Harmony

Proceedings of the Australasian Computer Music Conference 2014

Hosted by The Faculty of the Victorian College of the Arts (VCA) and the Melbourne Conservatorium of Music (MCM).

9th – 13th of July 2014

THE UNIVERSITY OF MELBOURNE

FACULTY OF VCA & MCM

acma

Australasian Computer Music Association

Proceedings of the Australasian Computer Music Conference 2014, Melbourne, Australia

Keynote Speakers:
Professor Barry Conyngham A.M.
Paul D. Miller (aka DJ SPOOKY That Subliminal Kid)

Paper Jury:
Roger Alsop
Ted Apel
Jim Barbour
Ross Bencina
Warren Burt
Paul Doornbusch
Richard Dudas
Angelo Fraietta
David Hirst
Matt Hitchcock
Cat Hope
Tim Kreger
Charles Martin
Peter McIlwain
Timothy Opie
Mark Pedersen
Lindsay Vickery

Music Review:
Roger Alsop
Warren Burt
Susan Frykberg
Timothy Opie

Conference Organisers:
Timothy Opie
Roger Alsop

Concert/Technical support:
Mark Pedersen
Tim Johnson

With special thanks to:
The Faculty of the Victorian College of the Arts (VCA)
The Melbourne Conservatorium of Music (MCM)
Box Hill Institute (BHI)
The Australian Forum for Acoustic Ecology (AFAE)

Published by
The Australasian Computer Music Association

http://acma.asn.au

July 2014

ISSN 1448-7780

Proceedings edited by Timothy Opie

All correspondence with authors should be sent directly to the authors.

General correspondence for ACMA should be sent to info@acma.asn.au

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 Australian Government requirements for fully-refereed research papers.

Contents

Refereed Papers:

Artist Talks (non-refereed):

Posters (non-refereed):

Programs and Concert Notes

Refereed Papers

Including:

Full Papers

Piece and Paper

SINUSOIDALITY ANALYSIS AND NOISE SYNTHESIS IN PHASE VOCODER BASED TIME-STRETCHING

Dr. Ted Apel

New Zealand School of Music
Victoria University of Wellington, New Zealand
ted.apel@nzsm.ac.nz

ABSTRACT

A novel extension to the phase vocoder method of representing sound is presented in which the "sinusoidality" of spectral energy is estimated during analysis and employed to add noise to a time-stretched phase vocoder synthesis. Three methods of sinusoidality analysis are presented as well as a sinusoid and noise synthesis method which extends the phase vocoder method. This method allows for the noise characteristics of the original sound to be better maintained during time-stretching.

1. INTRODUCTION

The phase vocoder technique of sound analysis and synthesis is well known for its ability to "time stretch" a sound (Moorer 1978). However, when a sound is time-stretched with this technique, the noise aspects of the sound tend to become pitched. Under extreme time lengthening, all noisy aspects of the original sound are transformed into stable sinusoidal components. This behavior is consistent with the phase vocoder's modeling of short-time Fourier transform (STFT) energy as exclusively sinusoidal energy. The purpose of this research is to extend the phase vocoder (PV) representation of monophonic sounds to allow for the original noise characteristics to be maintained during PV lengthening.

In order to derive the noise characteristics of sound in a PV representation, methods of analyzing the "sinusoidality" of STFT channels is employed. STFT channels with high sinusoidality are composed of predominantly sinusoidal energy and should exhibit sinusoidal characteristics during synthesis. STFT channels with low sinusoidality have predominantly noise energy and should be synthesized with noise characteristics. In this paper, three techniques for measuring the sinusoidality of STFT channels are presented, and a new method of combining pitched and noisy components during PV synthesis using the sinusoidality is presented.

2. SINUSOIDALITY ANALYSIS

The phase vocoder is built upon the Fourier transform. The Fourier transform analyzes a signal for sinusoidal components, that is, the output consists of coefficients to a sinusoidal basis function. The sinusoidal nature of Fourier analysis is ill-suited to the analysis of noisy sounds or sounds with significant noise components because these components are analyzed as many rapidly varying sinusoidal components. While this analysis can be used to reconstruct the original sound, temporal manipulation of this Fourier energy using phase vocoder techniques typically results in the noise characteristics of the sound taking on a pitched or sinusoidal character.

In order to alleviate this problem of noisy components of a sound being misinterpreted in phase vocoder manipulations, we will attempt to analyze the phase vocoder representation for its noise characteristics, and to create these noise levels during a phase vocoder synthesis. This analysis will consist of calculating the sinusoidality of each Fourier analysis channel and, using this measure, determining noise levels for each channel during synthesis.

The sinusoidality of an STFT channel represents the degree to which the energy of each spectral bin consists of sinusoidal based energy. A low sinusoidality measure indicates that the energy in that band is based on random or noise signals. Here we will follow the notation and conventions set by Peeters and Rodet (Peeters and Rodet 1998) where $\Gamma(n,k)$ is the sinusoidality of spectral bin k at time frame n, and $\Gamma(n,k)$ varies between high sinusoidality of 1 and 0 for low sinusoidality or high-noise content.

In the next section, three methods of estimating sinusoidality are presented which are used for estimating noise levels. Two of these methods (Charpentier and Phase Acceleration) are adapted from established algorithms, while the third is a novel method based on the harmonic structure of musical sounds.

2.1. Charpentier Sinusoidality

The phase vocoder method calculates an approximation of the instantaneous frequencies of spectral components from the difference in phase between consecutive spectral frames. These frequency values can be thought of as refinements to the nominal center frequency values of each bin. Charpentier devised a pitch detection algorithm that groups spectral bins based on their similar instantaneous frequencies (Charpentier 1986). Charpentier notes that a sinusoid will exhibit energy in at least three adjacent spectral bins, and that the instantaneous frequency of these bins will be correlated around the true frequency of that sinusoidal component. Figure 1 shows the amplitude spectrum

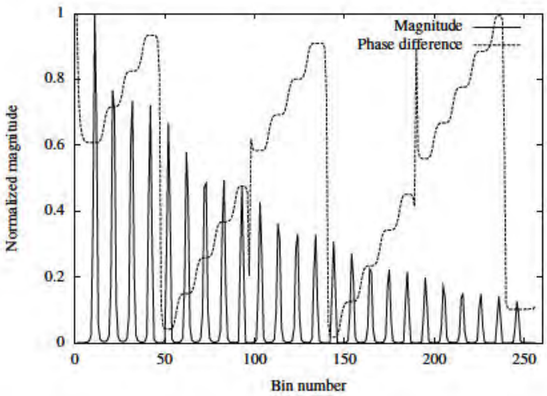

Figure 1. Amplitude spectrum and phase difference spectrum of a harmonic sound. Visible are the similar phase difference values around the sinusoidal energy peaks.

of a synthetically generated harmonic sound and the corresponding phase difference spectrum both scaled between 0 and 1. It can be clearly seen that the phase difference values stabilize around the areas with sinusoidal energy. Dressler (Dressler 2006) formalizes Charpentier's phase difference method as a decimal offset $\kappa(k)$ to the integer bin number k. Here $\kappa(k)$ is calculated from the offset between the measured phase $\theta_n(k)$ and the expected phase calculated from the prior phase of that bin, here called $\theta_{(n-1)}(k)$.

$$\kappa(k) = \left(\frac{N}{2\pi(SR)}\right)(\theta_n(k) - \theta_{(n-1)}(k)), \quad (1)$$

Where N is the hop size and SR is the sampling-rate. Charpentier uses each value of $\kappa(k)$ to detect harmonics of a signal by declaring a bin k to contain a harmonic when $\kappa(k-1)$ and $\kappa(k+1)$ are "sufficiently close" to $\kappa(k)$. We interpret this idea here to mean that for each spectral bin k,

$$SC(k) = |\kappa(k) - \kappa(k-1)| + |\kappa(k) - \kappa(k+1)|. \quad (2)$$

Here we can see that three points that cluster around a particular instantaneous frequency value will produce a low coefficient for that bin, and conversely, large variations in instantaneous frequency between spectral bins will produce a larger coefficient.

We can put this measure into our sinusoidality framework by normalizing the results and subtracting from one.

$$\Gamma_c(k) = 1 - \frac{SC(k)}{\underset{k}{\operatorname{argmax}}(SC(k))}. \quad (3)$$

As discussed below, the Charpentier sinusoidality method devised here exhibits the lowest error on one of our test signals with a single sine tone and broadband noise.

2.2. Phase Acceleration Sinusoidality

Our second phase spectrum based sinusoidality measure also employs the phase difference method of instantaneous frequency computation. Settel and Lippe devised a "band-limited frequency dependent noise gate" as a method of separating stable STFT channels from non-stable channels (Settel and Lippe 1994). Their method defines a threshold for changes in the phase vocoder's instantaneous frequency values between spectral frames. Because sinusoidal components should have a relatively constant instantaneous frequency, these slowly changing components are typically sinusoidal and pitched. A mathematical presentation of the Settel and Lippe method is presented by Arfib, Keiler, and Zölzer (Arfib, Keiler, and Zölzer 2002). Here we formulate this idea in terms of a variable sinusoidality instead of a threshold for each bin, and in terms of "phase acceleration" because the difference between instantaneous frequency values can be thought of as the acceleration of the phases of that frame.

The instantaneous frequency is calculated for each channel from the phase difference between frames n and $(n-1)$ by,

$$\Delta\theta_n(k) = \theta_n(k) - \theta_{(n-1)}(k) \quad (4)$$

where $\theta(k)$ are the principle values of the phase and are bounded by π and $-\pi$. The phase acceleration, $\Delta\Delta\theta_n(k)$, is the difference between the instantaneous frequency and the prior instantaneous frequency,

$$\Delta\Delta\theta_n(k) = \Delta\theta_n(k) - \Delta\theta_{(n-1)}(k). \quad (5)$$

The above two equations are combined to show the phase acceleration in terms of phase:

$$\Delta\Delta\theta_n(k) = \left(\theta_n(k) - \theta_{(n-1)}(k)\right) - \left(\theta_{(n-1)}(k) - \theta_{(n-2)}(k)\right), \quad (6)$$

so that,

$$\Delta\Delta\theta_n(k) = \theta_n(k) - 2\theta_{(n-1)}(k) + \theta_{(n-2)}(k). \quad (7)$$

We can normalize this phase acceleration to our nominal sinusoidality range of zero to one to create our final sinusoidality coefficients:

$$\Gamma_{pa}(n,k) = \left(1 - \left(\frac{\Delta\Delta\theta_n(k)}{2\pi}\right)\right)^p, \quad (8)$$

where $\Gamma_{pa}(n,k)$ is the resultant sinusoidality spectrum for frame n. This spectrum shows values near 1 for stable channels and values tending toward 0 for unstable channels. This sinusoidality measure tends to produce sinusoidality coefficients near one for both stable and non-stable components. The variable p in the above equation is used to counteract this tendency. By raising each coefficient to a small integer power of itself, typically $p = 4$, this overall tendency is significantly abated.

Duxbury et al. proposed a frequency dependent threshold to improve the Settel and Lippe method (Duxbury et al. 2001). They note that for any given value of Γ_{pa}, low frequency channels tend to be selected as stable and high frequency channels selected as non-stable. They alleviate this tendency by choosing a different threshold value for each octave sub-band created with six constant-Q filters.

Here we propose a new method of reducing frequency dependency in terms of our sinusoidality spectrum $\Gamma_{pa}(k)$. In order to reduce the sinusoidality measure $\Gamma_{pa}(k)$ for high k, we will change the slope of the sinusoidality spectrum by multiplying each $\Gamma_{pa}(k)$ by a scaled version of the channel number k. Our new stability spectrum is:

$$\Gamma_{pa}(k) = \left(1 - \left(\frac{\Delta\Delta\theta_n(k)}{2\pi}\right)\right)^p (M)\left(\frac{k}{k_{max}}\right), \quad (9)$$

where M is a slope constant with a heuristically determined value of approximately -1.024. This correction to the stability spectrum avoids the constant-Q filter calculation and multiple threshold calculations of the Duxbury et al. method.

2.3. Harmonic Sum Spectrum Sinusoidality

The measure of the overall "harmonicity" of a STFT frame is a frequently calculated feature, that typically requires the explicit calculation of a single fundamental frequency (f_0) by peak picking or other parametric techniques (Sarlo 2004). As our project here is based on the traditional PV parameters of amplitude, phase and/or instantaneous frequency, those methods that require higher order analysis such as spectral peak picking or fundamental frequency analysis are precluded. With these restrictions, we have created a new non-parametric sinusoidality analysis technique for STFT frames that relies on the harmonic relationships present in the sound, but without requiring explicit computation of these frequencies.

Musical instrument sounds typically have a predominantly harmonic structure. The spectral components of a musical tone that are in a harmonic relationship are the predominantly sinusoidally based energy of a spectral frame. Conversely, we can consider spectral energy that is not in a harmonic relationship to be noise based. This is the basis of the sinusoidality coefficient measure presented here.

The harmonic product spectrum (HPS) technique is a method of finding the fundamental frequency energy of a STFT by combining the energy of the harmonics with each other at the location of the fundamental frequency (Schroeder 1968). The harmonic product spectrum is the product of the power spectrum multiplied by successively compressed versions of the same spectrum. These spectra are down-sampled along the frequency axis by consecutive integer amounts. The resultant product reinforces the amplitude of the fundamental frequency. A related spectrum, the harmonic sum spectrum (HSS), with similar output, sums these down-sampled spectra instead of multiplying them (Alonso, David, and Richard 2003). The harmonic sum spectrum is given by:

$$HSS(k) = \sum_{r=1}^{K} |X(k_r)|^2 \quad \text{where} \quad k_r = (r)(k). \quad (10)$$

Here K is the number of down-sampled spectra typically around six.

Conceptually, our proposed method is the reverse of the HSS. To the normalized power spectra we add each integer multiple up-sampled copy of the spectra. These up-sampled spectra are calculated by interpolating new spectra at integer multiples of the original spectra. That is:

$$\Gamma_h(k) = \sum_{r=1}^{K} |X(k_r)|^r \quad \text{where} \quad k_r = k/r. \quad (11)$$

Spectral peaks in each of these up-sampled spectra are proportionately wider than those in the original power spectra. For example, the spectral peaks in the $r = 2$ spectra are twice as wide as the original. In order to slim these peaks back to an approximation of their original width, the up-sampled spectra are raised to increasing powers and normalized to their highest peak. In order to heuristically calculate this power term, a Hann window is interpolated to twice its length and raised to integer powers. These are normalized to the original height of the Hanning window. It was found that the window raised to the 4th power was slightly wider than the original window and was slightly narrower when raised to the 5th power. So, the $r = 2$ spectra would be taken to the 4th power in order to reduce the width of its peaks to approximate the original width. However, this 4th power was found to produce banks that are too narrow to use in practice. The first up-sampled spectrum is set to the 2nd power in our algorithm, and each subsequent up-sampled length is raised to the next power as can be seen in equation 11. This produces approximately equal width peaks for six to twelve up-sampled spectra K. Finally, the slope correction devised for the phase acceleration sinusoidality method is used to reduce the sinusoidality of higher harmonics. Figure 2 shows a normalized power spectrum for a violin tone and the corresponding harmonic sum sinusoidality for this spectra. The harmonics higher than those of the power spectra are visible in the sinusoidality spectrum.

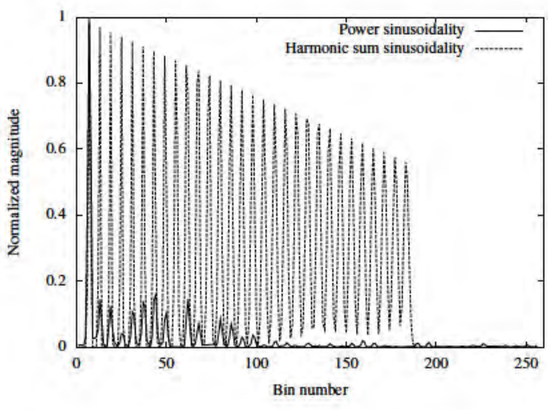

Figure 2. Power spectrum sinusoidality and harmonic sum sinusoidality for a a single frame of a violin tone.

