifted to an investigation of more fundamental, dynamic and fuller-body movement.

Different parts of our bodies have varying degrees of freedom. The relationship between the hip and torso is particularly dynamic as this area of the body allows movement on its axes in an unconstrained fashion. It also provides a fundamental reflection of core strength and flexibility. For these reasons the relationship between hip and torso was identified as a more appropriate source of input for the envisioned system.

Input

In order to explore the changing dynamic between hip and torso, a pair of disks – one attached to the hip and one attached to the torso above the waist, were created. The aim was to make this dynamic relationship highly visible and so provide an input that was easy for the viewer to 'read'.

Twelve simple, soft, digital switches were then positioned at equidistant intervals around the perimeter of the disks – one part of each switch on the upper disk, the complementary part of the switch on the lower. The switches could then be activated when the corresponding parts of the upper and lower disks connected, as achieved through bending the body at the waist. (see Figure 2).

Figure 2. Wilde, triggering different switches

Construction: Soft vs Hard Electronics

'Hard' and 'soft' electronics are both used in the *hipDisk* interface.

'Hard electronics' is a name given to traditional electronic circuits and componentry. Typically constructed of hard plastics, metals, silicon, etc., the components are literally hard, in a tactile sense, rigid and angular. They are often brittle and are unsuitable for any use that requires flexibility or stretch.

Soft electronics replace wires and other conductive surfaces with conductive fabrics and threads that incorporate metal filaments into their grain or weave yet remain flexible and soft, in a tactile sense. This flexibility and softness make them ideal for use on the body in that they can be worn, rather than merely attached or placed next to the body.

Soft circuits, clearly, have different properties to traditional, hard circuits. The use of soft switches in *hipDisk* was conditioned by the need to ensure contact between two surfaces that would meet at an angle, in a system where the central, vertical axis (in the case of *hipDisk*, the spine) allowed for horizontal displacement.

It was originally envisioned that the soft switches would provide variable resistance, so provide the ability to measure velocity and express continuous musical gesture, but this was determined to be impractical from a technical standpoint. The conductive qualities of the fabrics used, combined with the nature of contact allowed by the *hipDisk* interface, did not allow a high enough resolution of electrical signal to read pressure, or amount of surface touching. The interface was, as a result (and in consideration of the development time-frame), limited to a series of twelve on-off switches.

The reduction, from variable input to a simple on-off triggering system, affected the range and quality of possible output for the *hipDisk* system, and so impacted the subsequent development of the interface.

Output

Within the larger research project any form of output could have been deemed appropriate for an interface designed to explore full-body gesture. The choice of sound was conditioned by a number of considerations.

Typically the body doesn't make sound when it moves, but sound, being an independent sensory modality, provides an ideal contrast to the highly visible, movement-based input of the *hipDisk* interface. Through the use of sound, a clear, identifiable separation between input and output – as perceived by the audience or viewer – could be achieved.

In addition, sound output can be quite simple or highly complex while still allowing for complexity of use. The development timeframe made this particularly attractive.

While the author's previous work has focussed on the creation of live performance and interface design, particularly in the field of wearable and portable costumes and interactive sculptural elements, sound was not entirely out of scope. Admittedly the final sonic output of *hipDisk* is highly simplistic, but, as mentioned above and as will be clarified below, this was not unconsidered.

Once sound was chosen as the most appropriate output modality, design of the system could then be undertaken. In order to clarify the author's thinking about *hipDisk* as a sonic interface, consultation was undertaken with a number of Reskin participants who had undergone traditional musical training. Though they didn't necessarily agree, the choice of a twelve-note interface was based on the following determinants:

- the need to place the switches evenly around a circular formation, in an easily recognisable and actionable pattern.
- the range of musical possibilities provided by a chromatic scale, as opposed to a five- or eight-note system, or any other single key or mode
- the physical precision required to bend the torso into twelve distinct positions, as opposed to eight positions, or any other number

This final point was driven by the author's desire to create a system that could be played by a novice, yet through mastery, allow greater complexity. It is the author's impression that this quality is intrinsic to many musical instruments. The example of a recorder comes to mind. Simple tunes can be played on a recorder by novices yet an entire canon of highly complex and beautiful music exists for the same instrument.

While this is conceptually sound, the final interface has a learning curve so steep that, to date, no actual tunes or songs have been learnt by the author. In addition, the precision of physical positioning, required to trigger the individual switches, is not consistent. While the centre-front, centre-back and side switches are easy, consistent identification and triggering of the other eight switches has proven to be highly problematic. It would seem that an eight-note system would perhaps be more viable, simply because it would be a little easier to learn and operate.

Design Challenges

A number of design issues needed to be addressed in order to construct the envisioned interface, particularly considering the time-frame within which a fully functioning, robust, prototype needed to be realised.

Two fundamental issues were how to keep the disks horizontal, and how to maintain an ideal distance between them. The source of the sound output also needed to be addressed as it was considered important that the system be autonomous – that the wearer of the interface be able to move around freely within a self-contained system, with no need for additional computers or amplification beyond that which was to be worn on the body.

It is perhaps useful to point out that an underlying driver in the ongoing work of the author is the need for simplicity in interface design. In reflection of this, it was considered important that all elements of the interface be functional responses to the technical and/or conceptual development of the work. The resulting interface, as a result, steers clear of decorative addition.

It should also be noted that a number of aesthetic issues are yet to be addressed. In developing the interface further, different materials would perhaps be employed, and the design, necessarily, refined and further developed.

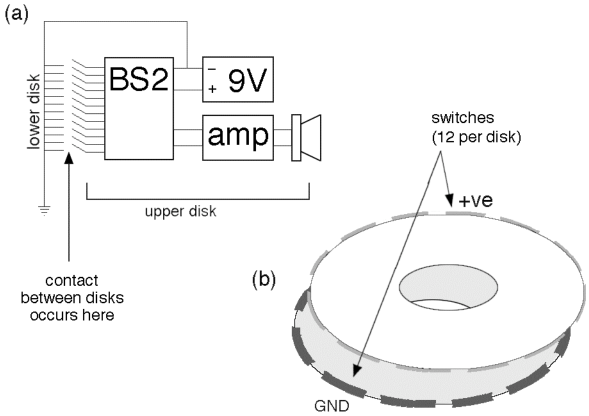


Figure 3. (a) schematic
(b) placement of switches around disks

Technical Overview

The sounds output by *hipDisk* are simple tones created digitally by the Basic Stamp 2 microcontroller, triggered by a series of twelve custom-built soft switches, amplified, then output through a simple speaker. The entire system is self-contained. (see Figure 3).

Construction

The current iteration of *hipDisk* is made of polypropylene sheets, cut and connected with rivets and reinforced with aluminium strips in order to maintain a perpendicular relationship between the waist- and hip-bands and the corresponding horizontal disks. Attached to the exterior of the waistband, on the upper disk, are two transparent plastic boxes that contain all of the electronics – the box on the right contains the Basic Stamp, that on the left a 9V battery, a small amplifier and an on-off switch (see Figure 4). All wires are held down with a combination of white electrical tape and white duct tape, except where they enter into the boxes containing the circuitry.

Lycra tubes filled with cotton braid and wadding are mounted on the interior faces of each disk, to support the soft switches (see Figure 5). The switches consist of two layers of conductive fabric sewn together with conductive thread, wrapped at

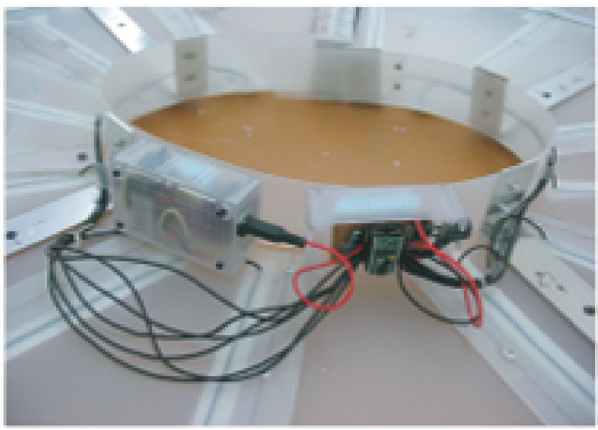


Figure 4. upper disk - detail

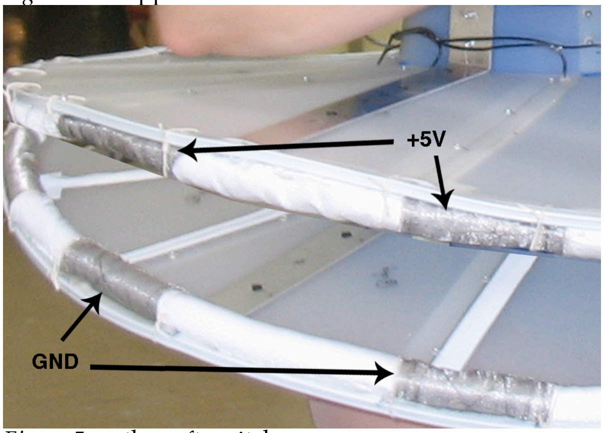


Figure 5. the soft switches

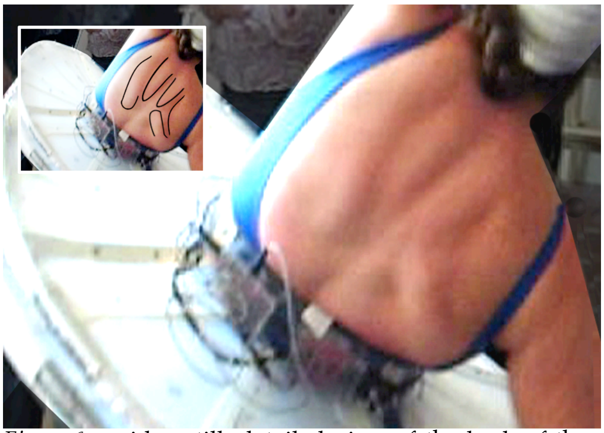


Figure 6. video still: detailed view of the back of the wearer while performing with *hipDisk*

specific points around the lycra tubing. The softness maximizes the potential for contact by providing greater surface contact under pressure.

Conductive fabrics and threads have varying amounts of resistance depending on the surface area and amount of thread in contact – the more thread in contact, the lower the resistance. While this resistance would not impact the functioning of the switches in any direct way (as the switches were binary), the impact of this additional resistance had to be taken into account in the design of the overall circuit.

Sound output is through a small speaker mounted into the bodice of the costume of the wearer. When *hipDisk* was demonstrated in the context of Reskin, it was found that the cleavage of the wearer provide a suitable amplification chamber for the speaker. As a result the sound output capabilities of the system were enhanced. It should be noted though that his aspect of the interface would necessarily vary with each performer.

Outcomes

A number of surprising outcomes have resulted from the creation of the *hipDisk*. The interface seems to embody an inherent interdependence between composition and choreography that leads the wearer to perform at the outer limits of his or her physical capabilities.

In 'reaching' for contact to be made by the two disks, the performer's body seems to be thrown almost unconsciously into a variety of positions that exist purely in order to support the necessary physical extension that will result in contact.

Rhythm of desired output is also closely linked to core strength and flexibility – the tempo of the produced melodies are limited only by the wearer's ability to move from one position to another.

Though *hipDisk* was designed specifically to inspire people to swing their hips and explore and extend the full range of movement available to them, the physical repercussions of this were not fully understood in advance. It was only after public presentation of the interface that this became apparent.

Video footage of the first demonstration of *hipDisk* (see Figure 6), clearly shows that the wearer's back muscles were highly engaged to perform the necessary dynamic movements. The following day muscle strain also indicated that a strenuous workout had been undertaken the day prior.

In light of this, it is recommended that a full physical warm-up be undertaken before use of the *hipDisk*, and additional stretching and cool-down exercises be performed after use.

Despite the fact that *hipDisk* does not appear as an exercise interface, the performer's focus on doing whatever is necessary to achieve the desired sound output seems to extend, over time, the body's capabilities to do so. Their ability to move from one position to another in a timely manner is enhanced through use. As a result, core strength and flexibility are positively affected through an extension of the outer limits of the wearer's physical capabilities. The *hipDisk* thus provides surprising, yet not undesirable health enhancing aspects for the wearer.

In addition, use of the interface seems to be highly compelling for audiences.

Audience Reception

hipDisk has been shown in four radically different contexts to date (at the time of writing). The first was a gallery opening; the second an academic-style conference; the third an informal presentation to representatives of a government funding organisation; the fourth, at Dorkbot Melbourne to a group of technically savvy arts professionals.

Each of these contexts was radically different, yet the resulting interface was found to be strangely compelling in each instance, almost without exception.

It is difficult to qualify what it is that makes *hipDisk* so compelling without conducting rigorous audience surveys. Some informal questioning and conversation has been undertaken to date with participants in each of the aforementioned fora so an attempt will be made to discuss the question despite the lack of statistical information.

The design goals, in creating *hipDisk*, were:

- to explore ways of using conductive fabrics to measure or track physical movement, or changes in the body's dynamic, so that these changes could be used to actuate digitally-mediated events
- to inspire people to swing their hips and explore and extend the full range of movement available to them
- to make a musical instrument that was at once simple, so able to be played by a novice, and sufficiently complex, so that mastery would afford musical complexity.

It is the author's belief that the first goal has not impacted on audience response as it has been driven by purely functional concerns, and the use of soft conductive substrates is not actually evident to an audience member without a detailed inspection of the interface.

The impact of the third goal can perhaps only be fully explored once a level of mastery has been achieved, though it should be noted that in presentation of the interface the suggestion of playing actual tunes with the *hipDisk* does elicit much laughter and encouragement. In performance there is currently no discussion, or clear evidence, of this aspect of the interface's potential, unless we accept that any audience member who has ever heard a piano, keyboard or recorder, could arguably extrapolate this potential from the available information. As a consequence it's difficult to argue the impact of this goal until further development has been undertaken.

The second goal, to inspire people to swing their hips and explore and extend the full range of movement available to them, seems to be that which has impacted most favourably on audience response.

To clarify: The interface seems to embody an inherent interdependence between composition and choreography that can only be demonstrated with an inordinate amount of effort. The wearer is required to perform at the outer limits of their physical capabilities for the simple result of a single audible tone. Added to this, rhythm of desired output is closely linked to core strength and flexibility – the tempo of the resulting melody is limited by the wearer's ability to move from one position to another. The physical difficulty presented by this seems to compound the imbalance in input and output energy. Finally, in 'reaching' for contact to be made by the two disks, the performer's body seems to be thrown almost unconsciously into a variety of rather strange positions that exist purely in order to support the necessary physical extension that will result in contact.

This combination of strange body positions and extreme imbalances in input and output energy seems to be both humorous and compelling.

Other unintentionally attractive aspects of the device seem to be:

- the choice of costume and persona – reminiscent of a character from a circus. Though there seems to be no clear logic behind the choice of bathing costume and cap, the original idea was conceived when the author became conscious of the pleasure experienced swinging her hips whilst swimming laps. The incongruity of this choice seems to support the strangely humorous and compelling nature of the interface as a whole.
- the ungainly manner of physical input necessary to achieve what could arguably be (and certainly sometimes is) highly pleasing and beautiful audible output
- the fact that the author, when wearing the interface, seems compelled to smile broadly, continuously, throughout her performance.
- and finally, the fact that anyone could even think of such an interface, let alone realise it and perform in it seemed to be particularly compelling.

hipDisk demands, conditions and requires total freedom of physicality on the part of the wearer. It cannot be used in a restrained manner. In fact, the effort required to use *hipDisk* effectively removes any thoughts of appearance, as well as conscious thoughts of the external realities of physical displacement at the time of use.

Activities that require total freedom from physical constraint, as a rule, seem to be outside the boundaries of most people's limits, so perhaps the desire to exist in an unconstrained manner is what is highly compelling? Is this not what leads people to undertake extreme sports and other physically demanding, extreme or dangerous activities?

Many adults, in fact, asked the author if she would be constructing *hipDisks* for children. Though it's beyond the scope of this paper it would seem that being unconstrained can be equated with freedom, the kind of freedom perhaps experienced by, or equated as being possible for, children who do not yet need to live by adults' rules.

Future Directions

Future directions include aesthetically resolving the current interface, and creating an extensible version that can fit multiple-sized wearers. The author's intention is then to work on a series of performances, beginning with the realisation of the standard tunes: "the Girl From Ipanema" and "Do You

Know The Way To San Jose". The objective is to create compelling performances that explore the potential of an interlinked choreographic compositional system, as well as to explore the possibilities made available by numerous *hipDisk*ed performers. For example, four *hipDisks* make it possible to play Jazz chords, or to provide back-up for solo singers or players, to harmonize, etc.

Within the context of the larger research framework, the author will also be exploring more subtle and embedded interfaces for the measurement of physical gesture, and the subsequent triggering and more complex control of sound output.

Conclusion

Many decisions were made during the realisation of *hipDisk* that were conditioned by the constraints of the context in which it was created. The interface is a necessarily highly simplistic input-output system. The overwhelmingly positive response to *hipDisk* has led the author to question what it might be that would give such a simplistic interface so much power to excite, enthuse, captivate. It seems clear to the author that the focus on gestural input, the body, it's dynamic and affordances have contributed in no small part to the success of the work.

Acknowledgements

Thanks to the ANAT for support to attend the Reskin wearable technologies lab.

Thanks to Cinnamon Lee, Michael Yuen, Somaya Langley and Alistair Riddell for their input into the interface design.

Special thanks to Ross Bencina.

Figures and Tables

Figure 1. *hipDisk*, demonstrated by the author.
Figure 2. Wilde, triggering different switches
Figure 3. (a) schematic
(b) placement of switches around disks
Figure 4. upper disk - detail
Figure 5. the soft switches
Figure 6. video still: detailed view of the back of the wearer while performing with *hipDisk*

References

ANAT – The Australian Network for Art and Technology. [online] accessible at: http://www.anat.org.au

Bardos, L., Korinek, S., Lee, E., and Borchers, J. 2005, 'Bangarama: Creating Music With Headbanging' in *Proceedings of the 2005 International Conference on New Interfaces for Musical Expression (NIME05)*, Vancouver, BC, Canada

Dance Dance Revolution: [online] accessible at: http://www.musicineverydirection.com/ see also Smith, J. (2004)

Dorkbot Melbourne. [online] accessible at: http://www.dorkbot.org/dorkbotmelbourne

Farringdon, J., Moore, A., Tilbury, N., Church, J., and Biemond, P. 'Wearable Sensor Badge & Sensor Jacket for Context Awareness' in *The Third International Symposium on Wearable Computers, 1999.* Digest of Papers.