2.4. Other Sinusoidality Analysis Methods

In addition to these three methods, many other sinusoidality analysis methods have been developed and compared in a prior study (Apel 2008). In addition, several other methods have been proposed and not adapted to the PV time-stretching framework discussed here (Zivanovic et al. 2007; Dubnov 1999).

The three sinusoidality measures presented here have been shown to be effective in separating sinusoidal from noise energy for differing synthetic test signals (Apel 2008). The Charpentier method performs well with a single sine tone in noise. The phase acceleration measures performs well for inharmonic tones in noise. The harmonic sum measure performs better than the others for a harmonic tone. They are each suited to analyzing different types of sounds. Even when the sounds are restricted to monophonic tones, no one sinusoidality measure presents itself as ideal. In the following section, methods of adding noise to a PV representation will be presented that rely on the sinusoidality coefficients for each PV frame in order to determine the amount of noise to be added to each channel. We employ the three different measures depending on the type of sound being analyzed.

3. SINUSOID AND NOISE PHASE VOCODER SYNTHESIS

A method of synthesizing time-stretched sounds from phase vocoder analyses and sinusoidality coefficients for each spectral frame is presented in this section. In this method, the percentage of energy of each spectral bin that is noise-based is multiplied by a spectral domain noise signal. Before describing our sinusoid and noise synthesis algorithm, two existing alternative methods will be considered.

A method of segregating spectral energy separates the spectral channels into two groups, one the "sinusoidal" channels, and the other the "noisal" channels. For each frame, the individual bins are determined as belonging to one of two groups. This is the scheme proposed by Lippe and Settle for segregating bins (Settel and Lippe 1994; Magnus 2001).

A threshold is set for each bin over which the bin is labeled as sinusoidal. Conversely, if the sinusoidal measure is below the threshold, the channel is labeled as noise based energy. Thus,

$$X_s(k) = (\Gamma_n(k) \geq T), \tag{12}$$

where $\Gamma_n(k)$ is a frame of sinusoidality measures between 0 and 1 for each channel, T is a stability threshold between 0 and 1, and $X_s(k)$ is a phase vocoder spectral frame with only the stable channels above the threshold not set to zero. Noise is then added to the spectra by a method shown in section 3.1. This method is less suited than the method proposed below because of the sharp cutoffs between adjacent bins created from the thresholding of energy.

A second method of adding noise to a PV spectrum involves noise modulation of each spectral bin to a degree determined by the sinusoidality of each bin. This method provides a single or unified spectral representation, as opposed to the dual representations in which the sinusoidal spectra and noise spectra are separated for processing and recombined during synthesis. While the technique is well suited for creating a unified representation of noise in sinusoidal modeling synthesis, it is ill sited for PV based systems in which sinusoidal energy is necessarily spread across several bins of the spectrum [1]. For example, a single sinusoid in a PV spectra will occupy at least three consecutive bins. Increasing the noise "bandwidth" of each of these bins would put energy in adjoining channels. A system of partitioning the STFT spectral bins, as proposed by Dolson and Laroach (Laroche and Dolson 1999) could perhaps be of use here, but we preclude it here based on its increased parametric nature.

A system based on this second method has been developed by Liao, Roebel, and Su in order to time-stretch gaussian noise as a test case in the creation of a phase vocoder based sound texture time-stretching method (Liao, Roebel, and Su 2012). Employing a temporal correlation function to maintain the statistical properties of the STFT phase spectrum, their method successfully time-stretches purely noise signals suggesting that it may become suitable for time-stretching complex sounds with noise and pitched components.

3.1. Dual Model Synthesis

Our method of segregating sinusoidal energy from noise based energy in a spectral representation divides the energy of each spectral bin into two parts corresponding to the sinusoidal energy and the noise-based energy. These separated spectra are processed separately and recombined during synthesis. Figure 3 shows our complete analysis synthesis system for time-stretching noise and pitched sounds. This process is discussed below.

The process starts with the STFT analysis data as a series of amplitude and phase spectra frames along with corresponding sinusoidality coefficients for each spectral bin. For each frame n, a new amplitude spectrum is created by using a sinusoidality coefficient spectra $\Gamma_n(k)$ to scale the amount of sinusoidal energy present in each bin. For each bin k of each spectral frame n,

$$|S_n(k)| = (\Gamma_n(k))\,(|X_n(k)|), \tag{13}$$

where $|S_n(k)|$ is a new magnitude spectrum, here called "sinusoidal magnitude spectrum", in which each bin's amplitude is scaled by the corresponding sinusoidality of that bin.

Next, the corresponding "noise magnitude spectrum," $|N_n(k)|$, is calculated by subtracting the sinusoidal magnitude spectrum from the unaltered magnitude spectrum,

$$|N_n(k)| = |X_n(k)| - |S_n(k)|. \tag{14}$$

As can be seen here, the original magnitude spectrum can be recreated by adding the noise magnitude spectrum and the sinusoidal magnitude spectrum.

[1] This distinction is analogous to the distinction between how noise and sinusoids are modeled in a dual manner in SMS modeling (Serra and Smith 1990) and how they are represented in a unified representation in the Fitz "bandwidth enhanced" method (Fitz and Haken 1995). In Fitz's system the individual sinusoids of a sinusoidal model of sound are modulated with noise to increase their bandwidth. This process increases the noise level of each sinusoidal component.

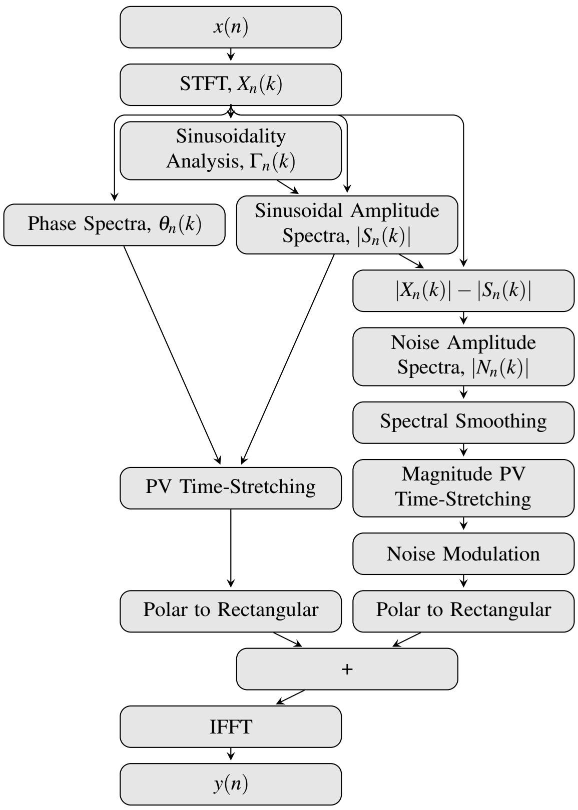

Figure 3. Sinusoidality analysis and noise synthesis system.

The noise magnitude spectrum is subject to a large variance in bin amplitude values between spectral frames. While this behavior accurately reflects the character of noise energy, it produces unwanted sonic artifacts when these random fluctuations are subject to PV time-stretching. For this reason the noise magnitude spectra are filtered in order to smooth each channel in time and frequency. The time frame smoothing is achieved by recursively averaging past frames of $|N_n(k)|$. Martin suggests a spectral noise smoothing filter for use on ambient stationary noise signals (Martin 1994),

$$|NS_n(k)| = \alpha|NS_{(n-1)}(k)| + (1-\alpha)|N_n(k)|, \tag{15}$$

where $|NS_n(k)|$ is the smoothed noise magnitude spectrum, and the smoothing constant α is typically set, according to Martin, between 0.9 and 0.95. As the dynamic character of musical sounds is typically present in the noisy aspects of sound, we use a smoothing constant of 0.4, which is much smaller than the Martin suggestion.

In addition to this temporal smoothing, we smooth the noise magnitude spectrum across bins. This is simply achieved by using a running average of β bins, where $8 < \beta > 12$,

$$|NS_n(k)| = \frac{\sum_{l=-\beta/2}^{\beta/2} |N_n(k+l)|}{\beta}. \tag{16}$$

Both of these techniques are used to smooth our noise magnitude spectra.

Our smoothed noise magnitude spectra $|NS_n(k)|$ are time-stretched separately from the sinusoidal magnitude spectra above. Since they are only amplitude spectra without corresponding phase or instantaneous frequency spectra, the amplitude spectra are simply interpolated between spectral frames. These new noise magnitude spectra are then each multiplied by a different STFT analysis of white noise,

$$N_n(k) = |NS_n(k)|P_n(k), \tag{17}$$

where $P_n(k)$ is a new STFT of white noise for each n of the time stretched sound. Each new noise spectral frame is added to its corresponding sinusoidal spectral frame,

$$Y_n(k) = S_n(k) + N_n(k), \tag{18}$$

where $S_n(k)$ is the complex sinusoidal spectral frame produced by a traditional phase vocoder time stretch of the sinusoidal magnitude spectrum $|S_n(k)|$ with the original phase and derived instantaneous frequency values. $Y_n(k)$ is the resultant spectral frame that is inverted to the time domain by the IFFT and appropriate windowing. The final new sound $y(n)$ is shown at the bottom of figure 3.

Several differing sounds both musical and environmental were time-stretched with the new sinusoidality analysis and noise synthesis method. In each case, the original unaltered sound is followed by a traditional phase vocoder time-stretching with 8 times the normal length. Then, the new time stretched sound with the noise characteristics preserved is listed. Sample sounds can be found on the website:

`http://vud.org/fppv/`

4. CONCLUSION

In each case, it can be heard that the generation of noise as part of the phase vocoder time-stretching creates a sound that more closely resembles the noise characteristics of the unaltered sound as compared to the traditional phase vocoder time-stretching method. On occasion, the two parts of the time-stretched sound, sinusoidal and noise based, do not fuse perceptually as they do in the original sound. This is no doubt attributable to the dual nature of our synthesis system. Future work could consist of devising methods of combining the two modes of synthesis into a unified representation. Improvements to the sinusoidality algorithms could consist of developing a single optimized measure for most sound types or automating the selection of sinusoidality measures based on analyzing the source sound.

References

Alonso, M., B. David, and G. Richard. 2003. "A Study of Tempo Tracking Algorithms from Polyphonic Music Signals." *4th COST 276 Workshop* .

Apel, T. 2008. "Feature Preservation and Negated Music in a Phase Vocoder Sound Representation." Ph.D. thesis, University of California, San Diego.

Arfib, D., F. Keiler, and U. Zölzer, (editors) . 2002. *DAFX - Digital Audio Effects*. John Wiley and Sons, LTD.

Charpentier, F. 1986. "Pitch detection using the short-term phase spectrum." *Acoustics, Speech, and Signal Processing, IEEE International Conference on ICASSP* 11:113– 116.

Dressler, K. 2006. "Sinusoidal extraction using an efficient implementation of a multi-resolution FFT." *Proceedings of 9th International Conference on Digital Audio Effects (DAFx-06)* :247–252.

Dubnov, S. 1999. "HOS Method for Phase Characterization in Sinusoidal Models with Applications for Speech and Audio." *IEEE Signal Processing Workshop on Higher-Order Statistics* :1–5.

Duxbury, C., M. Davies, and M. Sandler. 2001. "Separation of Transient Information in Musical Audio using Multiresolution Analysis Techniques." *Proceedings of the COST G-6 Conference on Digital Audio Effects (DAFX-01)* .

Fitz, K., and L. Haken. 1995. "Bandwidth Enhanced Sinusoidal Modeling in Lemur." *Proceedings International Computer Music Conference* .

Laroche, J., and M. Dolson. 1999. "New phase-vocoder techniques for real-time pitch shifting, chorusing, harmonizing, and other exotic audio modifications." *Journal of the Audio Engineering Society* 47(11):928–936.

Liao, W.-H., A. Roebel, and A. W. Su. 2012. "On Stretching Gaussian Noises with the Phase Vocoder." *Proceedings of the 15th Conference on Digital Audio Effects (DAFx-12)* .

Magnus, C. 2001. "Real-Time Separation of Periodic and Non-periodic Signal Components." Unpublished paper.

Martin, R. 1994. "Spectral Subtraction Based on Minimum Statistics." *Proceedings of EUSIPCO-94 Seventh European Signal Processing Conference* .

Moorer, J. A. 1978. "Use of Phase Vocoder in Computer Music Applications." *Journal of the Audio Engineering Society* 26(1-2):42–45.

Peeters, G., and X. Rodet. 1998. "Signal Characterization in terms of Sinusoidal and Non-Sinusoidal Components." *Proceedings DAFX98* .

Sarlo, J. 2004. "Real-time Pitched/Unpitched Separation of Monophonic Timbre Components." *Proceedings International Computer Music Conference* .

Schroeder, M. 1968. "Period Histogram and Product Spectrum: New Methods for Fundamental-Frequency Measurement." *The Journal of the Acoustical Society of America* 43(4):829–834.

Serra, X., and J. Smith. 1990. "Spectral Modeling Synthesis: A Sound Analysis/Synthesis System Based on a Deterministic Plus Stochastic Decomposition." *Computer Music Journal* 14:12–24.

Settel, J., and C. Lippe. 1994. "Real-time Musical Applications using the FFT-based Resynthesis." *Proceedings International Computer Music Conference* .

Zivanovic, M., A. Roebel, and X. Rodet. 2007. "Adaptive Threshold Determination for Spectral Peak Classification." *Proceedings of Ninth International Conference on Digital Audio Effects* .

HEIGHT PERCEPTION: LOCATION, MOVEMENT AND IMMERSION IN THE VERTICAL HEMISPHERE

James L. Barbour

Independent Spatial Audio researcher
PhD candidate, Spatial Information Architecture Laboratory,
SIAL Sound Studios, School of Architecture and Design,
RMIT University, Melbourne, Australia
jimbarbour54@gmail.com

ABSTRACT

Recent interest in auditory perception incorporating height in acoustic space has developed rapidly over the last few years along with techniques for the production of sound in the vertical hemisphere. While several cinema formats now incorporate overhead loudspeakers, sound artists have been using them for decades, usually in site specific configurations. This paper investigates sound source localization, source movement, spatial impression, immersion and envelopment in 3D soundfields using an array with horizontal and elevated loudspeakers and suggests guidelines for generating perceptually useful 3D environments.

1. INTRODUCTION

Our understanding of auditory perception of sound sources in 3D acoustic space has matured with the psychoacoustic principles involved well defined. On the horizontal plane, inter-aural differences in level (ILD) and time (ITD) provide solid cues for perceptual localization of sound sources. For sources above and below the horizontal plane, spectral cues generated by reflections within the pinna are the strongest cues, particularly where there are no inter-aural differences. Williams [1] identified the complex nature of pinna related ILD and the impact they have on the perception of sound sources at different angles of elevation, see figure 1.

However, compared to inter-aural differences, these elevation spectral cues are weak and vary considerably between listeners, making them impossible to predict and impractical to implement in production. However they do partially account for high frequency sounds being perceived at higher elevations than low frequency sounds when both are reproduced at ear height.

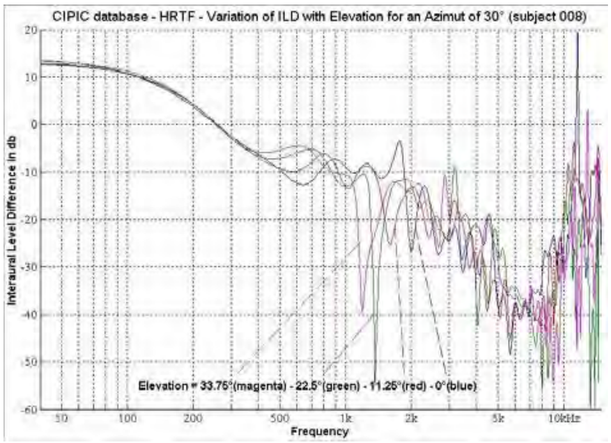

Figure 1: ILD spectral variations for one subject at different elevation angles (after Williams [1]) [2]

Attempts to harness these cues to create virtual height loudspeakers have had very limited success, and in the opinion of the author, never come close to creating the impact of overhead sound sources and immersive environments possible with speakers mounted above the listening position. Consequently, the application of these psychoacoustic principles in loudspeaker arrays incorporating overhead loudspeaker positions requires more investigation.