Gibbs, P., and Asada, H. 2005, 'Wearable Conductive Fiber Sensors for Multi-Axis Human Joint Angle Measurements' in the *Journal of NeuroEngineering and Rehabilitation 2005*, 2:7 doi:10.1186/1743-0003-2-7

Goto, S. 2006, 'The Case Study of An Application of The System, "BodySuit" and "RoboticMusic" - Its Introduction and Aesthetics.' *NIME 2006*: 292-295

Helmer, R. 2007, 'Wearable Instrument Shirt (WIS)' in *personal correspondence*

Hewitt, D., & Stevenson, I. 2003 'E-mic: Extended Mic-stand Interface Controller' in *Proceedings of the 2003 Conference on New Interfaces for Musical Expression (NIME-03)*, Montreal, Canada

Jordà, S. 2005, 'Digital Lutherie: Crafting musical computers for new musics' performance and improvisation' *PhD Thesis* submitted to Departament de Tecnologia Universitat Pompeu Fabra, Spain.

Kagami, K. 'Head Holder' in *The Fashion of Architecture: CONSTRUCTING the Architecture of Fashion*. (exhibition. Bradley Quinn, curator) AIA Center for Architecture, New York January 11 – March 11, 2006

Reskin [online] accessible at: http://www.anat.org.au/reskin

Singer, E. 2003, 'SonicBanana: A Novel Bend-Sensor-Based MIDI Controller' in *Proceedings of the 2003 Conference on New Interfaces for Musical Expression (NIME-03)*, Montreal, Canada

Smith, J. 2004, 'I Can See Tomorrow In Your Dance: A Study of *Dance Dance Revolution* and Music Video Games' in *Journal of Popular Music Studies 16 (1)*, 58–84. doi:10.1111/j.0022-4146.2004.00011.x

Waiswisz, M. 1985, 'The Hands, a Set of Remote MIDI-Controllers' in *Proceedings of the 1985 International Computer Music Conference*. San Francisco, California. International Computer Music Association, pages 313-318.

Wilde, D., Birkmayer, S. 2004, 'Dress and Ange: coercing the address of highly personal body-centric issues' in *Personal and Ubiquitous Computing (2004)* 8: 264–273 DOI 10.1007/s00779-004-0287-6

Zigelbaum, J., Millner, A., Desai, B., Ishi, H. 2006, 'BodyBeats: Whole-Body, Musical Interfaces for Children' in *CHI 2006*, April 22–27, 2006, Montréal, Québec, Canada.

Sonia Wilkie
MARCS Auditory Laboratories
University of Western Sydney
Locked Bag 1797,
South Penrith 1797,
Australia
s.wilkie@uws.edu.au

Catherine Stevens
MARCS Auditory Laboratories
University of Western Sydney
Locked Bag 1797,
South Penrith 1797,
Australia
kj.stevens@uws.edu.au

Psychoacoustic variables for temporal manipulation of the Sound Induced Illusory Flash

Abstract

Psychological research on cross-modal auditory visual perception has focused on the influence of visual information on sensory information. There is a lacuna in using an auditory stimulus to manipulate other sensory information. The Sound Induced Illusory flash is one illusory paradigm that uses the auditory system to bias other sensory information. However, more research is needed into the different conditions under which the Sound Induced Illusory Flash manifests and is enhanced or reduced.

The experiment reported here investigates the effect of new auditory variables on the Sound Induced Illusory Flash. The variables to be discussed include the use of varying timescales, pitch intervals, and harmonic relationships. The ultimate aim is to develop the illusory effect as a basis for new multi-media techniques and creative applications for the temporal manipulation and spatialisation of visual objects.

Introduction

Under certain conditions, particular combinations of stimuli – auditory and/or visual - can give rise to non-veridical perception or an illusion. For example in the Müller-Lyer illusion, the horizontal lines are physically equal but the line with the outward arrows at the end is generally perceived to be longer (figure 1.)

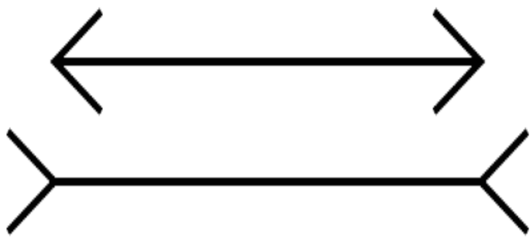


Figure 1. The *Müller-Lyer illusion illustrating that although the lines are the same length, the line with the outward arrows is often perceived as longer.*

This paper investigates non-veridical perception where auditory stimuli influences visual perception. Research on cross-modal interactions in auditory and visual perception have focused predominantly on conditions where visual stimuli are used to manipulate auditory perception (McGurk and Macdonald, 1976) (Alais and Burr, 2004). In such experiments, results suggest that vision is the dominant sense. However, the Sound Induced Illusory Flash illusion (Shams, Kamitani and Shimojo, 2002) exploits the capacity of the auditory system to distort visual perception.

Conditions that give rise to this cross-modal illusion involve presentation of a visual stimulus consisting of a single white dot that is flashed once in the participant's peripheral visual field. This is accompanied by an auditory stimulus consisting of multiple beeps of sound. The participant's task is to judge the number of times the dot is presented. The dual presentation of auditory and visual stimuli appears to emanate from a single source creating confusion about the number of physical flashes presented and gives rise to the percept of the white dot flashing equivalent to the number of auditory beeps. For example, the dot presented once with three beeps is perceived as three presentations of the dot. This illusory percept appears to occur because of the superior resolution of the auditory system for rhythmic perception, which in this case overrides visual information.

The Sound Induced Illusory Flash illusion is a recent discovery (Shams, Kamitani and Shimojo, 2000) with limited associated research. Whilst recent research (e.g. Shams, Iwaki, Chawla and Bhattacharya, 2005; Shams, 2005; Shams, Kamitani, and Shimojo, 2002) has focused on the underlying neural activity during the illusion, only the initial studies (Shams, 2002) (Shimojo and Shams, 2001) (Shimojo, Scheier, Nijhawan, Shams, Kamitani, and Watanabe, 2001) explored basic structural variables and their effects on the illusion. It is these structural, especially acoustic, variables that are of interest for further exploration of the illusion experience. Such research outcomes provide new opportunities for creative application and multimedia transmission techniques where auditory stimuli might influence visual perception.

Auditory Stimulus Variables and the Sound Induced Illusory Flash – A Review

Rhythm

The variables manipulated in previous research consist primarily of various combinations of the number of auditory and visual stimuli presented (Shams, Kamitani, and Shimojo, 2002), with only minor adjustments to stimuli between experiments (including pitch frequency at 1 kHz or 3.5 kHz; the transmission of auditory stimuli through headphones or speakers; and background screen colour of grey or black). These minute alterations of the stimuli were not considered by Shams and colleagues as significant variables, hence there was little discussion of their effects or interactions.

Brief discussion of the visual stimulus suggested that the illusory percept is stronger when the visual dots are placed in the periphery of the visual field rather than fovea (Shimojo, Scheier, Nijhawan, Shams, Kamitani, and Watanabe, 2001). However, this variable has not been developed such as exact spatial location in 360° peripheral vision; the number of (and multiple) dots presented; nor correlations with perceived auditory spatialisation.

A broad examination of research into auditory perception, reveals that rhythm of stimuli and timescale, may be important variables. Research has investigated the duration of the stimulus gap (between auditory and visual stimulus) and its length before illusory fragmentation (Shams, 2002). Whilst this research explores the elementary formations of rhythm, research has not investigated the gap size that would cause perceptual fusion between the two stimuli; nor has it investigated the combination of various durations to form rhythmic motifs, and the potential for auditory rhythm motifs transpose to create visual rhythms. Shams (2005) did note that
pilot research…

> *"confirmed that whether beeps and flashes were presented simultaneously or with slight temporal offset made little difference to behavioural reports of illusory perception"* (Pg 1923).

To have confidence in this important conclusion, there is a need for closer and systematic examination of the spatial disparity and duration of the stimulus gap for both large and short durations and their effect on fusion.

Manipulation of Fine-Grained Time Scale

Time scale is an important variable that has been employed and manipulated in many illusory paradigms. The capture of sensory information is accrued by the provision of timescale; therefore, misperception sometimes occurs from insufficient time to acquire sensory information. Employment of the micro time scales as a variable includes Microsounds (Roads, 1996; 2001) with stimuli generated at 600ms or less; the Octave / Scale / Chromatic Illusion (Deutsch, 1981; 1975) generate auditory stimuli at 250ms durations; The Illusory Continuity of Tones (Bregman, 1999) generates noise at 50ms or less; and the Auditory Driving of Visual Flicker (Shipley, 1964) generates auditory stimuli at 150ms. The micro time scale and limited duration of the auditory stimulus employed by the Sound Induced Illusory Flash is an important variable, as the illusion fragments when the beeps expand in duration from 70ms onwards to 100ms.