Another consideration in testing for 3D localization is whether or not to allow participants to move their head during testing. In research into elevation and azimuth localization reported by Ashby et al [3], accuracy was measured for 3 different cases of self-managed head movement: no movement, forced movement with the head rotating +/-30º azimuth continually, or free movement in any direction. Their results stated that 'For noise programme items, head movements significantly increase elevation localization accuracy' and '…increase azimuth localization accuracy.' They also note 'that pinna cues are less effective than head movement at resolving front-back confusion.'

Head movement (azimuth and elevation) transforms our pinna into a dynamic filter bank of multiple band pass and band reject filters, each with variable centre frequency, variable Q and variable amplitude, thereby endowing an elevated stationary source with perceptual motion and reduced localization blur. As a normal listening environment allows free movement, this position was adopted by the author for this study.

Following on from a previous study by the author of phantom images in the vertical hemisphere [4], this study uses a 3D loudspeaker array to investigate auditory perception of sound source location and movement around and over a listener. The study also investigates spatial impression, immersion and envelopment in 3D reverberant sound fields which may fully envelop the listener, or which appear to be localized to segments of acoustic space.

2. METHODOLOGY

A 3D loudspeaker array is being used for perceptual testing. The Equidome [5] is a 12.2 loudspeaker array with a ring of six speakers on the horizontal plane, six speakers in an elevated ring positioned 60 degrees up from the lower ring and two sub woofers on the floor either side of the listener. One pair of speakers in each ring is on the frontal plane, (+/- 90º) and all others are at 60º angles from that pair, ie. +/-30º and +/-150º. All speakers used are Bose cubes which have been equalized to be consistent in timbre and amplitude, and are spaced on a 1.5 metre radius from the listening position in a small room with substantial wall and ceiling treatment to minimize acoustic anomalies, see figure 2. All speakers were visible to the participants which is consistent with a normal home listening environment.

Figure 2: The Equidome

Nuendo software allows placement and automated movement of multiple, simultaneous sound sources anywhere throughout the array and Waves 360 convolution reverb allows the creation of multiple early and late reflections, each of which may be placed individually. Perceptual testing using a small group of expert listeners and controlled double-blind procedures has been completed and a quantitative and qualitative analysis has been undertaken.

3. RESULTS

3.1. Location

The first stage of testing focussed on static sound source locations around and over the listeners using male speech and a vibraslap percussion sample. Results indicate that differences in the accuracy of the perceived location, spectral changes and localization blur depend on whether the source is located totally within one speaker, hereafter called in-speaker, as a phantom image created by amplitude panning between two speakers, or as an image created through amplitude panning between four speakers.

When the sound sources are located in-speaker, it was clear that all participants perceived the sounds at that speaker position with no localization blur and with a spectral character which became the benchmark for subsequent comparison. For locations on the horizontal plane, limited head movement was observed for participants to be confident in selecting a location. When the in-speaker location was on the elevated plane, participants moved their heads using small turning and tilting motions, and successfully located the sources at the front, side or rear speaker positions.

Using amplitude panning to locate sources halfway between two speaker positions on the horizontal plane with the usual 3dB panning law, the results were very accurate for front centre and rear centre, with slightly reduced accuracy and a small localization blur for side locations, for example left front to left side or left side to left rear. This is consistent with knowledge of phantom images for a 60º stereo speaker pair where inter-aural time and level differences work their magic but become less effective at lateral locations. It is noted that speaker locations on the frontal plane at +/- 90º are very important for localization around the listener when combined with a rear pair at +/-150º and overcome poor horizontal localization when using the standard 5.1 arrangement of +/-110º.

When the sources are amplitude panned halfway between a horizontal and elevated speaker pair, for example horizontal front left to elevated front left (60º elevated), azimuth angle estimation was very good but elevation angle results were inconsistent with a tendency to localize at the elevated speaker rather than at a mid-point. Head movement was used by all participants and observed localization blur was greater, while some participants also perceived a spectral change.

When sources were panned between elevated only pairs, elevation perception was consistently accurate but azimuth angle estimation was less accurate with some front-rear confusion observed and localization blur greater than for the other two speaker phantom locations. The most accurate localization was for elevated left side and elevated right side speakers which are on the frontal plane, where inter-aural differences combine more effectively with elevation spectral changes.

The final test results were obtained when the phantom image was generated using amplitude panning between four speakers, for example between horizontal front left and right and elevated front left and right, including directly overhead where there is no speaker. In these cases, the phantom location was perceived with good azimuth angle estimation with a small localization blur increasing away from front centre and rear centre. Elevation angle perception again tended to be consistently located at a higher angle and was always above the horizontal plane. Head movement assisted in locating the sources and the top centre position was accurately located by all participants. Spectral changes were again noted by several participants.

Location Discussion

Previous studies of average angular localization errors, notably in Best at al [6] and Hammershoi [7], were based on studies conducted with a moving speaker or an array with a very large number of speakers. This removed any difficulties posed by phantom image generation but there were inaccuracies caused by methodologies used for indicating the location of a perceived sound source, with limitations due to head movement tracking, nose mounted pointing devices and other techniques. It was noted that the angular error increased as sources move away from the front, and up and over the top.

From the study presented here, using an array with a limited number of speakers, in-speaker locations were accurately located in both the horizontal and elevated sectors of the vertical hemisphere. When the sound source location was generated using amplitude panning between 2 speakers on the horizontal plane, the phantom image was perceived with good azimuth accuracy and little blur. A phantom image generated on the elevated plane was perceived with good elevation angle estimation but poor azimuth angle estimation, while on the vertical planes between 2 speakers, the accuracy was good for azimuth but poorer for elevation, though the sound sources were always perceived to be elevated. Four speaker phantom images generally had good azimuth perception but greater blur for elevation, though again they were always perceived to be elevated. In particular, for elevated sources on the frontal plane including directly overhead, the coherence of elevation spectral cues and inter-aural differences results in good accuracy with reduced localization blur.

	Azimuth accuracy	**Elevation accuracy**	**Localization blur**
in-speaker location	excellent	excellent	none
phantom image: **2 speaker horizontal only**	very good	n/a	low
phantom image: **2 speaker elevated only**	some front/rear confusion	very good	medium
phantom image: **2 speaker vertical plane**	very good	ok, tends to elevated positions	medium
phantom image: **4 speaker horizontal and elevated**	good	ok, tends to elevated positions	medium, worse at rear elevated postions

Table 1 3D localization

3.2. Movement

For the second test, a stereo sample of broadband noise was created using inter-channel amplitude panning using the auto-pan function within Adobe Audition with the 3dB panning law. The noise source panned from left to right and back to the left over 4 seconds and for testing, the left channel is allocated to one speaker and the right to a second speaker. Trajectories have been mapped around and over the listeners with an angular range of 60º, and have included horizontal only paths, elevated only paths, vertical paths from a horizontal speaker to an elevated speaker, across the top on the frontal plane and the median plane, and diagonal paths including horizontal front left to elevated front right.

Results indicate a clear sense of movement along the path in all sectors of 3D acoustic space, but with the start and finish points subject to similar errors to the static locations. When the path crosses the median plane, for example from horizontal front left to horizontal front right, there was a clear sense for all participants of a point source moving between the 2 speakers, consistent with our understanding of stereo phantom images and inter-aural differences. This result was also achieved for all participants for horizontal rear left to right which is to be expected. It was also interesting to note that where the path crossed the median plane at elevated positions, for example from elevated left side to elevated right side, there was 100% accuracy in identifying the path of movement across the top, reflecting the coherence of inter-aural cues and elevation spectral cues.

When the paths were along the horizontal plane only between front and side or side and rear locations, with the start and finish point in-speaker, perception of the path was again very accurate for all participants. However, when the path was along the elevated plane only, there was confusion between front and rear in-speaker locations, which were partially resolved by head turning, but indicated a lack of confidence in locations where spectral cues are weak.

Paths which started at horizontal speakers and moved vertically to elevated speakers were consistently identified with accurate azimuth and a clear sense of movement up and down. Of particular interest with this group of paths was the observation by several participants that the noise source no longer appeared to be a point source moving along a path but appeared now to be a noise source fading in and out at each speaker, with significant blur at the mid-point of the transition. This observation is consistent with no inter-aural change between speaker locations and only spectral cues providing localization information.

The final group of paths tested were diagonal trajectories, for example between horizontal front left and elevated front right. Where these paths crossed the median plane, there was 100% accuracy in identifying the path both for azimuth and elevation, indicating solid inter-aural cues and spectral cues. When the path was around the listener, for example horizontal front left to elevated side left, there was again very good accuracy in both azimuth and elevation perception. However, when the path was wholly at the rear but did not cross the median plane, and also when the path was along the median plane, there was some confusion with the exact starting and finishing locations, consistent with previous location results. There was, nevertheless, a clear perception of movement.

Movement Discussion

The perception of movement for a broadband noise source is clear throughout the vertical hemisphere when the source is panned between 2 in-speaker locations. When the sound source crosses the median plane along the path, there is very accurate perception of the azimuth angle. The elevation angle is also very clear everywhere except rear positions where there is increased localization blur. As identified by Williams [1], diagonal paths between horizontal and elevated speakers are also clear for most paths, though again there is increased blur for start and end locations at the rear. Vertical paths at all azimuth angles tend to exhibit a transition which is not as smooth between the two end points, as elevations cues are not as accurate as inter-aural cues, and the sound source tends to become two isolated sources fading in and out at each point.

It should be clearly understood that this accuracy in localization and movement directly overhead is only possible with speakers elevated 60º from the horizontal plane. If they are below this level, the gap across the top is too large for meaningful phantom locations or movement, though it may be suitable for enhanced 3D reverberant soundfields.

	Azimuth angle perception	Elevation angle perception	Localization blur
horizontal only path across **median plane**	excellent	n/a	none
elevated only path across **median plane**	excellent	very good	low
diagonal path across **median plane**	excellent	very good	low
horizontal only path along **sides**	good	n/a	low
elevated only path along **sides**	ok, some front/rear confusion	good	medium
diagonal path at **sides**	good	good	medium
median plane path	ok	ok	medium, worse at rear
vertical path	very good	ok, tends to isolated images top and bottom	medium, worse at middle point

Table 2 3D movement perception

3.3. Spatial Impression, Immersion and Envelopment

Early and late reflection patterns are generated in virtually all acoustic environments and we perceive them as integral parts of the listening experience. There are two broad categories of reverberant soundfields: those generated by static sound sources in a reverberant space, for example a performance on stage in a concert hall, and urban or country environments where a reverberant soundfield is generated in a smaller localizable sector in acoustic space, and where multiple, different reverberant soundfields may exist simultaneously. There are interesting questions raised about both categories when we consider creating or reproducing similar soundfields, which include the significance of height in the perception of acoustic space.

As Theile and Wittek [8] suggested, '…the perception of distances and depth mostly depends on early reflections,…(and) perception is particularly stable when the reflections come in from the original directions of the upper half space.' Control of early and late reflections and the directions from which they arrive at the listener is of critical importance in generating realistic immersive environments within a 3D speaker array.

The final stage of testing has investigated perception of the reverberant space generated with different balances between horizontal and elevated reflection patterns for a fully immersive and enveloping static environment, and for a reverberant space created around a static source in a small localizable sector in acoustic space. There has also been an investigation into the balance of early reflections between the horizontal plane and the elevated plane for the generation of clear spatial impression, and the balance of late reflections between horizontal and elevated for immersion and envelopment.

The sound sample used for this sequence of tests was a 4 second acoustic guitar phrase recorded in anechoic conditions which was convolved in the Waves 360 reverb software to generate a series of 4 channel early reflections (ER) and separate 4 channel late reflections (LR), a technique inspired by the Spacebuilder concept [9]. Each of the ER and LR components, along with the mono original sound were individually placed around the listeners. For example, the guitar was panned halfway between horizontal front left and horizontal side left speakers, and the early and late reflections were panned in-speaker at horizontal front left and side left and elevated front left and side left. This created an acoustic space located in only one sector of the vertical hemisphere, and with appropriate changes to the locations, any sector could be activated with the direct sound source and reflections. Using software automation, the direct, ER and LR components can be moved around the listener creating a sense of movement of a reverberant soundfield in acoustic space.

Results indicate that all participants found the acoustic space generated was very clear and could easily identify in which sector it was located. The participants were also able to clearly perceive the added sense of immersion generated by the elevated speakers. Two special cases were also tested: all components including the direct sound were panned to elevated speakers only, or to horizontal speakers only. When all component sources were elevated, this was clearly identified as totally overhead and separate from the horizontal plane. When all the components were horizontal only, participants felt fully enveloped in the sound.

As a further test of the capacity to perceive individual sectors of reverberation, a 2 minute percussion piece was played to participants. This piece had been generated by recording six separate percussion instruments in near anechoic conditions and generating distinctly different ER and LR 4 channel reverberation patterns for each part. These parts and their corresponding reverb were panned to each of the six sectors around the participants. All participants could clearly perceive a different spatial impression and reverb identity for each part and felt immersed in the overall soundfield.

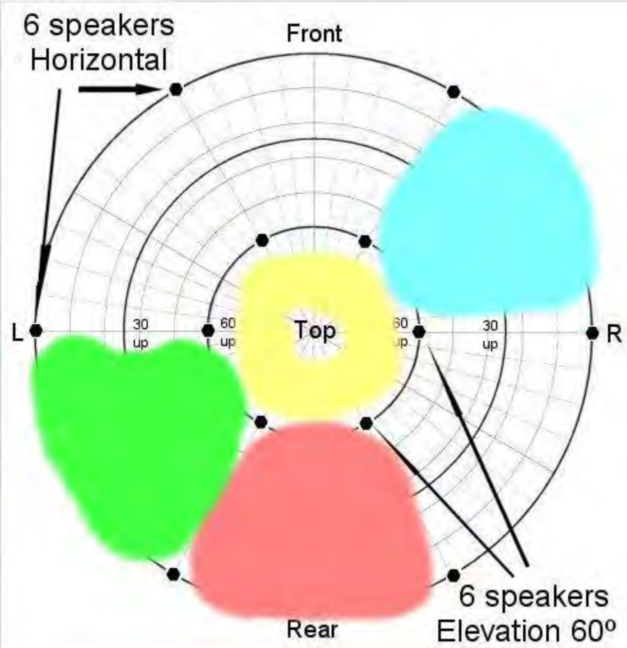

Figure 3: 4 instances of unique spatial impression and reverb identity in 4 separate sectors

These results confirm that it is possible to generate localizable soundfields, each with a different spatial impression and reverb identity incorporating a sense of height, in different sectors around and over a listener and that multiple instances of acoustic space may be perceived simultaneously in separate sectors of the total soundfield around a listener, see figure 3.

4. DISCUSSION AND CONCLUSIONS

This paper has investigated whether it is possible to generate perceptually accurate 3D soundfields with sounds located around and above a listener. An array of six speakers on the horizontal plane and 6 speakers on an elevated plane was used to generate stationary and moving sound sources and to generate reverberation.

When stationary sound sources are panned to in-speaker locations, they can be located with excellent accuracy including behind and above the listener. However, when the sound source is created using phantom images, there is variation in localization accuracy. For phantom images on the horizontal plane, accuracy is very good, but for phantom images on the elevated plane, elevation perception is good but azimuth perception suffers some front/rear confusion which head turning is unable to fully resolve. When phantom images are generated on vertical planes between speakers, azimuth localization is good but elevation perception tends to the elevated speaker position.

Movement perception is very good in all areas of the array but accurate localization of start and finish positions depends on the trajectory. When the path crossed the median plane and inter-aural differences are produced, there is very good accuracy in locating the path, including across the front and back, over the top and diagonally up or down. However, similarly to stationary source perception, it is difficult to create accurate localization for sounds moving vertically between speaker layers even though a sense of movement is still strong.

Using convolution reverb software to generate multiple early and late reflections, which are then placed around and over the listener, allows the creation of 3D soundfields which can accurately emulate the perception of a concert hall, fully immersing the listener with strong spatial impression and enveloping reverberation. If, however, the direct sound, early and late reflections are individually panned to only one part of the array, a localized soundfield with distinct spatial impression and reverb identity can be created in a sector of the soundfield. This can be very useful for a sound artist creating unique acoustic spaces around a listener, or it can enhance the mix for a movie.

Chris Scarabosio, sound mixer for the film ***Elysium***, when talking about mixing for the new cinema formats of Dolby Atmos [11] and Auro 3D [12], reported in Audio Technology magazine [10] '…using reverb not to create more ambience but to localize the ambience.' He was referring to one scene where he created an ambient space around the music mix in one sector of 3D acoustic space, while using the full 3D speaker array of both Atmos and Auro 3D for other components of the sound design.

This study has shown that a 3D loudspeaker array with speakers elevated 60º from the horizontal plane can generate perceptually useful localization of stationary and moving sound sources around and above listeners. The array can also produce 3D soundfields with clear spatial impression, immersion and envelopment in a single acoustic space or produce multiple, simultaneous acoustic spaces around and above the listener, each of which may have unique spatial impression and immersive identity. The use of a 3D loudspeaker array offers exciting opportunities to sound designers and sound artists to explore the creative potential of immersive aural environments.

5. REFERENCES

[1] Williams, M., *The Psychoacoustic Testing of the Multiformat Microphone Array Design, and the Basic Isosceles Triangle Structure of the Array and the Loudspeaker Reproduction Configuration*, AES paper 8839

[2] CIPIC database can be downloaded at: http://interface.cipic.ucdavis.edu/sound/hrtf html

[3] Ashby, T, Mason, R and Brookes, T, *Head Movements in Three Dimensional Localisation*, AES paper 8881

[4] Barbour, J, *Elevation Perception: Phantom Images in the Vertical Hemisphere*, AES 24th International Conference, paper no.14

[5] Barbour, J, *The Equidome, a Personal Spatial Reproduction Array*, AES paper 8598

[6] Best, V et al, *A Meta-analysis of Localization Errors Made in the Anechoic Free Field, Principles and Applications of Spatial Hearing*, World Scientific, 2011

[7] Hammershoi, D, *Localization Capacity of Human Listeners, Principles and Applications of Spatial Hearing*, World Scientific, 2011

[8] Theile, G and Wittek, H, *Principles in Surround Recordings with Height,* AES paper 8403

[9] Woszczyk, W et al, *Space Builder: An Impulse Response-Based Tool for Immersive 22.2 Channel Ambiance Design*, AES 40th International Conference: Spatial Audio: Sense the Sound of Space, Paper Number: 7-1

[10] Audio Technology magazine, issue 97, p56, Alchemedia Publishing Pty. Ltd.

[11] Dolby Atmos, http://www.dolby.com/us/en/consumer/technology/movie/dolby-atmos.html

[12] Auro 3D, http://www.auro-3d.com/

[13] Kim, S et al, *Subjective Evaluation of Multichannel Sound with Surround-Height Channels*, AES paper 9003

[14] Pulkki, V, *Virtual sound source positioning using vector base amplitude panning*, JAES, vol 45(6), pp 456-466, June 1997

INTERFACING REAL-TIME AUDIO AND FILE I/O

Ross Bencina
portaudio.com
rossb@audiomulch.com

ABSTRACT

Programming a computer to record or play a sound file in real-time is not as easy as it may seem. The naive approach is to call file I/O APIs from within the routine that handles real-time audio I/O. This leads to audible glitches whenever the time taken to access a file exceeds the time available to deliver a buffer of real-time audio. This paper describes an approach to streaming file playback and recording that operates correctly under these conditions. It performs file I/O in a separate thread, buffers audio data to mask file I/O delays, and uses asynchronous message passing and lock-free queues for inter-thread communication.

1. INTRODUCTION

Real-time audio I/O and file I/O are input/output methods with radically different timing properties. Software that records or plays sound files must perform both types of I/O. To ensure that a constant stream of audio samples is fed to and from the audio hardware, real-time audio programs must adhere to a strict schedule. By contrast, file I/O operates on a best-effort basis, guaranteeing only average-case throughput, not the timeliness of individual operations. These differing timing properties make it difficult to interface between real-time audio and file I/O. Solutions typically involve performing file I/O in a separate thread. This poses further difficulties, as most general-purpose operating systems do not provide real-time-safe inter-thread communication facilities.

This paper presents an efficient and reliable method for implementing real-time file playback and recording. The method comprises streaming algorithms, an asynchronous message passing protocol, a system of lock-free queues, and a "server" thread that performs file I/O. A sample implementation in C++ is also available.[1]

The paper includes the following lock-free mechanisms, which readers may find generally useful for real-time audio programming:

- A light-weight, lock-free message architecture supporting actor-like stream objects. (Aside from the I/O server thread, no message loop, scheduler, or per-thread queues are used.)
- Real-time-safe creation and destruction of stream objects; including non-blocking tear-down when streams are awaiting asynchronous results.
- Light-weight, relaxed-order message queues for returning results to stream objects.

[1] https://github.com/RossBencina/RealTimeFileStreaming

2. CONTEXT

Real-time audio programs must adhere to strict time constraints, yet few non-audio system APIs guarantee timeliness. Here we explore these issues and introduce concepts and terms for later use.

2.1. Real-time audio terminology and time constraints

The operating system triggers a real-time audio program at fixed intervals to consume and/or produce buffers of audio samples. We refer to the user-defined function responsible for producing and/or consuming audio data as the **real-time audio routine**. The period of time between successive invocations of this routine is referred to as the **buffer period**. "Low-latency" interactive audio applications, such as music production and performance, ideally employ buffer periods between one and six milliseconds.

To ensure that real-time audio does not glitch, the time taken to execute the real-time audio routine must always be less than the buffer period. As a consequence, all operations performed by the real-time audio routine must be guaranteed to complete in short and bounded time. Informally, we refer to operations with this property as *real-time-safe*.

2.2. Library and system calls are not real-time-safe

General-purpose operating systems do not guarantee worst-case time bounds for system calls. There are many causes of unbounded time behaviour in such systems, including: algorithms with poor worst-case complexity, code that waits for hardware, code that performs blocking interactions with other threads, the thread scheduler and/or the virtual memory paging mechanism.

In particular, memory allocation and unconditional locks (mutexes) are not real-time-safe. Memory allocators often employ algorithms with poor worst-case time bounds. Locks may be subject to unbounded priority inversion.[2]

Furthermore, we can extrapolate that most system APIs are not real-time-safe because they directly allocate memory or use locks, or depend on code that does. An inevitable conclusion follows that any real-time audio routine must avoid calling system APIs. In fact, this is a stated requirement of many real-time audio I/O APIs.

[2] Of Windows, OS X and Linux, only Linux offers the option of real-time-safe mutexes. Windows uses a probabilistic priority inversion avoidance mechanism with no time guarantees. Apple's libc mutex implementation contains the comment "TODO: Priority inheritance stuff." Even on Linux, there is no guarantee that third-party code uses PTHREAD_PRIO_INHERIT.

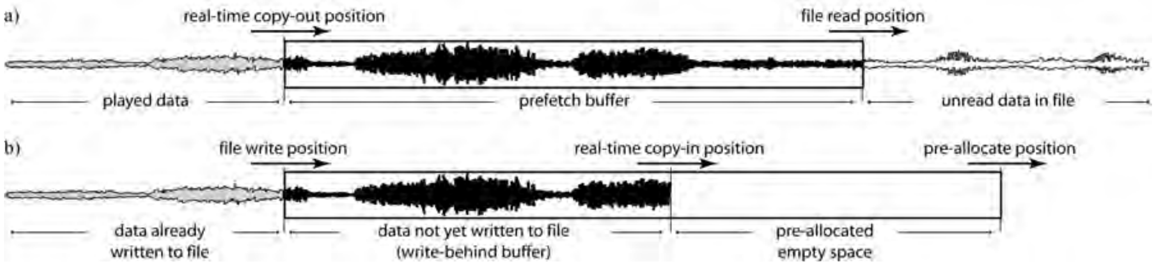

Figure 1: a) Playback buffering scheme, b) Record buffering scheme.

2.3. File I/O is not real-time-safe

File operations–such as opening, closing, reading and writing–take considerable time due to delays inherent to consumer-grade storage devices (hard disks, SSDs). Consider the following:

- Fast spinning disks exhibit seek times of the order of 5 to 7 milliseconds for a single I/O operation.
- SSDs are typically specified to complete 99% of operations in less than one millisecond, but do not guarantee worst-case time for all operations.
- Storage devices are a shared resource. When multiple tasks try to perform I/O simultaneously, operations will be queued and delayed as a consequence.
- Blocking file I/O functions such as `fread()` may wait for data to be read from the storage medium despite the use of anticipatory read-ahead by the operating system.

These considerations are amplified when multiple files are streamed simultaneously, as in a multi-track digital audio workstation.

3. STREAMING PLAYBACK AND RECORD

This section describes, in abstract, two separate processes that may be executed from a real-time audio routine: playing audio from a file and recording audio to a file. We assume a model where the real-time audio routine "pulls" data from the file during playback, and "pushes" data to the file during recording.

3.1. Playback: asynchronous reads and read-ahead buffering

Playback of audio data from a file involves a repeated process of reading audio data from the file and copying the data into an audio output buffer. For this process to be successful, the audio routine must feed audio data to the operating system within a prescribed time frame. Any delay will lead to audible gaps or glitches.

When compared to the time available to run a low-latency real-time audio routine, file I/O operations can take a long time to execute. To avoid stalling the audio routine it is necessary to perform I/O *asynchronously*. This entails splitting I/O operations into two phases: (1) initiate a request to perform an operation (e.g. read data from a file), (2) at a later time, I/O completes and the file data becomes available. Observe that a blocking I/O operation has been converted into two non-blocking operations (*initiate* and *complete*), which are assumed to be real-time-safe.

While asynchronous I/O avoids stalling the audio routine, the potential delay between I/O initiation and completion leaves open the problem of delivering an uninterrupted stream of audio data.

To achieve uninterrupted playback, the audio routine must continuously *prefetch* (read-ahead) audio data from the file into a buffer. In the event that an asynchronous file read operation is delayed, there must be sufficient audio data in the buffer to support uninterrupted playback. Figure 1a illustrates the process.

During streaming playback the real-time audio routine is responsible for two tasks: (1) perform read-ahead by issuing asynchronous file read requests to keep a prefetch buffer full; and (2) copy audio data from the prefetch buffer to the system's audio output buffer. Initially, the real-time audio routine outputs silence until the prefetch buffer becomes full.

3.2. Recording: prefetching buffer space, asynchronous writes and write-behind buffering

Recording audio data to a file involves copying the audio data from the system audio buffer to an intermediate buffer, then writing the buffered data to the file.

As with playback, file writes must be performed asynchronously to avoid stalling the real-time audio routine. The buffering scheme differs from the playback case in that the recording audio routine must always pre-allocate sufficient buffer space to capture incoming audio data. Once filled, the buffer space is passed off to the asynchronous file write operation, which writes the data to file (a process known as *write-behind*). When the write operation has completed, the filled buffer space can be reclaimed or reused. See Figure 1b.

The length of the pre-allocated space must be sufficient to mask the worst-case time taken to allocate more space. The write-behind buffer needs to be long enough to mask the worst-case delay for file write operations.

4. MESSAGE PROTOCOL AND ALGORITHMS

Our implementation strategy uses an "I/O server" thread to perform memory allocation and file I/O. Clients, such as the real-time audio routine, communicate with the I/O server using asynchronous messages. In this section we describe the message protocol for asynchronous file I/O, the streaming algorithms expressed in terms of the protocol, and requirements for a messaging infrastructure suitable for real-time use. In particular we require the availability of a real-time-safe mechanism for sending requests and receiving replies.

4.1. An asynchronous message protocol for file I/O

The asynchronous protocol comprises messages (requests) sent by a client to the I/O server and corresponding results sent by the server in reply. From the client's perspective, *I/O invocation* maps to sending a request and *I/O completion* maps to receiving a reply.

The notation used below reads as follows: `MSG a → b` denotes message `MSG` with parameter `a` sent asynchronously to the server and result/reply `b` later received asynchronously by the client. `MSG a → ◦` is similar except there is no reply and the operation is assumed to succeed. `(a,b,...)` denotes a collection of parameters or results. `a | b` denotes that either result `a` or result `b` is returned.

4.1.1. Opening and closing file handles

`OPEN_FILE (path, accessMode) → fileHandle | error`

Given a file path (a string) and an access mode (read-only, read-write, etc.), open the file and return a file handle or an error code (if an error occurs). The file handle is an opaque identifier used by the client to refer to the file when requesting I/O.

`CLOSE_FILE fileHandle → ◦`

At some point after receiving a valid file handle, the client must close the handle with a `CLOSE_FILE` message.

4.1.2. Data blocks

Read and write operations deal in pointers to server-allocated data blocks that hold regions of file data. All of a file's data blocks contain the same number of audio samples. Data blocks are addressed by their starting position within the file. In addition to the raw file data, each data block maintains book-keeping fields; such as an indication of which areas of the block contain valid data.

4.1.3. Read operations

`READ_BLOCK (fileHandle, position)`
`→ dataBlock | error`

Given a file handle and a position in the file, allocate a data block, read data from the file into the block, store the valid data range, and return the data block or an error code (if an error occurred).

`RELEASE_READ_BLOCK (fileHandle, dataBlock) → ◦`

Data blocks acquired with `READ_BLOCK` must be released using the `RELEASE_READ_BLOCK` message, which causes the data block to be deallocated.

4.1.4. Write operations

`ALLOCATE_WRITE_BLOCK (fileHandle, position)`
`→ dataBlock | error`

Given a file handle and a position in the file, allocate the data block in which the client will write data for a specified region of a file. If the specified region is a valid area of the file, fill the data block with data from the file (similar to read), otherwise leave the data area uninitialised and mark it as invalid. Return the data block or an error code (if an error occurs).

After acquiring a block with `ALLOCATE_WRITE_BLOCK`, the client must either commit or release the block using either `COMMIT_MODIFIED_WRITE_BLOCK` or `RELEASE_UNMODIFIED_WRITE_BLOCK`.

`COMMIT_MODIFIED_WRITE_BLOCK (fileHandle, dataBlock)`
`→ ◦`

`COMMIT_MODIFIED_WRITE_BLOCK` writes the block's data to the file and releases the data block.

`RELEASE_UNMODIFIED_WRITE_BLOCK`
`(fileHandle, dataBlock) → ◦`

`RELEASE_UNMODIFIED_WRITE_BLOCK` deallocates the data block without writing any data to the file. This message should only be used if the block was not modified.

4.1.5. Message ordering constraints

A client may only issue `READ_BLOCK` or `ALLOCATE_WRITE_BLOCK` requests for a particular file handle prior to sending the `CLOSE_FILE` message. This requires the server to process requests in first-in first-out (FIFO) order, otherwise the server might encounter a block request after having closed the file. `COMMIT_MODIFIED_WRITE_BLOCK` requests that extend the file length also require that requests be processed in order, unless the file system supports files with "holes."

The client may `RELEASE` or `COMMIT` acquired data blocks after sending the `CLOSE_FILE` message. The I/O server only closes the native file handle once all data blocks have been released or committed.

4.2. Streaming algorithms

We now present the playback and recording algorithms. The algorithms assume that a file handle has been obtained by sending the `OPEN_FILE` message and receiving the resulting file handle.