The structural variable of rhythm has been explored in some depth in current research on the Sound Induced Illusory Flash, although expansion of this variable would be problematic as the micro time scale of auditory stimulus presentation is too short to accurately perceive rhythm. Further, participants would require musical knowledge of simple rhythms to articulate the visual rhythms perceived.

Overall, significant gaps exist in the research of structural variables that may affect the Sound Induced Illusory Flash. Further exploration of the illusion is necessary to provide a greater understanding of the illusion and auditory-cross modal integration. More importantly, and for the present context, it offers potential for creative applications, and new multimedia techniques.

One way to explore the illusion is to examine other possible variables that have been manipulated in the generation of other uni-modal and cross-modal illusions that might be applied to the Sound Induced Illusory Flash.

Frequency and Pitch interval

Frequency as a stimulus variable (as apposed to the sensory dominant and intensively researched rhythm) for cross-modal manipulation was introduced by Marks' exploration on the Mediation of Brightness, Pitch, and Loudness (1974). This psychophysical research does not exhibit cross-modal manipulation, but cross-modal associations between auditory and visual stimulus (greyscale hue and pitch). However, the employment of pitch as a variable has been effective in uni-modal auditory illusions.

The Octave Illusion (Deutsch, 1974) (Deutsch, 1983), Scale Illusion (Deutsch, 1975) and the Chromatic illusion (Deutsch, 1987), pits the perceptual grouping principles of similarity of frequency and spatialisation against one another resulting in an illusory percept based on pitch proximity. Both of these variables – pitch proximity and spatial location – may translate to cross-modal illusions where the variables influence the judgement of visual stimuli and their effect on visual perception assessed.

Harmonic Relationship

The variable of intervallic close harmonic relationship is used to distort motion direction of a consecutive pitch. This variable is most notably exploited by the Tritone Paradox (Deutsch, 1986) that employs multiple layered frequencies at the octave; the Shepard Tone (Shepard, 1964) (Risset, 1972) (Risset, 1986) employs multiple layered frequencies at the octave or the augmented 5th that is further combined with variables of phasing and auditory fixation points.

There is a significant gap in current research on the application of pitch interval and spatialisation in cross-modal auditory visual perception. This gap is reflected in studies of the Sound Induced Illusory Flash that only employs the variables of micro time scales with allusions to rhythm, to distort visual temporal perception, with no investigation into pitch interval and spatialisation variables.

The vast majority of auditory illusions use rhythm and micro time scales as variables to manipulate perception. However, pitch interval and spatialisation were used for purposes of articulation, the generation of perceived motion and spatialisation, therefore there application as variables to articulate visual rhythm, create visual motion and spatialisation should translate to the Sound Induced Illusory Flash.

Experiment Design

Hypothesis

It is hypothesised that the application of pitch interval to the Sound Induced Illusory Flash generates two auditory fixation points corresponding to the high and low pitches (reflective of the Octave illusion), articulating the apparent dot to flicker accordingly and rhythmically, whilst emphasising the illusory effect.

An interval of an octave is employed, as this interval is the closest harmonically to the unison, and therefore the most conservative option for manipulation of the variable.

The application of the variable of transmission from monaural to binaural presentation of the auditory stimulus is hypothesised to further emphasise the two auditory fixation points corresponding to the high and low pitches, articulating the dot to flicker accordingly and emphasising the illusory effect of a single dot. When applied to multiple dots, it is hypothesised that the pitch associated dots will localise according to the spatialisation of the auditory stimuli.

Method

Participants

A sample of 40 participants naïve to the illusion were recruited. They were Psychology 1A students from the University of Western Sydney and received course credit for their participation. Participants were aged between 17 and 54 years (M = 21.175, SD = 6.68), with more female participants than male participants (36 female, 4 male). People reporting a hearing impairment, visual impairment (corrected to normal vision), severe migraines or epilepsy were excluded from testing.

Stimuli

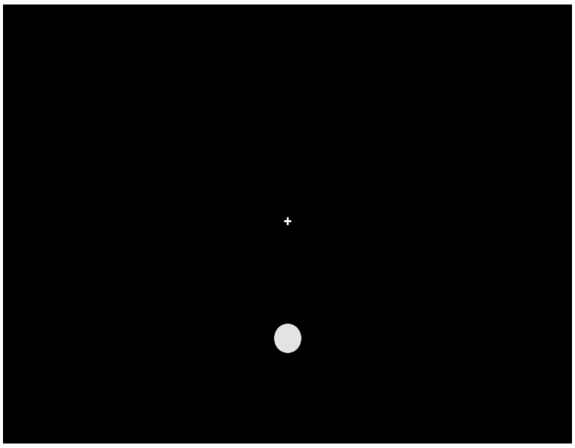


Figure 2. *Screen Capture of the visual stimulus. The fixation point is the centrally located cross, with the dot presented.*

The visual stimulus consisted of a centrally located fixation point, and a single white dot positioned below the centre of the screen that is located in the participants' peripheral vision. In Illusory trials the dot was presented 23ms after the auditory stimulus onset and for the duration of 17ms. The visual stimulus remained the same in the trials, with the independent variables concerning only the aural stimuli.

The auditory stimulus consisted of a sine tone generated every 50ms and lasted for a total of 7ms (attack 2ms, sustain 3ms, decay 2ms).

The variables manipulated were:

- The number of auditory stimulus beeps, presented at two, three, four and five generations.
- Intervallic frequency with the Unison (261Hz) versus the Octave (261 and 523Hz).
- Transmission of the sound through headphones with monaural versus binaural.

The interval variable (unison versus octave) was presented within subjects and the transmission variable (monaural versus binaural) between subjects.

Beeps	Unison Interval		Octave Interval	
Presented	Monaural	Binaural	Monaural	Binaural
2	X	X	X	X
3	X	X	X	X
4	X	X	X	X
5	X	X	X	X

Table 1. Experiment Conditions *Table.*

Yielding 16 conditions in total, each condition was presented 6 times. Catch trials were presented with equal chance of the stimulus being illusory or not. The catch trials consisted of the presentation of physical flashes equivalent to the number of auditory beeps, eg. two physical flashes with two auditory beeps; three physical flashes with three auditory beeps; four physical flashes with four auditory beeps; and five physical flashes with five auditory beeps. In catch trials, the dot was presented every 50ms and simultaneous with the auditory stimulus onset.

Equipment

Participants were located at computer workstations with their head positioned on a chin rest 40cm from the computer monitor and eye level with the fixation point. A Mac Pro G5 with a Diamond digital CRT monitor was used with sound transmitted through AKG K601 headphones. MAX / MSP was used to construct an application that generated the auditory and visual stimulus; presented the trials in a randomised and collected order (using the urn object); generated the questionnaire; and collected the participants responses in a text file.

Procedure

Participants were instructed to place their head on the chin rest and focus on the fixation point. The dot would be presented but to use their peripheral vision to count the number of times it was presented.

The task required them to state on a multiple choice questionnaire the number times the dot was presented, ranging from one event to five events. Further, with a time limit of 8 seconds to respond, participants had to state their confidence level of their decision, ranging from not confident, unsure, very confident.

Results

The experiment recovered the illusory effect with results suggesting that pitch interval and binaural transmission enhanced the illusory effect.

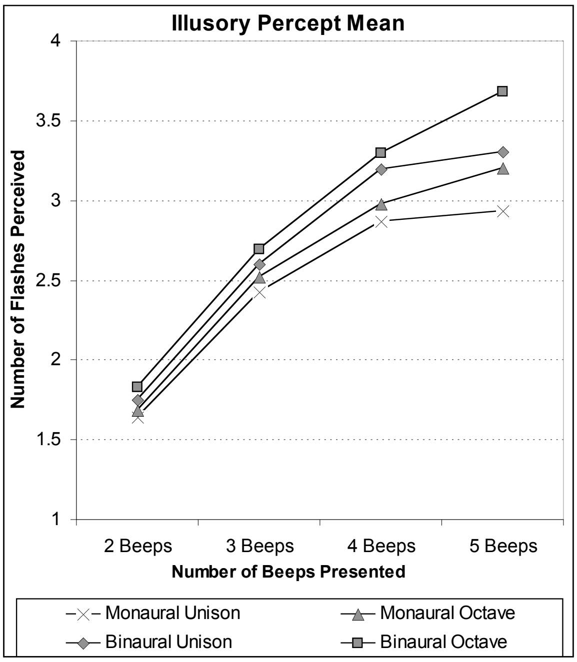


Figure 3. Graph illustrating the means of the perceived number of Flashes.

Participants had a greater performance accuracy mean for two beep presentations, that continued although degraded through to five beep presentations. This degradation of performance accuracy mean was also present in the catch trials, which suggest attentional issues.

Over the two to five beep presentations, the binaural octave condition consistently provided the highest mean, followed by the monaural octave and binaural unison conditions, with the monaural unison condition providing the lowest mean.