Algorithm: streaming file playback. Maintain a prefetch queue referencing N data blocks ordered by ascending file position. Maintain a playback index referring to the next audio sample to output. At each time-step, output a single audio sample. (See Figure 2a.)

1. [Setup.] Set the playback index to zero. Issue N `READ_BLOCK` requests for sequential data blocks. Order the block references in the prefetch queue, marking all blocks as *pending* until they have been received. Output silence until the replies to all N `READ_BLOCK` requests have been received. At each subsequent audio time-step perform the following three steps:
2. [Retiring replies.] Whenever a `READ_BLOCK` reply is received, mark the block as *received* or as *error* if an error result is received.
3. [Output an audio sample.] Output (copy out) the sample referred to by the playback index–this indexes into the front-most data block in the prefetch queue. Output silence if the block is *pending* or if an *error* is indicated.
4. [Advance playback index, maintain the prefetch queue.] Increment the playback index. If the index now points into the second block in the queue: (1) pop the first block from the queue and issue a `RELEASE_READ_BLOCK` request for that block; (2) issue a `READ_BLOCK` request for the block following the latest block requested so far. ∎

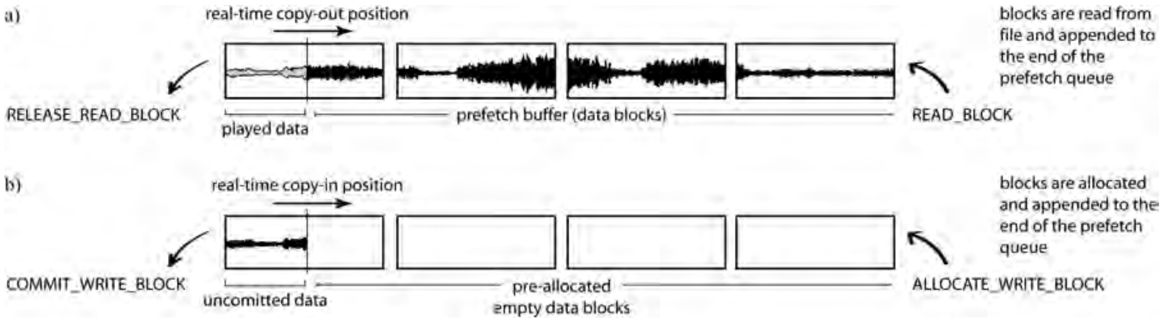

Figure 2: a) A prefetch queue of data blocks for playback, b) A prefetch queue of data blocks for recording

The algorithm for recording is similar to the above, except that data is copied in to the blocks rather than out of them (see Figure 2b). The messages are changed as follows: `READ_BLOCK` becomes `ALLOCATE_WRITE_BLOCK`. When data is copied into a block the block is marked as *modified.* Modified blocks are released using `COMMIT_WRITE_BLOCK`. Unmodified blocks are released using `RELEASE_UNMODIFIED_WRITE_BLOCK` (e.g. to flush the prefetch queue when recording stops).

4.3. Implementation requirements and desiderata

Any implementation of the messaging protocol must fulfil the following requirements:

- Allocating, deallocating, sending requests and receiving replies must be real-time-safe.
- The protocol requires the server to process events in FIFO order. Clients may receive replies in any order.

In addition, the following desiderata would contribute to the usability and flexibility of a solution:

- Support immediate disposal of streaming state, without having to wait for pending replies.
- Construction and tear-down of streaming state should be real-time-safe.
- Support creation and destruction of streams in any thread, and allow stream states to migrate between threads (e.g. create a stream in one thread and use it in another, or use a stream in different threads at different times). This implies being able to send requests and receive replies from any thread.
- Do not use thread-local storage or client-managed inter-thread queueing infrastructure.

The first point is desirable when combining asynchronous streams with imperative code. For instance it is desirable to be able to embed a stream in a C++ object, and to delete the object without having to wait for the stream to receive and process pending asynchronous replies.

The last three points express a desire for programming flexibility and minimising maintenance overhead. They seek to avoid the inconvenience of using rigid topologies of queues and threads, and/or requiring certain operations to be performed in certain threads.

Stream operations are not required to be thread-safe or re-entrant. Stream state may be passed between threads but not simultaneously accessed from multiple threads.

5. IMPLEMENTATION IN C++

Figure 3 shows the multi-threaded communication structure of the solution. The I/O Server thread has a queue (mailbox) for inbound requests. Client streams send requests to the server by enqueueing Request objects into the server's mailbox.

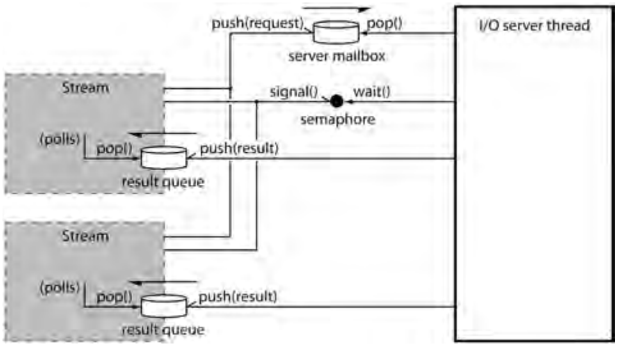

Figure 3. Multi-threaded communication structure showing two client streams.

Streams are not associated with a particular thread and may move freely between threads. In contrast to architectures that employ an incoming message queue for each thread, we use a separate result queue for each client stream. When the server completes a request, it posts the reply (if any) into the result queue specified by the request. Client streams poll their result queues as part of the playback or recording process. There is no requirement for clients to be signalled by the server.

5.1. Requests and message queues

We now turn our attention to the implementation of the Request objects and lock-free message queues.

5.1.1. Request objects

A single fixed-size Request data type (a C/C++ `struct`) is used to represent all message protocol requests and replies. Replies are returned to the client using the same Request instance that was used to initiate the request. The Request data type includes a tag indicating the specific request type (`OPEN_FILE`, `CLOSE_FILE`, etc.), a union containing the parameters for each request type, and fields common to all requests (link pointers, result code, a pointer to a client queue to return replies/results, and fields reserved for client-only use).

Each Request contains two “next” pointer fields that allow the Request to be simultaneously linked into two separate linked structures or queues. One next pointer is used for linking the request into client-server communication data structures, for example to enqueue the Request to the server, or to enqueue a reply to the client. A second client-only next pointer is used by the client to link Requests together into client-local data structures.

With the exception of client-only fields, the client should not modify a Request object from the time that it is sent to the server until its reply is received. When Request objects are not queued with the server, all Request fields are available for client use.

A global lock-free freelist of Request objects supports allocation and deallocation from any thread. The freelist is implemented using the "IBM Freelist" algorithm (IBM 1983; Treiber 1987; Michael and Scott 1998), a last-in first-out (LIFO) stack that supports non-blocking push and pop operations.

5.1.2. A lock-free "pop-all" LIFO stack

Listing 1 shows a lock-free LIFO stack algorithm that supports *push, is-empty* and *pop-all* operations.[1] It supports concurrent operations from multiple threads. We use this algorithm as the basis of our message queues.

```
struct Node { Node *next; };
struct Stack { Node *top; };
void init(Stack& s) { s.top = NULL; }
void push(Stack& s, Node *n, bool& wasEmpty) {
  do {
    Node *top = s.top;
    n->next = top;
    wasEmpty = (top==NULL);
    // CAS: atomic compare-and-swap
    // set s.top to n only if s.top == top
  } while(!CAS(&s.top, top, n));
}
bool is_empty(Stack& s) { return (s.top==NULL); }
Node *pop_all(Stack& s) {
  if (s.top==NULL) return; // don't modify if empty
  // XCHG: atomic exchange
  // set s.top to NULL, return old s.top
  return XCHG(&s.top,NULL);
}
```

Listing 1: "Pop-all" concurrent LIFO stack algorithm.

The algorithm works as follows: `s.top` points to the top of a LIFO stack of linked nodes terminated by a `NULL` value. The push algorithm (identical to the IBM freelist) attempts to link a node onto the top of the stack, and retries if a conflict with a concurrent operation is detected. The pop-all algorithm uses an atomic exchange operation to replace `top` with `NULL` and returns the previous contents of the stack–a `NULL`-terminated linked list in LIFO order.[2]

5.1.3. Sending requests to the server

In section 4.3 we established that messages should be received and processed by the I/O server in FIFO order, and that it is desirable to support posting requests to the server from arbitrary threads. To achieve these goals we use a multiple-producer single-consumer (MPSC) FIFO queue. A simple algorithm with this property is the so-called "reversed IBM freelist."

The algorithm can be described in terms of the pop-all LIFO stack from the previous section: Producers enqueue requests by pushing them onto the pop-all stack. The consumer maintains a separate consumer-local stack in FIFO order. Requests are dequeued from the consumer-local stack. When the consumer-local stack is empty, the consumer checks whether the pop-all stack is empty, if not it pops *all* items from the pop-all stack and reverses their order into the consumer-local FIFO queue, from whence further requests are dequeued. The result is a very simple lock-free MPSC FIFO queue.

When a client enqueues a new request it signals the server using a semaphore (or on Windows, an auto-reset Event Object). Since the server only waits on the semaphore when its lock-free stack is empty, the client need only signal the semaphore when it pushes a request onto an empty stack. The pop-all stack's push operation indicates when it has pushed onto an empty stack.

5.1.4. Receiving results from the server: result queues

Each stream has its own **result queue**, an object comprising a lock-free queue and a book-keeping counter. Results are only enqueued onto result queues by the I/O server, and dequeued (polled) by the stream that owns the result queue. The streaming algorithms do not require replies to be returned in order, so the queue does not need to guarantee delivery order. We use a lock-free single-producer single-consumer (SPSC) relaxed-order queue. It can be implemented similarly to the reversed queue from the previous section, with reversing omitted.

A result queue maintains a count of expected results. The client increments the count when it sends a request that expects a reply, and decrements the count when the result is received. When the count drops to zero no further results are expected. (Note that this scheme precludes requests with optional replies.)

We embed result queues directly in Request objects. Result queues only use one or two words of storage, and hence can be stored in the request-type-specific parameter area of a Request. This lets us allocate result queues using the real-time-safe global Request freelist and to send and receive them as messages. The latter feature is used by the result queue clean-up process described later.

5.1.5. The server process

The server waits (blocks) on a semaphore until new requests are posted to its queue. Received requests are processed sequentially in FIFO order.

In order to manage the lifetime of native file handles, the server associates a reference count with each open file. The count is incremented when the server returns a file handle to the client in response to `OPEN_FILE`, and whenever the server returns a block to the client. The count is decremented on `CLOSE_FILE` and whenever a block is received from the client. The native file handle is closed when the reference count drops to zero.

---

[1] The presentation here lacks memory fences and atomic loads and stores. Check the example code for a more complete implementation.

[2] Although of uncertain origin, this algorithm, as well as the reversing method discussed next, is well known to those familiar with the art of lock-free programming, see e.g.: https://groups.google.com/d/msg/lock-free/i0eE2-A7eIA/g745KEEx2JIJ

5.2. Client streams

We now present the client stream data structure and its life cycle: creation, maintenance, and destruction.

5.2.1. Client stream data structures

All client stream data structures are built by linking together Request objects. Request object allocation and deallocation is real-time-safe. So too is allocation and deallocation of stream data structures.

A fully constructed stream is illustrated in Figure 4. A stream comprises a result queue Request (for receiving results from the server), an `OPEN_FILE` Request (which is transformed to `CLOSE_FILE` when closing the stream), and zero or more block Requests linked into a tail-queue that serves as the stream's prefetch buffer. A stream is referenced by client code as an opaque pointer (handle). Internally the pointer points to the result queue.

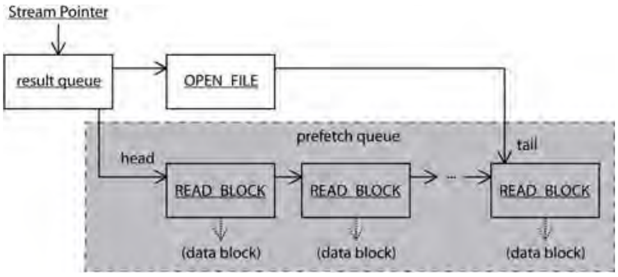

Figure 4: A fully constructed stream using linked Request objects. Each solid rectangle is a Request.

In addition to the result queue, `OPEN_FILE` Request and prefetch queue, a stream maintains the following state variables:

- The overall state of the stream; one of: {OPENING, OPEN_IDLE, OPEN_BUFFERING, OPEN_STREAMING, ERROR}.
- A count of pending blocks (those that have been requested but not yet returned to the stream).
- An index that tracks how much data has been copied to or from the front-most data block.

These state variables are stored in unused and client-use Request fields.

5.2.2. Creating and opening a stream

To create and open a stream, the client first creates two Request objects: a result queue and an `OPEN_FILE` request. The `OPEN_FILE` request is sent to the server, specifying the result queue as its return address. The stream state is set to OPENING, and an opaque stream pointer that points to the result queue is returned to the client. Some time later the `OPEN_FILE` request is returned in the result queue. When the stream gains control it dequeues the result and handles it as follows:

- If the `OPEN_FILE` request was successful, it is linked to the result queue. It contains the file handle that will be used for all block requests. The stream state is set to OPEN_BUFFERING and the prefetch queue pointers are cleared to indicate that the prefetch queue is empty.
- If the `OPEN_FILE` Request failed, the error result is copied to the result queue's error field, the stream state is set to ERROR, and the Request object is deallocated.

5.2.3. Maintaining the prefetch queue

The prefetch queue is represented using a singly-linked tail queue. A tail queue affords constant-time insertion at the back and constant-time removal from the front. The queue is formed from pending and completed `READ_BLOCK` or `ALLOCATE_WRITE_BLOCK` requests (see Figure 4).

The reading or writing process works by copying data from or to the front-most block in the prefetch queue. A numeric field in the front-most block request is used to track how much data has been copied so far. Once the process has finished with a block, the block is returned to the server and a new block is requested and inserted at the back of the queue. Block requests are linked into the queue as soon as they are created.

The block request's request type field is also used (overloaded) to track the state of each data block. The block state is one of: {PENDING, READY, MODIFIED, ERROR}. The request type (`READ_BLOCK` or `ALLOCATE_WRITE_BLOCK`) denotes the PENDING state. When a request's reply is received from the server, the block's state is set to READY or ERROR. The MODIFIED state is used to determine whether a recording stream should `COMMIT` or `RELEASE` the block.

In its simplest form, *seeking* clears the prefetch queue, returns all acquired blocks to the server, and issues `READ` or `ALLOCATE` block requests starting at the new seek position. Care must be taken with PENDING requests since they are yet to arrive in the result queue. They can be removed from the prefetch queue if flagged "return to the server on arrival."

Operations on the prefetch queue also affect the stream's global state. If the streaming process finds that the front-most block is PENDING, it signals an underflow condition and causes the stream state to be set to OPEN_BUFFERING. The stream state is set back to OPEN_STREAMING when all pending blocks have been received (this condition is detected using the pending blocks count). The BUFFERING and STREAMING states support behaviour that suspends playback while the stream is buffering. If the prefetch buffer is correctly sized, the BUFFERING state should only be encountered immediately after seeking.

5.2.4. Destroying a stream

To destroy a stream, all resources associated with the stream must be released. The destruction process needs to account for the case where Requests are in flight when the client instigates stream destruction. For example, the stream may be waiting for the reply to an `OPEN_FILE` request or to one or more `READ_BLOCK` requests. In these cases the result queue can not be destroyed until all results have been received. Then once received, results must be cleaned up: a file handle returned by `OPEN_FILE` must be closed, blocks acquired by `READ_BLOCK` or `ALLOCATE_WRITE_BLOCK` must be released.

The stream destruction algorithm makes use of three facts: (1) each request that returns a resource has a deterministic clean-up procedure, (2) the result queue can be sent as a message to the server, and (3) the result queue includes a counter indicating the number of replies that it is expecting. The counter makes it possible to determine when the result queue has received all replies, and hence when it can be safely destroyed.

Destruction of a stream that is in any of the OPEN_... states involves the following algorithm:

Algorithm: destroying an open stream.

1. [Handle block requests in the prefetch queue.] For each block request in the stream's prefetch queue: If the block request is READY or MODIFIED, transform the block request into a `RELEASE` or `COMMIT` message respectively, and send the request to the server (the server will release or commit the block). Discard references to PENDING block requests, they have yet to arrive in the result queue and will be cleaned up later. Deallocate requests in the ERROR state directly to the global freelist.
2. [Close the open file handle.] Transform the `OPEN_FILE` request into a `CLOSE_FILE` and send it to the server.
3. [Dispose the result queue.] If the result queue's expected result count is zero, deallocate the result queue's containing Request object to the global freelist. Otherwise, send the result queue to the server using a newly defined message type: **`CLEANUP_RESULT_QUEUE resultQueue`** → ∘
   This message delegates responsibility for cleaning up any pending requests to the server (see next section). ∎

Destruction of a stream in the OPENING state sends the result queue request to the server as a `CLEANUP_RESULT_QUEUE` request. Destruction of a stream in the ERROR state deallocates the result queue Request object to the global freelist.

5.2.5. Server processing of `CLEANUP_RESULT_QUEUE`

When the server receives a `CLEANUP_RESULT_QUEUE` request it pops any queued results from the result queue. For each popped result it decrements the result queue's expected result count and performs the result's compensating operation (e.g. `OPEN_FILE` transforms to `CLOSE_FILE`, `READ_BLOCK` transforms to `RELEASE_READ_BLOCK`). If the result queue's expected result count is zero, the result queue is deallocated. Otherwise, the result queue is marked with an *awaiting clean-up* flag. Subsequently, whenever the server completes a request, it checks the result queue's flag. If the flag is set, rather than enqueueing the reply, the server decrements the expected result count and performs the compensating operation. Once the expected result count drops to zero the result queue is deallocated to the global freelist.

The result queue clean-up process ensures that all resources associated with a stream are released. It also provides a non-blocking real-time-safe mechanism for destroying streams. As the result queue is itself a Request object, the `CLEANUP_RESULT_QUEUE` message can always be issued. It cannot fail due to failure to allocate a Request.

6. DISCUSSION AND FUTURE WORK

The presented algorithms for real-time stream operations have time complexity of either O(1) or O(N) in the length of the prefetch queue. Strategies can be devised to reduce the cost to O(1) by moving responsibility for prefetch queue creation and tear-down to the I/O server.

Ideally the protocol would not require Request objects to be allocated or deallocated while the stream is running. Request allocation and transport overhead can be reduced by altering the protocol to allow requests to be reused. Methods to achieve this include: always return results, provide messages that perform multiple operations (e.g. read N blocks), and combine messages (e.g. combine `RELEASE_BLOCK` and `READ_BLOCK` into a single Request).

The protocol described in section 4 does not communicate the file length or audio data format information to the client stream. If needed, this information can be communicated as additional results from `OPEN_FILE`.

As presented, the method does not reliably support multiple real-time streams. The seek process injects N consecutive `READ_BLOCK` operations into the I/O server queue. These operations may delay more urgent operations that are enqueued later. To address this, assign a deadline to each request and have the server perform operations in earliest-deadline-first (EDF) order.

The performance of the lock-free queues used here are sufficient for the expected contention rate. In a high-throughput scenario other lock-free and wait-free queue algorithms should be considered.

Some messages involve shared access to the Request object by both client and server. This does not result in any data races since access is limited to disjoint fields in each thread. It does however introduce the possibility of false-sharing. It would be interesting to investigate the impact of false-sharing on protocol performance.

We have omitted discussion of handling formatted sound files (.wav, .aiff, mp3, etc.). One way to handle formatted sound files is for the I/O server thread to perform formatted I/O using an existing sound file I/O library. Alternatively, the I/O server could be extended to implement sound file container parsing and audio format conversion.

This paper has presented a simplified model of a real-time file streaming system currently under development by the author. The system under development is designed to support multi-threaded access to streaming file I/O for a multi-core-capable real-time audio engine. Beyond what has been presented here, the system supports caching and sharing file handles and data blocks amongst multiple client streams. Requests may be prioritised, re-prioritised and cancelled. Multiple native I/O operations may be queued to the operating system concurrently. The system supports parsing sound file containers and transparent data format conversion. All of this is achieved using an asynchronous messaging model similar to the one presented here.

7. RELATED WORK

I/O has been an important aspect of computing from the beginning. Knuth (1997) surveys early work and describes I/O using linked chains of buffer descriptors.

Operating system kernels use queues of linked I/O requests (Comer 2011). Windows NT's *I/O request packets* (Russinovich et al. 2012) have their origins dating back at least to DEC's RSX11 operating system (Pellegrini and Cutler 1974). The Parameter Block API of Mac OS Classic (Apple 1985) is an asynchronous file I/O API that entails enqueueing parameter blocks (requests) to be processed asynchronously by the operating system.

Fixed-size I/O data blocks are commonly used for device-independent kernel interfaces (Tanenbaum 2001). Pai and colleagues (2000) conducted an extensive study of models for exchanging I/O buffers. Brustoloni (1997) provides a useful taxonomy of buffer sharing models.

The solution described here is an instance of the "Half-Sync/Half Async" design pattern (Schmidt and Cranor 1995), with the variation that clients do not block if data is unavailable. Asynchronous message queueing and fixed size allocation pools are standard techniques in real-time systems development (Douglass 2003).

For a gentle introduction to lock-free algorithms, see (Michael 2013). The *Synthesis* operating system kernel is notable for its use of lock-free data structures for queueing (Massalin and Pu 1992). The lock-free pop-all LIFO and reverse IBM freelist described in this paper, were encountered on the comp.programming.threads newsgroup during 2005-2008. Participants included Joe Seigh, Chris Thomasson and Dmitry Vyukov.

Lock-free ring buffers are used for transferring audio data and messages in real-time audio applications, primarily as a means to avoid priority inversion, as in for example PortAudio (Bencina and Burk 2001). Lock-free techniques for computer music systems were discussed by Fober and colleagues (2002). A concrete example of use is James McCartney's SuperCollider 3 synthesis server, scsynth (2002). See (Bencina 2011) for an analysis of the message passing techniques in scsynth. Another application of lock-free techniques in computer music is Shelton's real-time live coding environment (2011).

Anderson and colleagues (1997) describe a real-time video conferencing application using a scheme of lock-free queues for inter-thread communication. That publication is notable for its analysis of the hard-real-time-safety of lock-free queues on a uniprocessor. A related analysis is given by (Cho 2006). The author is not aware of similar analysis for multi-processor systems.

8. ACKNOWLEDGEMENTS

Many thanks to Tony Holzner, Scott Brewer, Phil Burk, Andrew Bencina and the anonymous reviewers for their helpful feedback on earlier drafts of this paper.

9. REFERENCES

Anderson, J. H. et al. 1997. "Real-Time Computing with Lock-Free Shared Objects." *ACM Transactions on Computer Systems.* 15(2):134–165.

Apple Computer, Inc. 1985. *Inside Macintosh, Volume II.* Reading, Massachusetts: Addison-Wesley, pp. II-97 – II-119.

Bencina, R. and P. Burk. 2001. PortAudio – an open source cross platform audio API. In *Proceedings of the 2001 International Computer Music Conference*.

Bencina, R. 2011. "Inside scsynth," In Wilson, Cottle, Collins eds. *The Super Collider Book,* Massachusetts: The MIT Press, pp. 721–740.

Brustoloni, J. C. 1997. "Effects of Data Passing Semantics and Operating System Structure on Network I/O Performance." Doctoral thesis. Carnegie-Mellon University Pittsburgh PA, School of Computer Science.

Cho, H. et al. 2006. "Lock-Free Synchronization for Dynamic Embedded Real-Time Systems." In proceedings *Design, Automation and Test in Europe, DATE '06.*

Comer, D. 2011. *Operating System Design - The Xinu Approach, Linksys Version.* CRC Press, pp. 373–381.

Douglass, B. P. 2003 *Real Time Design Patterns.* Boston: Addison-Wesley, sections 5.3 and 6.3.

Fober, D. et al. 2002. "Lock-Free Techniques for Concurrent Access to Shared Objects." *Actes des Journ'ees d'Informatique Musicale JIM2002*, Marseille, pp.143–150.

International Business Machines Corporation (IBM). 1983. *IBM System/370 Extended Architecture, Principles of Operation. First Edition,* pp. A-44 – A-45.

Knuth, D. E. 1997. *The Art of Computer Programming, Volume 1.* 3rd Ed. Upper Saddle River, NJ: Addison-Wesley, pp. 215–231.

Massalin, H. and C. Pu. 1992. "A Lock-Free Multiprocessor OS Kernel." ACM SIGOPS Operating Systems Review, 26(2):108.

McCartney, J. 2002. "Rethinking the computer music language: SuperCollider." *Computer Music Journal* 26(4) 61–68.

Michael, M. M. and M. L. Scott. 1998. "Nonblocking Algorithms and Preemption-Safe Locking on Multiprogrammed Shared Memory Multiprocessors." *Journal of Parallel and Distributed Computing* 51(1):1–26.

Michael, M. M. 2013. "The Balancing Act of Choosing Nonblocking Features." *ACM Queue* 11(7).

Pai, S. et al. 2000. "IO-Lite: A Unified I/O Buffering and Caching System." *ACM Transactions on Computer Systems.* 18(1):37–66.

Pellegrini, M. and Cutler, D. 1974. "RSX-11M Working Design Document," Digital Equipment Corp, Maynard Mass, p. 35.

Russinovich, M., et al. 2012. *Windows Internals.* 6th ed. Part 2. Redmond, Washington: Microsoft Press, p. 28.

Schmidt, D. C. and Cranor, C. D. 1995. "Half-Sync/Half-Async - An Architectural Pattern for Efficient and Well-structured Concurrent I/O." In *Proceedings of the 2nd Annual Conference on the Pattern Languages of Programs.*

Shelton, R. J. 2011. "A Lock-Free Environment for Computer Music: Concurrent Components for Computer Supported Cooperative Work." PhD thesis. The University of Melbourne, Department of Computer Science and Software Engineering.

Tanenbaum, A. S. 2001. "Modern Operating Systems," 2nd Ed. 2001, New Jersey: Prentice Hall. p. 298.

Treiber, R. K. 1986. "Systems Programming: Coping with Parallelism". Technical Report RJ 5118, IBM Almaden Research Center.

DIGITAL COMPOSITION SYSTEMS: USER EXPERIENCE STUDY

Liam Bray
The University of Sydney
Faculty of Architecture, Design and Planning
Sydney, Australia
liamhbray@gmail.com

Oliver Bown
The University of Sydney
Faculty of Architecture, Design and Planning
Sydney, Australia
oliver.bown@sydney.edu.au

ABSTRACT

Digital composition systems shape the way a composer creates a composition. Several novel methods of digital composition have demonstrated that open-ended possibilities in the design of composition interfaces can be expanded on. Generative non-linear processes have demonstrated their usefulness and ability to aid creative processes. This study investigates what tasks are effectively achieved and what problems arise when using two different types of digital composition systems, specifically focusing on linear and non-linear systems.

Keywords

Digital Composition Systems, User Study, Generative Composition

1. COMPOSITION SYSTEMS

Avid Protools and systems similar to it such as Apple Logic and Cubase are considered industry standard digital audio workstations (DAW's) amongst both amateur and professional music producers. Within this group of programs there is little deviation from what would be considered standard user interfaces. Design of these interfaces is closely abstracted from multitrack recorders and mixing consoles historically used in the analogue process of recording to magnetic tape (Theberge 2004). DAW's also have basis in western musical notation, allowing a user to listen to a linear series of sequenced events measured against a grid which represents time and musical measure. This is coupled with the literal abstraction of the piano into the piano roll, with both pitch and velocity being represented as linear sequences of events.

Development of novel composition systems has taken on a new form in recent years as platforms such as MaxMSP have now developed simple user friendly ways to integrate with industry standard DAW's. As designing tools for creating music has become easier, the demand for diversity in digital musical tools for both live music performance and composition has risen. Large online communities now collaborate to design musical tools. MaxMSP and Pure Data have large databases of patches available online with new and novel tools constantly being developed and refined. Popularity of high-level programming platforms for creative purposes has culminated in a collaboration between Cycling74 and Ableton. Max for Live (M4L) is the integration of MaxMSP into Ableton Live as a plug-in. This enables designers to create and distribute MaxMSP patches easily to an even broader audience and allows novice Ableton Live users to integrate experimental composition tools into their workflow easily.

Nodal [9] is an example of a novel composition tool that can be easily integrated into workflow within a DAW using MIDI event sequencing. Nodal is a non-linear composition system in which artists are able to create generative systems consisting of cyclic structures and networks of events. Generative systems such as nodal are able to facilitate complex emergent structures, but this is often at the expense of the system's usefulness to create compositions in a traditional sense and requires users to think differently about the composition process. McCormack [8] addresses this paradigm as the focus of Nodal's design.

> The aim of the work...is to design and build a generative composition tool that exhibits complex emergent behaviour, but at the same time offers the composer the ability to structure and control processes in a compositional sense.

Nodal's design is intended to provide a fundamentally different approach to composition while attempting to retain the system's usefulness through intuitive visual mapping and simple event rules. Nodal (figure1) displays musical events as a grid of nodes; the space between nodes represents time and the connection between nodes indicates the direction of the next event. Nodes have the ability to be sequential, thats is they continue to follow a path or in the case of multiple nodes flow in sequential order. Parallel nodes take a single event and send it on multiple paths at once and random nodes select a path at random.

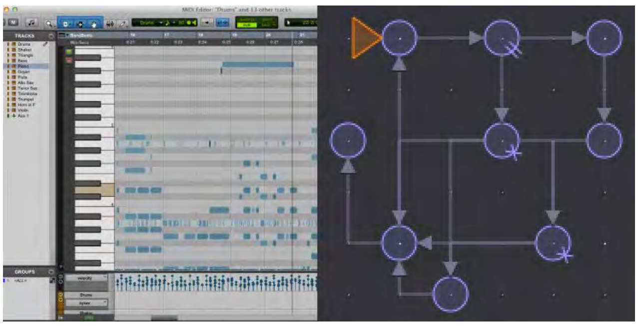

Figure 1: A Nodal network created by a participant

2. DESIGN PARADIGMS FOR COMPOSITION

Digital composition systems are designed to allow musicians and music producers to deal with extremely complex amounts of musical information. These complex digital representations are inherently difficult to manage for users. Gaining a better understanding of how users develop strategies to complete complex compositional tasks in both traditional DAWs and novel composition systems will help to identify some of the fundamental differences in design that can be attributed to the usefulness of managing complexity within the system.

For users, learning and identifying tools and limitations afforded by a system is an iterative process and is simultaneous to the discovery of the potential compositional results the system is able to achieve [11]. This design problem for DAWs is summarised by Resnick et al. as ”Low Threshold, High Ceiling, and Wide Walls”. Effective tool designs should be simple enough for new users to get started, complex enough for experts to create sophisticated results and should support a wide range of explorations, such as different ways to control interaction with the system. [14] Additionally the opacity or the ability of the user to create a conceptual model of the system is paramount in enabling users to develop strategies easily. For new users opaque systems can lead to uncertainty, a lack of understanding and frustration.

Traditional DAWs are often focused on recording a performance of MIDI (Musical Instrument Digital Interface) or audio, in realtime. However after this performance has been captured the system's primary function is to edit and manipulate this information. Some component of this is the documentation of a performance for listening, but arguably the primary function of a DAW is to enable the composer to continue the composition process digitally. Ableton Live is a good example of this; the DAW has the ability to record audio, but the primary focus of it's design is the creation of electronic, or elctro-acoustic music. This is an example of a reversal of the traditional composition process: the composer creates the notation before the music is performed, and the musical structure may be sketched in abstract form before a more concrete instantiation is required [3].