These results indicated that modifying the auditory stimulus with the octave interval and binaural transmission enhanced the illusory effect, suggesting that the contrast in auditory stimulus draws attention to, and articulates the auditory rhythm.

References

Alais, David. and Burr, David. 2004. "The Ventriloquist Effect Results from Near-Optimal Bimodal Integration" *Current Biology*. 14: 257 - 262.

Bhattacharya, Joydeep. Shams, Ladan. and Shimojo, Shinsuke. 2002. "Sound-Induced Illusory Fash Perception: Role of Gamma Band Responses" *Cognitive Neuroscience and Neuropsychology*. 13: 14.

Blauert, Jens. 1996. Spatial Hearing - Revised Edition: The Psychophysics of Human Sound Localization. The MIT Press.

Bregman, Albert. 1999. *Auditory Scene Analysis*. Cambridge Massachusetts: The MIT Press.

Chambers, Christopher D. and Mattingley, Jason B. 2002. "The Octave Illusion Revisited: Suppression or Fusion Between Ears" *Journal of Experimental Psychology: Human Perception and Performance*. 28: 6. 1288–1302.

Deutsch, Diana. 1998. *The Psychology of Music*. Second Edition. California: Academic Press.

Deutsch, Diana. 1974. "An Auditory Illusion" *Nature*. 251: 307-309.

Deutsch, Diana. 1975. "Two-channel listening to musical scales" *Journal of the Acoustical Society of America*. 57: 1156-1160.

Deutsch, Diana. 1978. "Delayed Pitch Comparisons and the Principle of Proximity" *Perception and Psychophysics*. 23, 3: 227-230.

Deutsch, Diana. 1981. "The Octave Illusion and Auditory Perceptual Integration" *Hearing Research and Theory*. 1: 99-142.

Deutsch, Diana. 1983. "The octave illusion in relation to handedness and familial handedness background", *Neuropsychologia*. 21: 289-293.

Deutsch, Diana. 1988. "The Semitone Paradox" *Music Perception*. 6, 2: 115-132.

Deutsch, Diana. and Roll, Philip L. 1976. "Seperate "What" and "Where" Decision Mechanisms in Processing a Dichotic Tonal Sequence" *Journal of Experimental Psychology: Human Perception and Performance*. 2, 1: 23-29.

Marks, Lawrence E. 1974. "On the Associations of Light and Sound: The Mediation of Brightness, Pitch, and Loudness", *American Journal of Psychology*. Vol 87: 173-188.

McGurk, Harry. and Macdonald, John. 1976. "Hearing lips and seeing voices" *Nature*. 264: 746 – 748.

Müller-Lyer, F.C. 1889. "Optiche Urteilstäuschungen", *Dubois-Reymonds Archiv für Physiologie*. 7: 263–270.

Risset, Jean-Claude. 1972. *Musical Acoustics*. Paris: IRCAM.

Risset, Jean-Claude. 1986. "Pitch and rhythm paradoxes: Comments on 'Auditory paradox based on fractal waveform'" *The journal of the Acoustical Society of America*. 80, 3:961-962.

Roads, Curtis. 1996. *The Computer Music Tutorial.* Cambridge, Massachusetts: The MIT Press.

Roads, Curtis. 2001. *Microsound.* Cambridge, Massachusetts: The MIT Press. 86-106, 330-351.

Shams, Ladan. 2002. "Integration in the Brain - The Subconscious Alteration of Visual Perception by Cross-Modal Integration" *Science and Consciousness Review*. 1: 1-4.

Shams, Ladan. 2005. "Sound induced flash illusion as an optimal percept" *Neuroreport.* vol 16: 17.

Shams, Ladan. Iwaki, Sunao. Chawla, Aman. and Bhattacharya, Joydeep. 2005. "Early Modulation of Visual Cortex by Sound: an MEG Study" *Neuroscience Letters*. 378: 76–81.

Shams, Ladan. Kamitani, Yukiyasu. and Shimojo, Shinsuke. 2002. "Visual Illusion Induced by Sound" *Cognitive Brain Research*. 14: 147-152.

Shams, Ladan. Kamitani, Yukiyasu. Thompson, Samuel. and Shimojo, Shinsuke. 2001. "Sound Alters Visual Evoked Potentials in Humans" *Cognitive Neuroscience and Neuropsychology*. 12, 17: 3849-3852.

Shepard, Roger. 1964. "Circularity in Judgements of Relative Pitch" *The journal of the Acoustical Society of America*. 36, 12: 2346-2353.

Shimojo, Shinsuke. Scheier, Christian.Nijhawan, Romi. Shams, Ladan. Kamitani, Yukiyasu. and Watanabe, Katsumi. 2001. "Beyond Perceptual Modality: Auditory Effects on Visual Perception" *Acoustic Science and Technology*. 22, 2: 61-67.

Shimojo, Shinsuke. and Shams, Ladan. 2001. "Sensory Modalities Are Not Separate Modalities: Plasticity and Interactions" *Current Opinion in Neurobiology*. 11: 505-509.

Shipley, T. 1964. "Auditory Flutter Driving of Visual Flicker" *Science.* vol 145: 1328-1330.

Ros Bandt
The Australian Centre
The University of Melbourne
137 Barry St,
Carlton, Vic, 3055
Australia
Email: r.bandt@unimelb.edu.au

From Bridges to BYOS: 10 years of Global and local electro-acoustic collaborations by Ros Bandt (Australia), with Johannes S. Sister-manns (Germany)

Artist Talk

Abstract

This paper traces the relationships of place, culture and identity in the shared electro-acoustic collaborations by Ros Bandt (Australia) and Johannes S. Sistermanns (Germany) over a ten year period. Two works A Global Bridge for Percy *for the Melbourne International festival in 1997 and the recent* BYOS, bring your own sounds, *2006, a collaborative work co-composed in the Melbourne studios of the ABC provide a platform for discussion of the collaborative process involving various degrees of freedom and control and decision making by both electro-acoustic composer performers*

Introduction

This artist talk considers the ten years of electroacoustic composition in relation to two collaborative works spanning the ten year period 1997-2006. It investigates the local worlds of both composer performers and sound artists, the Australian Ros Bandt and the German Johannes S. Sistermanns who met in Australia at the Transit Zeitglieich festival in 1993. They have shared their independent soundart practices in Australia, and Germany, performed in each others territory, and globally in the digital realtime broadcasting of works performed in the different nations and heard across the world in realtime, from Melbourne's Percy Grainger Museum to the Frankfurter Hochschule for Musicwissenscahft und Darstelle Kunst. The digital documentation of these works is a postscript of these visceral and virtual events.
http://www.abc.net.au/arts/lroom/globgar.htm

Bandt,(Sounding Spaces) and Sistermanns, (Soundplastic) have shared their local domains and ability to transcend and transform these through their international co-operative ventures. Sound Installation, soundplastic, performance and electroacoustic composition are components of the wider philosophical and artistic aims of each project, one Global, spanning Frankfurt and Melbourne in real time,
http://www.abc.net.au/arts/lroom/gardel.htm
from Melbourne's Percy Grainger museum to the Musikhochschule Frankfurt where Grainger studied , the other an intense 3 days in the ABC Melbourne studios with an engineer, each artist providing sonic material unseen by the other to share.

For each work Ros Bandt provides a graphic cartography of sound files, dense processes and temporal designs involved including flow diagrams, spatial mapping and scores. Both works were commissioned and recorded by the ABC. *A Global Bridge for Percy* is 50 minutes and can be heard on the above website, while *BYOS*, 16 minutes, a radiophonic work will be broadcast by ABC's New Music Uplate shortly. and can be heard in the audiotech here at the ACMA conference. Audible statements from each composer have been commissioned by Robyn Ravlich from audio arts about the process of BYOS and these will be broadcast as part of the programme.

Questions are posed, How do you share the concept of a work? Who does what and when? How do you agree on the many levelled sonic parameters for each sound and the more formal flow and morphology of events? What is the role of the technician/s in relation to the artistic directors? In what sense is the outcome of BYOS and Bridges a collaborative work? Who owns it?

How do the different cultures affect the possibilities of artistic creation? How is aesthetic collaboration arrived at?

The author is pursuing these questions in the future against a range of cultures as she leaves for Turkey later this year as the Tarhu nomad to investigate live electronics using audio mulch with Erdem Helvacioglu.

References

Bandt, Ros. 2001. "A Global Bridge for Percy", *Sound Sculpture: Intersections in Sound and Sculpture in Australian Artworks*, Craftsman House, pp132-140. Also see CD track 30

The cross cultural bowed spike fiddle:Tarhu www.spikefiddle.com

Transmission

David Kirkpatrick
djk_1200@hotmail.com

Paul Kopetko
paulkopetko@gmail.com

Artist Talk

Transmission is a collaborative computer music piece which attempts to cross the boundaries between the rigid, static nature of pre-recorded electronic music and the improvisation of live, in-the-moment performance. This is achieved through the design of a new computer-based musical instrument which was premiered at a University of Wollongong concert on the 1st of June 2007.

In *Transmission*, a computer tracks the movements of two performers on stage, whose physical location and bodily gestures are used to capture sounds being broadcast over the FM radio spectrum. These captured radio sounds are mixed with an array of pre-recorded sounds also triggered by gestures. The appropriation of FM radio sound during the performance is key to the improvisational aspect of the work, as the sonic 'found objects' which are captured will not only be different in every performance, but also virtually impossible to predict beforehand. It is thus up to the performers' musical intuition to capture and treat suitable ranges of live radio to combine with improvised selections from a bank of gesturally-selected pre-recorded music. In using the body to control a mergence of pre-composition and improvisation, a piece of computer music is created that bridges the gap between composer, performer and machine, begging the question of where one entity ends and the other begins.

The technical setup for the piece is as follows: two performers are monitored from above by a black and white CCTV camera. The output from the CCTV camera is connected to our computer system through a *Falcon Frame Grabber* PCI card, which then sends the video footage in real-time to Frieder Weiss' *EyeCon* software (recently used in a Sydney Opera House performance entitled *Glow* by the dance company *Chunky Move*). Our self-designed *EyeCon* patch detects the two performers and tracks their positions and movements, forwarding this data to *Pure Data* via *OSC*. The position information of the first performer is used by *Pure Data* to control the frequency scanner of a *Radio Xtreme USB FM radio tuner*, so that as the performer moves left and right along the stage, the tuner moves left and right along the FM frequency spectrum. More in-depth position and gestural information sent to *Pure Data* is converted into MIDI signals and routed to *Abelton Live*, which applies digital filtering effects to the radio signal in real-time. The second performer's movement and gesture cues are similarly sent through *Pure Data* to *Abelton Live* in order to trigger and process a bank of pre-recorded clips.

In our presentation at the conference we will discuss the artistic principles behind *Transmission*, as well as the technical process used to develop the computer-based instrument. We shall display documentation of our performance and propose ways in which a similar process could be used in future to create a virtually unlimited array of works encompassing sound, light and projection.

Life After the Cliff

Tim Kreger
tkreger@bigpond.net.au

Artist Talk

After 15 or so years of existing within the so-called Ivory Towers of academia I made the decision to leave the compound. For me the decision to do so was akin to jumping off a cliff, a cliff so high I couldn't see what was at the bottom. The last five years of my academic career was fraught with the fear of eventually being pushed off the cliff involuntarily so the decision to actually take the jump of my own free will was both liberating and to some degree empowering. I would like to outline my professional practice since leaving the university system and how many of the skills acquired within academia have assisted with my career as a freelance musician/programmer/etc. I would also like to discuss the competing imperatives of both academic and commercial environments. In many contexts the imperatives are the same however there are several which require radically different approaches.

I would like to do this with reference to the projects I have been involved in, to outline the varied array of activities and the challenges each project has presented.

Live Coding: New perspectives on generative art

Andrew Sorensen
MOSO Corporation
3/35 Kate St
Alderley, 4059
Australia
andrew@moso.com.au

Artist Talk

Live Coding is a relatively youthful performance practice that explores the definition and manipulation of process in a real-time improvisatory setting. To date, computer programming languages provide the most powerful formal languages which humanity has yet devised for the description and manipulation of process. It is these computer programming languages that Live Coding practitioners use to create and perform musical/visual works in a live performance setting. Typically practitioners build and manipulate generative structures during a performance, leaving these structures to "perform" while the performer creates new structures, either to replace or "perform" in counterpoint with existing material. Generative structures often start simply and are subsequently expanded and modified by the live programmer as the performance unfolds. Live programming usually involves the projection of the code enabling the audience to engage with the construction of the works internal processes.

For the past two years I have focused on Live Coding performance practice, working on the tools and ideas that allow me to explore code as a medium for real-time audio/visual expression. My current live coding practice is centred around base notions of periodicity, linearity, probability and recursion. These processes will be well known to many of you and have been fundamental to generative music making for many decades. The real-time constraints imposed by Live Coding has encouraged me to reengage with these base functions, providing an intimacy of process and a connection with medium, often lost through higher level abstraction.

This talk will discuss some of these ideas and provide practical demonstrations of some aspects of the discussion.

David Worrall
composer

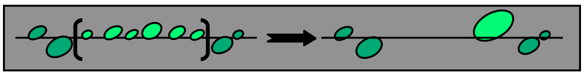
Canberra, Australia
worrall@avatar.com.au

Artist Talk: Sonification, Synthesis and SoniPy

1. Data sonification

One way of classifying data sonifications is according to where they are placed on a continuum with free data transformation on one end to perceptually linear mapping on the other.

Closer to the 'free data' end of the continuum would be the use digital documents as control data for some sound synthesis routines as arbitrarily determined by the sonifier. It is likely that the (possibly imagined) content of the documents would play a role in culturally contexturalising the resulting soundscape.



My current work sits closer to the other end of the continuum. I'm sonifying high-frequency intraday stockmarket data as delivered by a securities exchange trading-engine. Whilst there are utilitarian uses for such sonifications to assist decision-making participants and regulators, my primary concern is a cultural one: to provide a better 'feel' for market exchanges, which have increasingly infiltrated our lives at the same time as their operation has become more virtualised.

This work is undertaken as a PhD in data sonification at The University of Canberra, funded by them and the Capital Markets Cooperative Research Centre in Sydney.

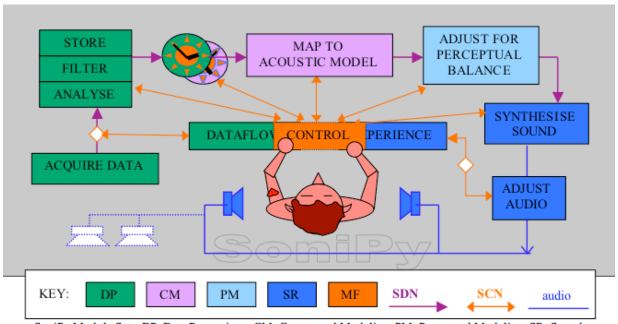

SoniPy, the singing serpents!

2. The SoniPy Project

As data sonification is an interdisciplinary activity, it is not surprising that a diverse range of research and production tools is required. My forays into the sonification of High Frequency Data (HFD) made it clear to me that, in general, music composition software was inadequate for the task of data sonification once handling the data itself became critical, as it does in the timely handling of multi-dimensional or multiplexed streams.

I'll further expand on the reasons for the these limitations and the criteria I established for finding, testing and integrating software components that can work together is a time-consuming and somewhat frustrating task.

The Sonipy Project aims to collect a set of tools by collating public-domain Python modules that are suitable for the purpose and integrating them using a modular framework. In doing so, the Project aims to provide a community resource for individual sonifiers who don't have the time, skill or inclination to independently go through the process of putting together such a coherently-functioning toolset.



SoniPy Module Sets: DP: Data Processing. CM: Conceptual Modeling, PM: Perceptual Modeling, SR: Sound Rendering, MF: Monitoring and Feedback, SDN: SoniPy Data Network, SCN: SoniPy Control Network.

A full discussion of the motivations and concepts involved is discussed in the ICAD 2007 paper[1] which is available for download from the SoniPy website: http://sonipy.sourceforge.net/.

SoniPy is established as an open-source community-based project. It is just beginning and contributions are welcome.

[1] Worrall, D. et al. (2007) *SoniPy: The design of an extendable software framework for sonification research and auditory display*.

Stephen Barrass

Sonic Communication Research Group
University of Canberra
Canberra, ACT 2601
Australia

stephen.barrass@canberra.edu.au

Studio Report 2007

Sonic Communications Research Group (SCRG)

<www.scrg.org.au>

About

The Sonic Communications Research Group (SCRG) is an interdisciplinary group of researchers and educators working with sound as a medium for creative communication.

The group has a practice-led approach that seeks to crossover and integrate theory from media studies, design, information science, psychology and the creative arts.

Our special expertise includes sonification, sonic-writing, haptic-audio, auditory interfaces, generative sound, music imagery, music cognition, auditory perception, and sound art, theory and practice.

SCRG is a participant in the ARC Research Network on Human Communication Sciences. <www.hcsnet.edu.au>

People

Faculty
A/Prof. Stephen Barrass (Co-Director, UC)
Prof. Roger Dean (Co-Director, UWS)
Prof Hazel Smith (UWS)
Dr. Mitchell Whitelaw (UC)
Dr. Sam Hinton (UC)

Postdoctoral Fellows
Dr. Freya Bailes (UC)

Postgraduate Students
David Worrall
Dempsey Chang
Robert Bell

Research Assistants
Michael Bylstra

Alumni
Michael Bylstra (hons 2006)
Somaya Langley (hons 2006)

Projects

Sonification of the Capital Markets
Worrall/Dean/Barrass/Whitelaw
A pilot projectwith the Capital Markets Collaborative Research Centre (CMCRC).

SoniPy – Worrall/Bylstra/Barrrass/Dean
An open source environment for data sonification research and auditory displays.
sonipy.sourceforge.net

Percussive Instrument Timbre Classification Hierarchy
Bell/Barrass
A system for analysing and classifying the timbres of percussion instruments.
percussionresearch.uniblogs.org

Realtime Computational Generation of Large Scale Musical Structure, ARC Discovery Grant, 2004-6.
Dean/Smith/Bailes
The project aims to develop efficient techniques for the generation and control of large-scale musical structure in real-time, and to use these to assess the degree to which the resultant structures are cognitively accessible to experienced musicians.