One tool used frequently in composition systems to help users visualise sequences of events is the playhead. In Protools the playhead is represented by the scrolling bar which travels from left to right. This bar can be placed in linear time represented by the notated scale and playback will begin where it is placed. In Nodal the playhead is simply the change in colour that takes place when an event is triggered.

The systems used in this study require the user to develop different strategies to achieve complex musical compositions. Protools becomes more complex with the layering and intensity of sounds as well as increased frequency of events or variation within events, but Nodal becomes quite complex with few events as both randomness and chance can be introduced with a minimal series of events. In both cases feedback is provided to the user both aurally and visually. This feedback is the primary means by which the composer iteratively works with the system. Liveness [3] describes the type of feedback a user receives from a system as well as how responsive that feedback is when a user performs an action. Liveness as a measure is useful in defining the principal differences in feedback between composition systems.

Nodal incorporates some basic aspects of a music tracker such as VSTrack [12] as composition in nodal requires the mapping of at least some events prior to the execution of a sequence. However in contrast to a tracker the non-linear generative nature of Nodal makes it harder to follow when more complex event sequences take place. Nodal as well as being executable is edit triggered: while a sequence of events is playing edits to note sequence, pitch and timing are able to be made.

Protools is divided into three different interfaces track view, mixer, and midi editor. This study focuses on the piano roll found in the midi editor. The piano roll can be used to record performances and in that sense it is a stream-driven environment that is constantly active and continually visualising notations as they are made [3]. It can also be used to map events which can then be executed. This is limited to a single sequence of events: requiring any further sequences in the same pitch range to be divided amongst additional tracks.

Athanasopoulos et al. [1] have demonstrated in a cross-cultural study of people from the U.K., Japan and Papua New Guinea that in the absence of general or musical literacy people describe variation in time and pitch idiosyncratically. This highlights that some of the fundamental principals of western musical notation aren't necessarily self-evident. One of these being the concept of an independent sound object [4] which is inherent in western notation. For people literate in western standard notation a cartesian approach to musical notation is primarily adopted. But what was found when the BenaBena people of Papua New Guinea described musical pitch variation over time is that the concept of an abstract sound object was not evident. Structural variations that corresponded to pitch variation were seen as irrelevant and the participants attempted to illustrate the sound's mode of production through the use of culturally significant representations of the sound. This raises many questions for the design of novel composition systems. The usefulness of a system will always have to contend with user's cultural understanding of musical notation and the abstract concepts that are implied by that understanding. Mapping generative systems within a fundamentally cartesian understanding of musical notation dictates limitations. Determining if novel composition methods are better at facilitating the management of complexity is beyond the scope of this study, instead what is emphasised is the importance of the continued development of composition systems to include more than what is currently considered useful.

3. METHOD

This study aims to further understand how users develop strategies to complete complex tasks in linear and non-linear composition systems by looking at users experience and self-assessed ability to complete tasks. Primarily the study is focused on assessing three criteria:

1. Is the user effectively able to achieve complex compositional tasks?

2. What is the users experience of the system?

3. What are the relationships between the characteristics of the system and the musician's experiences and actions? [7]

Participants where asked to use both Nodal and Protools to achieve simple and complex compositional tasks. Participants were self described as having a musical background of varying types and skill levels. Prior to the commencement of the study a short demonstration was provided for each system to ensure a base line of knowledge for all participants. Tasks required users to develop and execute a strategy for achieving the task and then to reflect on the system's design and their self-assessed level of ability to complete various components of the task. This was measured using two methods. Firstly a five point Likert scale from never to always with questions comparative to the Cognitive Dimensions of Notations framework [5] which focuses on comparing the user's reflections on the usability of computer based systems by identifying conceptual models employed by notational systems. It has similarly been applied to music typesetting as well as live-coding musical systems [2] Some questions included in the questionnaire were:

- I feel I am able to complete the task successfully
- I have total control over the system
- I have a strong sense of what I have to do to complete the task

After completing a task in both systems participants were asked to asses how they felt about the task as well their experience of each system comparatively. This was measured using a five point Likert scale from simple to frustrating. The questionnaire consisted of questions based on the psychometric scale used to measure flow. [13] Some questions included in the questionnaire were:

- Find the tools I need to use to complete operations
- Create new ideas I have spontaneously thought of
- Get the system to do what I want

Participants were also filmed during the study, this was to determine length of time they took to complete tasks, review strategies participants developed and iterations of compositions as they developed as well as record comments participants made about the process during the task. The outcome of these two methods of analysis and the video documentation is a detailed picture of a user's self-assessed experience and strategy when attempting to complete a task within a composition system.

The study consisted of seven participants of varying levels of musical ability and familiarity with DAWs. Each participant was asked to complete two tasks in both Protools and Nodal. The first task was to create a soundscape for a video game. The soundscape had to be sufficiently long enough that it would not become boring or repetitive when used in the video game, requiring participants to create variations on a theme. The second task was to create an improvised solo to the twelve bar blues. The twelve bar blues was provided for the participants, and they were allowed to copy and paste from the existing patterns to create new ones. This task required participants to pay careful attention to melodic structure, harmony, disharmony and timing.

4. RESULTS

4.1 Reorganisation Of Events

The study identified that regardless of the participants level of experience with composition systems, they found reorganising sounds in order to modify the order of something they had composed harder in Nodal than in Protools. (Figure 2)

For example, participants with more than one year experience using DAW's (Protools and Ableton) did not find it any easier than participants with less than one year or no prior experience using DAWs. This result identifies an inherent design in non-linear systems. Manipulation of events within linear timeline based systems involves direct manipulation of events, such as the alignment and placement of notes on a piano roll. Whereas generative systems are designed to allow users to set up systems of events which then take place over time, which is the indirect manipulation of events. One implication of this is deciphering the outcomes of notations, in order to further understand how to achieve a desired result, is a slow and difficult process. This is supported by the study's results. When participants were asked if they felt they have total control over the system when completing a task, 57% answered frequently in protools, compared to 28% in Nodal. Once a Nodal Network becomes sufficiently complex it becomes difficult to visually track events. This lowers the amount of tangible feedback provided by the system.

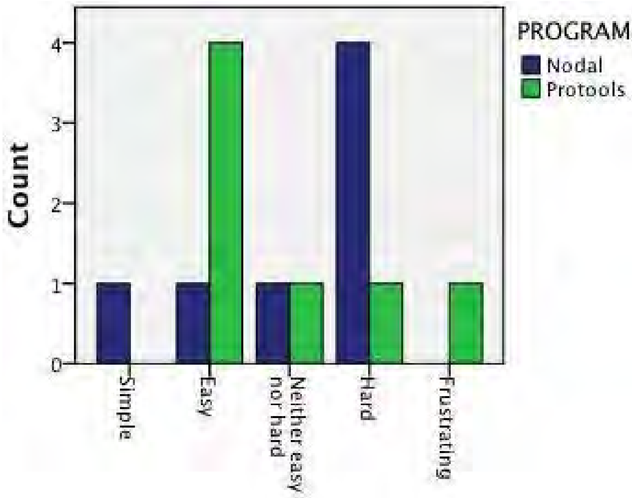

Figure 2: Reorganise sounds and modify the order of something I have made

4.2 Error Correction

The study additionally identified that regardless of aparticipants level of experience using DAWs, they found correcting mistakes in Protools easier than in Nodal.

Indirect manipulation of events within non-linear systems often adds to opacity. In Nodal it is most evident when modifying events during playback. The ability of Nodal to create complex sequences of events based on simple rules is intuitive when composing a network, but once playback has begun the sequence becomes difficult to follow, particularly when multiple events are taking place at once. Visualisation of the playhead, the change in colour of a node when triggered is the primary form of visual feed back. This is hard to follow when the system becomes sufficiently complex. Protools timeline and piano roll also contains the potential to become sufficiently complex to make the visual tracking of events difficult. But coupled with the ability to stop and start playback from a selected point within the timeline allows for quick and iterative changes to compositions.

4.3 Ease Of Use

Athanasopoulos et al. [1] study showed that individuals familiar with reading left to right and modern western musical notation were more likely to represent musical information in that way. This potentially means that familiarity of participants with the piano, and the left to right linear playhead in Protools means participants are more likely to intuitively understand Protools UI. This is reflected in the user study's result to the question "I do things easily without thinking".

71.4% of participants using Protools answered frequently compared to 28.57% when using Nodal. Users with little or no experience with DAWs will still have some level of understanding and a cultural knowledge that might aid them in interfacing with the system. Nodal's less familiar mapping requires more of the user, this is expressed by Murray et al. [11] who show that users are simultaneously learning and discovering the potential compositional results attainable with the system. While the results of this study are not statistically significant because of the small number of participants, this result also points to the potential that the piano roll is in fact an extremely intuitive way to represent the division of pitch. Sequential division of frequencies in a 2D plane could be well represented by the piano roll.

4.4 User Experience

Throughout the study participants were told they could ask questions and converse about what they are doing. During these conversations some common themes emerged. The primary theme that developed was that even though users feel less confident in their ability, feel as though they have less control and have to think harder about what they would like to do, they find composition in Nodal more rewarding. Additionally people using Nodal seem to be more absorbed in the task they are completing. This is evident in the response to the statement "Time seemed to pass quickly", 85.7% of participants using Nodal responded to the question with Always, comparative to 42.9% when using Protools.

Generative systems can often take on a game-like design. There are many examples of video games which incorporate generative systems into there design such as Iwai's Otocky[10]. User experience and usability create different criteria for user interface design. Games such as Otocky require more effort and are extremely limiting for creating compositions but can conversely result in a more enjoyable and fun experience for users [6]. Nodal's novel approach to composition is somewhat comparable, particularly in terms of users' conscious and unconscious control of events. Users retain less immediate control of events but find composition more enjoyable. This is supported by the response to the statement "using the system is extremely rewarding" where 71.4% of users responded always. Additionally, of the participants who self identified themselves as inexperienced with generative and digital composition systems but as experienced musicians, most had trouble initially understanding the concepts of a generative system and how it could be legitimately useful. But after using the system once or twice they remarked on how they could see how Nodal's approach to sequencing events was interesting, several participants expressed there excitement in the creative potential Nodal has. However the participants who did express their interest in Nodal did not necessarily create more complex or musically interesting results.

4.5 Strategies In Nodal

Participants who were filmed adopted several strategies to solve the composition tasks they were given. Strategies were consistently developed by participants as the tasks went on. When participants began the second task (improvisation to the twelve bar blues) in each system they usually took on a similar style of strategy as they did in the first task (compose a sound scape for a video-game). One result was that users who were familiar with linear DAW's seemed to understand how to create more musically complex results in both systems, even when they were less experienced musicians. The two figures (3 and 4) show the results of two participants who demonstrated this finding. Figure 3. is a system composed by a participant who self-assessed as a non-musician but who had experience using popular DAWs such as protools. The participant begun with a simple set of events and created a loop, with a melodic sequence. Then they added a second instrument and then added in the parallel nodes. The result was an exponential gain of events, this is as parallel events were triggered more events were created and then triggered again and so on. An infinite loop of events triggering at all times was the outcome. To solve this the participant begun to explore the idea of a dead end, or a drain for the excess events. It took several iterations of drains until the system was sufficiently consistent with creation of events and draining of events. The participant has effectively created a cyclic composition that creates variations in six note cycles. The network incorporates a drain which removes the additional events created by the parallel nodes and a second instrument which triggers at a separate point in the network to the first. This demonstrates an in-depth understanding of Nodal and the potential it has to great complex variation. This system illustrates how simply and efficiently Nodal can become complex. Figure 4. is a system composed by a self-assessed proficient musician. The participant has developed an understanding of rhythm and repetition in Nodal but has not developed the network beyond a linear sequence of events. This result became consistent across the participants who where inexperienced with DAWs but were experienced musicians. While the result is not statistically significant it suggests that as trained musicians have developed an advanced and habitual composition process, they may find incorporating a simple and less controlled methodology into their composition process a challenge at first. Another possible reason for this result could be that Nodal has some core concepts built into it that are commonly found in most digital composition systems. MIDI is considered a standard communication protocol across the vast majority of digital composition systems.

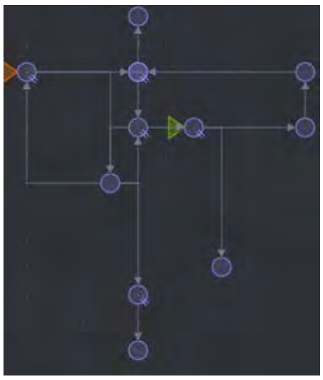

Figure 3: complex system produced by an experienced participant with little musical background

5. DISCUSSION

Nodal has shown to be effective in facilitating complex compositions with simple tools which are easy to understand and use. This has been achieved through the use of intuitive mapping of events and simple structural concepts. This study identifies the potential usefulness of non-linear and generative systems such as Nodal to facilitate a better user experience for musicians when composing. While

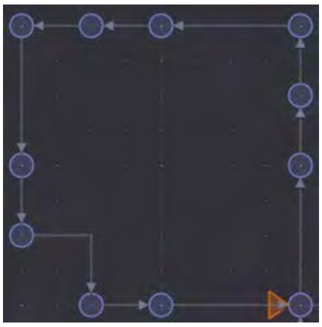

Figure 4: Simple system, produced by a user with little previous DAW experience but with significant musical experience

DAWs such as Protools are clear and precise tools for creating compositions they lack the compelling ability to organise events as non-linear networks of events. Protools piano roll and linear track view are both more immediate and intuitive but less interesting to new users and often less creative to inexperienced musicians. To create more effective tools that incorporate non-linear deign the identification of potential design problems is paramount. Participants in this study have identified several key design problems that can be broadly applied to non-linear composition systems.

- The ability to track events through the use of the representation of the playhead is key in user's ability to correct errors and make simple changes to the structure of events.
- Users potentially find non-linear composition tools more rewarding, but they are not necessarily as useful in creating linear compositions.
- The potential of non-linear systems is not immediately obvious
- Experience with generative systems does not necessarily make it easier to understand Nodal as a first time user.
- Indirect manipulation of events makes visualising the sequence of events harder once the system is sufficiently complex.

One key design factor that has become clear in this study is that the design of composition systems which are only linear or non-linear is limiting. Incorporation of both elements is what could potentially create powerful tools for flexible and complex creativity within a composition. Nodal and Protools both integrate MIDI as an input and output method for communicating with other systems, as well as using terminology and functionality internally that is specific to MIDI functionality. In Nodal you have the ability to define pitch by reference to the note and octave relative to a piano (C6,D4 etc...), this is identical in principal to the piano roll in Protools. Both have inherent design affordances, being able to map sequential events between both paradigms could allow composers to maintain flexible yet fine control over the composition. Generative systems are unique systems which when set in motion contain some degree of autonomy contributing to the completed work of art. Designing digital composition systems which allow users to create autonomously and are at the same time able to allow for fine control of events could enable composers to use generative processes without sacrificing control.

There is a tension between usability and user experience in the design of composition systems. The goal of composition systems is to both facilitate complex and creative tasks. In fulfilment of these goals a composition system should be effective and efficient at completing tasks, easy to learn and to remember. But at the same time it must also be motivating, helpful, rewarding and entertaining so as to be conducive to creative outcomes.

6. FURTHER STUDY

Study into similar composition systems will allow for the furthered verification of these design problems. Beyond that, interviewing experienced electronic artists who are using generative processes in their compositions could potentially identify additional factors which musicians experience when using these systems. Understanding the potential benefits of experimental composition systems will broaden the scope of design of tools for composers, aiding the creative process and creating better end user experience.

7. REFERENCES

[1] G. Athanasopoulos and N. Moran. Cross-cultural representations of musical shape. *Empirical Musicology Review*, 8(3):185–199, 2013.