Cloud Musak - Bylstra/Barrass
An ambient sonification of clouds and sky taken from webcam and internet weather stations.
cloudmusak.blogspot.com

The Suspect Backpack - Langley/Whitelaw
A wearable mobile sonic media art experience, providing first-person interaction and engagement in public and private spaces. This [prototype] work utilises proximity sensors, a laptop, a pair of speakers and a set of headphones to navigate two spoken-word sound environments.
suspectbackpack.blogspot.com

Audio-Visual Augmented Reality (AVIARy)
– Barrass/Morse
An Augmented Reality system for moulding musical and visual gestalts from algorithmic materials layered in space and time.
http://www.petermorse.com.au/old/PMcomhtml/petermorse.html

Creative Works

These are just a sampling of performances, installations and other other creative audio works by members of SCRG in recent times. There are significant works that I haven't yet collated in time for this studio report. I will have a more complete collection to show at ACMA.

Langley S. (2007) The Suspect Backpack, Dorkbot, Melbourne, May 2007 criticalsenses.blogspot.com

Dean R. (2006) Sonic Intermedia Performance, Cultural Studies Association of Australasia Annual Conference, December 2006. www.unaustralia.com/speakers.php

Barrass, S. (2006) OpShop, Experimenta Under the Radar, Institute for Contemporary Art, London, December 2006 www.ica.org.uk/Experimenta%20Under%20the%20Radar+12289.twl

Barrass, S., Bylstra M. and Kuzmanovic, M. (2006) Fruit Turntable Custard, S-crepe Happenings, Canberra, January 2006 www.criticalsenses.com/screpe

Whitelaw, M. and Gaffney N. (2006) Boom, S-crepe Happenings, Canberra, January 2006 www.criticalsenses.com/screpe

Barrass, S. (2005) Augmented Reality Crazy Frog phone-tone remix, Bimbimbie Sound and Intermedia Performance, University of Canberra, October 2005

Dean, R. and Smith, H. (2005) Time the Magician, Bimbimbie Sound and Intermedia Performance, University of Canberra, October 2005

Whitelaw, M. (2005) Arcadia Sonic Mosaic, Sound and Intermedia Performance, University of Canberra, October 2005

Barrass, S., Riddell, A., Fitton, A., Barrass T., Morse, P., and Onaclov (2005) Edible Audience, Liquid Architecture 6, National Gallery of Australia, Canberra, July 2005. www.nga.gov.au/LiquidArchitecture

Publications

Bailes, Freya (2007). A response to Andrea Halpern's commentary. *Empirical Musicology Review*. 2(2), 62-64.

Barrass, Stephen and Whitelaw, Mitchell and Bailes, Freya (2007). Listening to the Mind Listening: An Analysis of Sonification Reviews, Designs and Correpondences. *Leonardo Music Journal*. 16, 13-19.

Bailes, Freya (2007). Timbre as an elusive component of imagery for music.. *Empirical Musicology Review*. 2(1), 21-34.

Barrass, Stephen, Barrass, Tim (2006). Musical creativity in collaborative virtual environments. *Virtual Reality*. 10, 149--157.

Barrass, Stephen, Whitelaw, Mitchell, Potard, Guillaume (2006). Listening to the mind listening. *Media International Australia incorporating Culture and Policy*. 2006, 60-67.

Dean, Roger, Whitelaw, Mitchell, Smith, Hazel, Worrall, David (2006). The mirage of real-time algorithmic synaesthesia: Some compositional mechanisms and research agendas in computer music and sonification. *Contemporary Music Review*. 25, 311--326.

Barrass, Stephen (2006). Haptic-Audio Narrative: From Physical Simulation to Imaginative Stimulation. Springer *Lecture Notes in Computer Science*. 4129, 157-165.

Barrass, Stephen (2005). A perceptual framework for the auditory display of scientific data. *ACM Trans. Appl. Percept.*. 2, 389--402.

Bailes, Freya, Dean, Roger T. (2005). Structural Judgements in the Perception of Computer-Generated Music. *2nd International Conference of the Asia Pacific Society for the Cognitive Science of Music* (APSCOM). 155-160.

Whitelaw, Mitchell (2005). System Stories and Model Worlds: A Critical Approach to Generative Art. *Readme 100: Temporary Software Art Factory*. 135--154.

Whitelaw, Mitchell (2004). Hearing Pure Data: Aesthetics and Ideals of Data-Sound. *Unsorted: Thoughts on the Information Arts: An A to Z for Sonic Acts X*. 45--54.

Whitelaw, Mitchell (2004). *Metacreation: Art and Artificial Life*, MIT Press.

Nesbitt, K. V., Barrass, S. (2004). Finding trading patterns in stock market data. *Computer Graphics and Applications*, IEEE. 24, 45--55. Abstract

Whitelaw, Mitchell (2003). Sound Particles and Microsonic Materialism. *Contemporary Music Review*. 22, 93--100.

Barrass, Stephen (1998). Auditory Information Design. Ph.D. Thesis. Australian National University

Greg Schiemer
Fazel Naghdy
Mark Havryliv (APAI)
Timothy Hurd
University of Wollongong
Australia
mhavryliv@gmail.com,
{fazel, schiemer}@uow.edu.au

Studio Report

Abstract

The Haptic Carillon project is an ARC Linkage (APAI) with Industry partners, the National Capital Authority and Olympic Carillon International. Our objective is to create an electronic practice clavier with the sound and feel of a real carillon. This will solve a problem that has plagued carillonists for centuries, namely, the inability to practice their instrument in private. Progress has been made synthesising haptic characteristics of the traditional clavier. We propose to present a comparative demonstration of the haptic and original carillon clavier. This will happen in the clavier chamber of the National Carillon, Aspen Island on Lake Burley Griffin.

Studio Report

Warren Burt
Illawarra Institute of TAFE
PO Box U-27
Wollongong
NSW, 2500

Abstract

At the Illawarra Institute of TAFE, under the leadership of head teacher Michael Barkl, we have set up a series of computer labs and teaching facilities for music technology that allows us to teach a wide variety of computer music skills in a number of contexts. A professional recording studio environment allows us to teach traditional multi-track recording skills, while a series of individual recording rooms allows students to work on individual projects. Portable recording equipment allows students to record both concerts and environmental sounds. A lab with 19 computers fitted out with a wide variety of software enables us to teach not only vocational sound skills, but advanced computer music concepts as well. The studio setups allow us, while adhering to the VET (vocational education and training) syllabi of the TAFE courses, to also move beyond them, giving students professional experience at the cutting edge of developments in music technology.

Australian Sound Design Project Interactive Map and Website CD launch

Ros Bandt
Australian Sound Design Project.
The Australian Centre
The University of Melbourne
137 Barry St
Carlton 3053
Australia
www.sounddesign.unimelb.edu.au
r.bandt@unimelb.edu.au

Abstract

The Melbourne City recently commissioned the Australian Sound Design Project to document soundoing artworks in Melbourne's CBD. Ros Bandt and Iain Mott documented the following works and arranged them on an interactive map for ease of access for the City. A cross platform CD of the website including this new map is being launched nationally at this national computer music event.

Introduction

THE AUSTRALIAN SOUND DESIGN PROJECT: UPDATE

The Australian Sound Design Project continues to lead the University of Melbourne in pioneering sound research and making it audible online. The site has been showcased and exhibited as one of the University's most innovative projects in annual design innovation fests. It has appeared in publications as an example of innovative sound research.

During 2006-7 the website has grown to include 147 works in its online digital gallery, with 98,478 hits on the site during the year. Most activity was from Australia followed by the Netherlands, Japan, France and the USA Educational. The faculty of Arts awarded Dr Ros Bandt the excellence for research Award for 2006 which featured this innovative digital publishing along with her other books, writings and original sounding artworks.

City of Melbourne Collaboration and Commission

The Australian Sound Design has completed a major new commission for the Melbourne City Council publishing seventeen new sound works that fall under their custodianship. These include, a) newly commissioned public sound works, b) works from the Council's collection and c) works in the Melbourne CBD. The online publication will encourage sound to be recognised as an important part of cultural heritage. Iain Mott and Ros Bandt documented the works on site and liaised with the artists to publish the works in image, sound video and text. Iain Mott developed an interactive map as a portal to these works. You can see the works in site on a map of the City of Melbourne and zoom in to the detailed documentation directly for each work.

Browse works on map of Melbourne. At

http://www.sounddesign.unimelb.edu.au/site/worldkit/index.html

Be sure to go to the digital gallery for each site.
The 17 CBD works newly documented are

- Birrarung Wilam 2006-
- City Talking 2006
- Federation Bells 2002-
- Lie of the Land 1997-
- Life Coach 2006-2007
- Musical Decking 2002
- No Answer 2006-2007
- Proximities2006-
- Reeds 2000
- Rice Paddies 2001
- Scar-A Stolen Vision 2001
- Southgate Soundscape 1992-2006
- Spine 1.2 2006
- Transient Frequencies
- Tilly Aston Bell 1999-
- Wall Piano 2002
- Weather Harp 2003-2006

The Interactive Map was designed by Iain Mott using worldkit and an aerial map supplied by the Melbourne City Council. Red Circles are used to locate the fifteen works which are currently audible and red squares for the two decommissioned works for which we have documentation. This is an important archival facility as well as an information portal for visiting current works. By using the zoom buttons, one can travel down to see the position of the work in the site. By clicking on the red circles and squares you can navigate to the full documentation of each work directly seeing images, reading artist's statements and listening to the sound and watching videos.

All features of the search engines of the sound design site can be accessed for cross referencing by date, work, artist, site, title, location, stylistic feature.

This map is now being used in subjects at the Australian Centre *Australia Now* and for the VCA *Sound Art in the Environment at the Centre for ideas*.

Research and Development of the ASDP Site

In 2006 the site was recommended for inclusion in the National Research Network for Data Sustainability. At the University of Melbourne, Gavan McCarthy—director of Austehc and the team who invented the OHRM data archival methodology—is now using the ASDP site data to trial innovative aspects of the newest upgrade of the OHRM. The site is currently being reconfigured in OHRM 5 sponsored by Austehc.

In 2004-5 the National Library mined the site data for its lasting value through its *Music Australia* project, contributing $6000 to have the information transposed into a meta-data compatible format with the National Library system.

Credits

I would like to thank the City of Melbourne and the Australian Centre at the University of Melbourne for facilitating this work and especially am indebted to my former research assistant, Iain Mott who has worked tirelessly to produce the map and accompany me on the bicycle route to carry out the field work audio-visual documentation on site around the City of Melbourne over the past year.

Michael Bylstra
Sonic Communications Research Group
University of Canberra
University Drive
Bruce ACT 2617
Australia
mbylstra@digitalfeast.com.au

Cloudscape Radio: The Ambient Sonification of Clouds

Poster

The everyday auditory experience for the city dweller is one that is becoming more and more removed from that of a natural setting. The auditory environment of an office, for example, is simple, repetitive and monotonous, and lacking the elements of complexity, variation and flux found in nature. Sonification, however, can be used to collect data from nature and present it aurally, in order to create new acoustic environments as rich as natural ones. Sonifications of nature data can both facilitate a better understanding of nature and provide new aesthetic experiences. Examples of nature data sonification include Garth Paine's *Reeds* (Paine), a sonification of local weather data, Carrie Bodle's Listening Up (Bodle), a sonification of upper atmosphere ions and electrons, and Andrea Polli's *Atmospherics* (Polli), a sonification of storm simulation data.

The complexity, beauty and variety of form that clouds possess inspired the idea of transforming live cloud images into sound – to create a sonic window to the outside world for indoor environments. By installing a live cloud sonification system as an ambient sound system in a place where we carry out our daily affairs, we can maintain a constant awareness of the activity of the natural world which would not be possible visually. Rather than simply eliminating noise pollution and creating barren acoustic environments, we can mask unwanted noises and create acoustic environments that create a sense of place for the individual. (Wrightson, 2000)

It was reasoned that the experience of listening to clouds should be more akin to listening to natural sounds such as wind, rain or seashore sounds, than to listening to conventional music. It was decided that the sound generated should be abstract, non-representational and should not contain any conventional musical pitch, harmony, tempo, or timbre. It was intended that an understanding of the relationship between the various types of clouds and the various types of clouds could be learnt through experience, without the need of iconic or metaphoric representations.

Cloudscape Radio, a system for the sonification of cloud imagery, was developed where images are captured from a USB Webcam and sonified in real-time. Additionally, the capability of sonifying images from publicly available webcams on the internet was built into the software. A large database of Internet webcams, which were pointed at the skies, was collated, so that users can tune into live sonifications of cloudscapes from all over the world, at any time.

A mapping from image data to sound was chosen where the x-axis is mapped to time, the y-axis is mapped to frequency and the single-channel pixel value is mapped to amplitude (Fig. 1), similar to the mapping used in 'the vOICe' image sonifier (Miejer, 1992). However, rather than synthesizing tones consisting of inharmonically related sinusoids, we chose to synthesize 'coloured-noise (noise with a varying spectral envelope) (Hanna, 2004), where each row is mapped to a narrow-band noise signal. Through research into psychoacoustics, we intended to develop an ideal coloured-noise synthesizer that had the capability of perfectly synthesizing any sound within the set of all perceivable coloured-noise sounds.

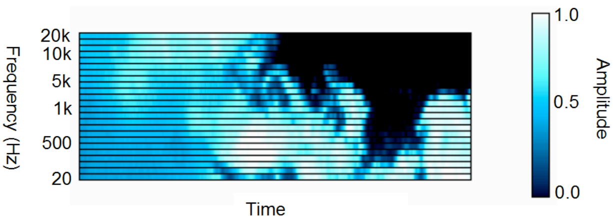


Figure 1. *Mapping from image to sound*

A method for creating a noise signal with an arbitrarily detailed (but static) spectral envelope was developed where a signal containing thousands of sinusoids is pre-rendered using an Inverse-FFT and then looped. By adding a number of narrow-band noise signals generated using this method, coloured-noise with a dynamic spectral envelope can be generated in real-time. Through informal experiments it was discovered that the maximum frequency resolution of hearing for coloured-noise signals is equivalent to the total possible number of concurrent, non-overlapping critical bands, which is 24 according to the Bark scale (Smith, 1999). By controlling the amplitude envelopes of 24 looping, pre-rendered narrowband noise samples, the entire gamut of perceptible coloured-noise signals can be efficiently generated in real-time. By mapping live cloud images to our coloured-noise synthesizer, the complex and dynamic properties of clouds can be preserved to create a rich and constantly evolving acoustic environment.

Installation Requirements

- A PC running Windows XP with the Java Runtime Environment (JRE) 1.4 or greater installed.
- Internet access
- Speakers
- (optional) A USB webcam

References

Andrea Polli, A., & Van Knowe, G. "Atmospherics/Weather works: The sonification of meteorological data", http://www.andreapolli.com/studio/atmospherics/

Bodle, C. "Sonification: Listening up ", http://www.sonifications.com

Hanna, P., Louis, N., Desainte-Catherine, M., & Benois-Pineau, J. 2004, "Audio features for noisy sound segmentation", *Proceedings of the International Conference on Music Information Retrieval,* Barcelona, Spain

Miejer, Peter. 1992. "An experimental system for auditory image representations", *IEEE Transactions on Biomedical Engineering*, 39(2): 112-121.

Paine, G. "Reeds: An interactive installation", http://www.activatedspace.com/Installations/Reeds/

Smith, J. O., & Abel, J. S. 1999. "Bark and ERB bilinear transforms", *IEEE Transactions on Speech and Audio Processing*, 7(6): 697-708.

Wrightson, Kendall. 2000. "An introduction to acoustic Ecology", *Soundscape*, 1(1): 12-13.

Thorin Kerr
thorin.kerr@gmail.com

A Computer Assisted Composition system for a live coding environment

Poster

I am developing a Computer Assisted Composition (CAC) system for use in the Impromptu interactive programming environment, developed by Andrew Sorenson. The Impromptu interface utilises the Scheme programming language, offering the composer/performer a flexible starting point to construct almost anything the Lisp programming language will allow. I am developing a library of algorithmic tools at a 'higher level' of abstraction with a view to assist in the generation of musical structures.

Live coding is the performance practice of creating artistic work by writing and modifying the computer programs that generate the work during the performance. There are emerging aesthetics in the field of live coding which I will need to consider in relation to the development of my CAC system. For example, there is a preference amongst live coders to allow code to be clear enough such that underlying algorithms can be deciphered.

There is also some debate concerning how 'high level' such a system should be. At one end of the 'high level' spectrum, packages with pre-scripted interfaces are generally eschewed by live coders largely because they don't offer access to underlying algorithms, and/or are not transparent in their presentation of those algorithms. At the 'low-level' end, building 'from scratch' with original code can be daunting, and not always advantageous in a performance situation. These are some of the issues I am contemplating in the design of my CAC system.

Any system of compositional tools will ultimately reflect the compositional ideology of the developer. Given this, rather than attempt to create a 'musically neutral' system intended to accommodate existing compositional practices, I am pursuing a compositional strategy which my system is aimed to reflect. My compositional strategy is still being formulated, however one of my concerns is to provide access and control to both micro and macro musical structures. While there are many exceptions, in general, the distinction between micro and macro structures has divided the approach to composition of many CAC's. By offering consistent methods to control musical structure at any level, my system aims to blur the distinction between micro and macro structures.

I am interested in defining the relationship between structures. This has evolved from an ongoing interest in linguistics and 'concrete poetry'. Of particular interest are poetic techniques that are based upon - sometimes ambiguous - relationships between texts. In such cases it is the relationship itself that becomes the basis of the work, as opposed to a specific meaning denoted by the text itself. Employing techniques such as these offers useful possibilities for a compositional approach to musical structures.

In a live coding context my CAC system would offer the composer/performer the potential to construct and organise material on a number of levels at once. While the compositional system employed reflects the interests of the developer, it is hoped that the tools provided are flexible enough, and of interest to other composers who want to realise their own compositional approach in a live performance context.