[2] A. Blackwell and N. Collins. The Programming Language as a Musical Instrument 1 Introduction. (June):120–130, 2005.

[3] L. Church, C. Nash, and A. F. Blackwell. Liveness in Notation Use : From Music to Programming. *PPIG 2010*, 2010.

[4] Z. Eitan. Musical objects , cross-domain correspondences , and cultural choice : Commentary on " Cross-cultural representations of musical shape " by George Athanasopoulos and Nikki Moran. *Empirical Musicology Review*, 8(3):204–207, 2013.

[5] T. Green and A. Blackwell. *Cognitive Dimensions of Information Artefacts : a tutorial*, 1998.

[6] H. S. Jennifer Preece, Yvonne Rogers. *Interaction design : beyond human-computer interaction.* John Wiley & Sons, Inc, New York, NY, 2002.

[7] A. Johnston. Beyond Evaluation : Linking Practice and Theory in New Musical Interface Design. In A. R. Jensenius, A. Tveit, R. I. Godø y, and D. Overholt, editors, *Proceedings of the 2011 International Conference on New Interfaces for Musical Expression (NIME2011)*, pages 280–283, Oslo, Norway, 2011. University of Oslo and Norwegian Academy of Music.

[8] J. Mccormack, P. Mcilwain, A. Lane, and A. Dorin. Generative Composition with Nodal. *Workshop on Music and Artificial Life (part of ECAL 2007)*, 2007.

[9] P. Mcilwain, J. McCormack, A. Dorin, and A. Lane. Composing With Nodal Networks. *Proceedings of the 2006 Australasian Computer Music Conference*, 2006.

[10] K. Montminy. Otocky, 2006.

[11] T. Murray-Browne, D. Mainstone, N. Bryan-Kinns, and M. D. Plumbley. The medium is the message : Composing instruments and performing mappings. *Proceedings of the International Conference on New Interfaces for Musical Expression*, pages 56–59, 2011.

[12] C. Nash. *VSTrack: Tracking Software for VST Hosts.* Mphil thesis, Trinity College, 2004.

[13] C. Nash, A. Blackwell, W. G. Building, and J. J. T. Avenue. Tracking Virtuosity and Flow in Computer Music. *Proceedings of the International Computer Music Conference 2011, University of Huddersfield, UK*, pages 575–582, 2011.

[14] B. Shneiderman, G. Fischer, M. Czerwinski, and B. Myers. A workshop sponsored by the National Science Foundation. *A workshop sponsored by the National Science Foundation*, 1(September), 2005.

COMPOSITIONAL IMPLICATIONS OF TOUCH SCREEN TECHNOLOGY

Warren Burt
Box Hill Institute
Centre for Creative Industries

ABSTRACT

Different ways of performing live improvisatory computer music using touch screen technologies are discussed and examples of performing with these are shown. Performing with these screens give new performing and composing possibilities, and these are discussed, along with compositional details of works made with this technology.

1. XENAKIS AND GABURO

As early as 1952, in his diary, Iannis Xenakis wrote "Relearning to touch sound with our hands – that's the heart, the essence of music!"[1] and his lifelong involvement with creating music through the hand ("The hand is the organ of the body that is closest to the brain.")[2] culminated in the UPIC system, where waveforms, gestures and musical forms could be specified on a drawing tablet.

Kenneth Gaburo's work of the 1980s involved him in an intense use of physical processes and sensory deprivation experiments as a means of generating musical material for his compositions. He maintained that his body would be a better, and more intrinsic, source of quasi-random information than any external source could be. His works Antiphony IX (orchestra, children and electronics) and Re-run (electronics) were both generated with sensory-deprivation processes and these are described in his papers ISIT, and LA.[3]

2. A DILEMMA AND AN INTEREST

I've become dissatisfied with small physical controllers and netbooks – although they are very convenient, they limit the physicality of one's performance to a very small area, and they don't allow for whole body engagement in performance, some things I've been searching for for several decades now. I've also become dissatisfied (or maybe that's frustrated due to lack of access to them and time to work with them) with controllers that simply read the position of the body in space, something I had worked with extensively in the 1980s and 90s, such as the 3DIS system and the Buchla Lightning.[4] (Warren Burt "Fair Exchanges: *Writings on Dance* No. 18-19,Melbourne, Winter 1999, pp. 51-71. Warren Burt: "Computer as Part of Improvisatory Theatrical Performance. Or What I Did with Eva and Bill in July and August 2002." in *Chroma* No. 32, Brisbane, August 2002, pp. 9- 14.)

Most recently I've become intrigued with the possibilities of touch-screen technology. It's relatively cheap, easily accessible, and it seems to offer a number of possibilities for whole-body performance, as well as offering potentials for compositional resources I haven't encountered before. Whenever I encounter a new piece of equipment or software, my first question is "What are the compositional possibilities of this?" and "How can it extend what I already do or offer me new possibilities?"

For the past couple of years I've been investigating what I could do with touch-screen technology. I've done this on 3 platforms – an Apple iPad4, an ASUS Vivo Tab Windows 8 tablet/netbook, and a Samsung Galaxy Tab2 7.0 Android Tablet. Additionally, I've used a series of used iPhone and Android mobile phones as auxiliary sound-makers in performances. This paper will detail some of the pieces I've made with this technology and what some of the implications of this technology are.

---

[1] Makhi Xenakis, *Meteorites*, in "Iannis Xenakis: Architecct, Composer, Visionary", p 130, The Drawing Center, New York, 2010, issuu.com/drawingcenter/docs/drawingpapers88_xenakis (last accessed April 20, 2014)

[2] ibid, Sharon Kanach, *"Music to be seen Tracing Xenakis's Creative Process"*, p 99

[3] Kenneth Gaburo, *CONCERNING PHYSICALITY Harry Partch's BEWITCHED*;(aesthetic philosophic discourse); Percussive Arts Magazine (Research Edition), V.23, #3; March, 1985
Kenneth Gaburo, *ISIT* (a philosophical argument with reference to the concept of subject-object; also (synchronously), a text sound-performance composition); Pub. International Synergy Journal, V. 1, #2; Los Angeles; 1987 http://www.angelfire.com/mn/gaburo/isit.html
Kenneth Gaburo, *LA*: (a philosophical argument for Sound As Spirit); Pub. Perspectives of New Music; V. 25, #1; 1987

[4] Warren Burt "Fair Exchanges: *Writings on Dance* No. 18-19,Melbourne, Winter 1999, pp. 51-71.
Warren Burt: "Computer as Part of Improvisatory Theatrical Performance. Or What I Did with Eva and Bill in July and August 2002." in *Chroma* No. 32, Brisbane, August 2002, pp. 9- 14.)

3. SEVERAL KINDS OF INTERACTION

3.1. Drawing

Drawing – there are a number of programs which allow one to draw either spectra or pitch/time contours, such as HighC, Coagula, and Virtual ANS. Already, like Xenakis, I had found working with a drawing tablet instead of a mouse was very useful with these programs. Now with touch-screen technology, even just a single-touch screen, I find I can make spectrograph drawings and draw pitch/time contours with my fingers. Immediately after I started this I realized that I was doing a kind of sonic fingerpainting. I found the musical gestures I was making with my hand were very different from the kinds of gestures I had been inscribing with a mouse. They featured more elaborate shapes that curved back on themselves with more facility than any mouse movement could. Here is an example of a pitch/time score made with HighC.

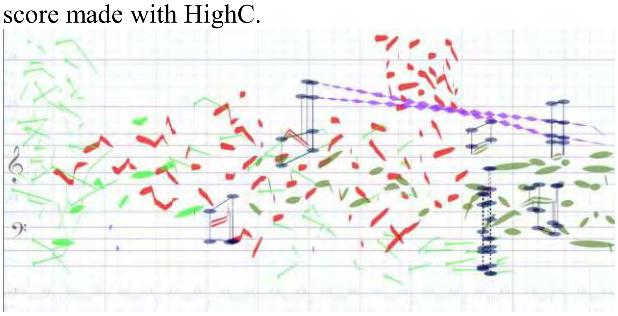

Figure 1: Warren Burt: Ides of March score made with HighC.

I made a composition with drawings like this over several days – it's called "Audio Fingerpainting." It's the first section of what I hope will be a much longer work. http://www.warrenburt.com/storage/AudioFingerpaintingMix.mp3

3.2. Multiple Tablets / Multiple Apps

For the launch of Linda Kouvaras' book "Loading the Silence: Australian Sound Art in the Post-Digital Age"[5], I was asked to give a speech / do a small performance. This performance took place in the Grainger Museum of the University of Melbourne. Surrounded by exhibits of Percy's experiments with Free Music, it occurred to me that I could do a performance with multiple tablets which would have sound sources that would reflect various aspects of Grainger's (and others') experiments. I had 5 different tablets. The Windows8 tablet had Ross Bencina's AudioMulch on it, which was running a very simple program that played pre-programmed glides of sine waves, emulating Grainger's 1950s Free Music experiments. All I had to do was touch "start" at the proper time, and the Windows8 tablet would play a short gliding tone sequence as an homage. I had two Android tablets, both of which were running the John Cage Prepared Piano App. Since Cage was working on his prepared piano work simultaneously with Grainger's Free Music experiments, I thought this would be appropriate. Further, having the app on two tablets at once meant I could have a kind of polyphony and virtuosic-sounding gesture that wouldn't have been possible with just one tablet. Finally, on the iPad and an iPhone, I was running the Thumbjam app. On the iPhone I had a sampled Grand Piano tuned to 36 tones per octave equal temperament, like Grainger's Butterfly Piano experimental instrument. On the iPad, I had Thumbjam playing a series of samples drawn from Grainger's own recordings of his Butterfly Piano experiments. Having samples of these allowed me to insert Percy's actual gestures into the mix.[6]

Having five tablets allowed me to play as if I had an extended percussion sampler available. The aspects of single-hits and brief visiting of particular sound worlds encouraged me to leap rapidly between the tablets, making a kind of live fantasy that I don't think I would have been able to make in any other way. Here's a short excerpt from a video of a performance made after the launching.

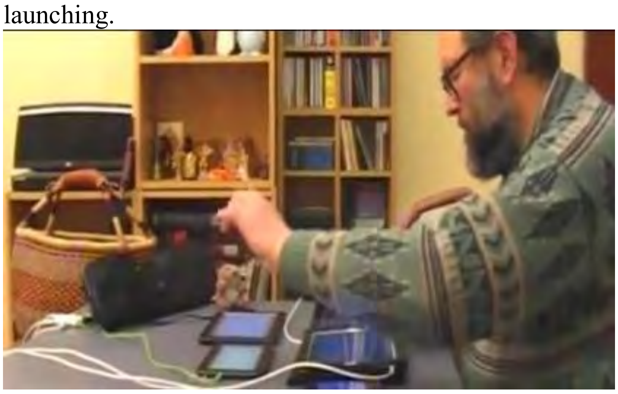

Figure 2: Warren Burt: Launching Piece, June 2013 http://www.warrenburt.com/journal/2013/6/21/launching-piece-composed-and-performed-by-warren-burt-a-new html (view 3:00-4:00)

3.3. Microtonal Keyboard Work

A number of apps have come along which allow one to set up microtonal keyboards of various kinds, some of which allow one to route MIDI from their outputs to other synthesizers. Some of these programs are SoundSquares, Musix Pro, and Lemur, which offers 2 options: one can either program one's own experimental keyboard arrays, or use their new Hexapads template. In all these cases, one can set up two dimensional arrays of either rectangular or hexagonal keys which allow one different harmonic and performance possibilities than a traditional keyboard can. For example, in "Morning on Princes Pier," I use the SoundSquares app to play a series of 7-note scales derived from Ervin Wilson's "The Scales of Mt Meru" series of papers.[7] I route the output

[5] Linda Kouvaras, *"Loading the Silence Australian Sound Art in the Post-Digital Age"* Ashgate, London, 2013

[6] "Picking Up The Threads of a History More Extensive Than Previously Known: Percy Grainger's Work with Music Technology." - in "*Proceedings of the 2004 ACMC Conference*", ACMA, Wellington.

[7] http://www.anaphoria.com/wilsonintroMERU.html (last accessed April 20, 2014)

of the iPad into an Alesis Air F/X, which allows me to move my hand above the unit, controlling the effects. My left hand performs pitches, and my right hand controls effects. By setting the two units a good distance apart, I can have wide-reaching stance, which will allow me much more freedom of movement than performing on one device would. Notice here that I've bought a stand which allows me to put the iPad on a mic stand and perform with it in a free standing manner. This allows me to have an upright posture, which I think will be a more engaging way of performing than sitting down will. That is, by standing, I can move around more, and in a more expressive manner, than I could if I were seated. This is more engaging for me, and I hope also for an audience.

Figure 3: Warren Burt: Morning on Princes Pier 2013 for iPad and Alesis AirF/X
http://www.warrenburt.com/journal/2013/5/13/concert-at-box-hill-institute-may-9-2013.html (view 7:20-8:20)

In "Lucas (2-1) C Right Drone" I assemble a custom rectangular keyboard in Lemur which allows me to play different 7 and 5 note subsets of a particular microtonal scale I derived in 2006 from Ervin Wilsons "Scales of Mt Meru" paper. Different "pages" of the Lemur keyboard allow me to freely switch between different ways of dividing up the scale. Further, there are two parallel keyboards for each harmonic division of the scale: One consists of toggle switches which turn on and off sustained pitches, while the other consists of momentary switches with only produce sound for the duration of touch, as a normal keyboard would. With these parallel keyboards I can assemble complexes of sustained tones as well as melodic filigree above them. And I can freely "modulate" between different harmonic worlds which are present in my 12-note microtonal scale. In this piece, the custom keyboard is on the iPad in the Lemur app. The MIDI from the iPad is routed into a Windows computer, which is running LinPlug's Spectral softsynth. More recently, I've also routed the MIDI from the Lemur keyboard internally on the iPad to the Sunrizer softsynth on the iPad, a way of performing this piece with only one piece of gear.

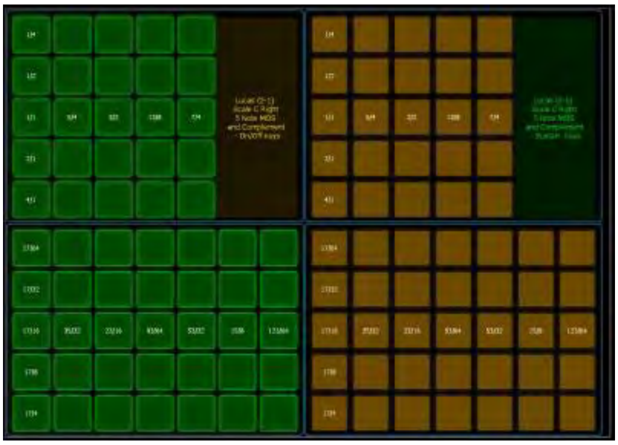

Figure 4: Lemur Keyboard layout on iPad for "Lucas (2-1) C Right Drone" Warren Burt 2014 http://youtu.be/JPi9iqz_Hvk (view 0:10-1:10)

Most recently, I've been working with Hexagonal keyboard arrangements on the iPad using both Musix Pro as well as with the Hexapad Layouts in Lemur. (I was just about to sit down for several months and program my own hexagonal keyboard program, and then not one, but two programs to do that showed up within weeks of each other. This has happened several times in the past few years. There are so many people now developing things that it's almost as if you want something, all you have to do is wait a while, and an app to do it will be developed by someone, somewhere.) Both programs allow some re-configuring of the keyboards, and I've been finding that applying these "generalized keyboard" layouts to non-equal tempered microtonal scales has been helping me to find harmonic resources in microtonal scales and making them easy to play in a way that even working with a physical hexagonal keyboard (such as the Microzone keyboard) wouldn't. The Microzone, at least the one that the University of Wollongong had, was very hard to play – the keys were physically hard to press down – after 15 minutes playing it, I had very sore wrists. The hexagonal layouts on the iPad on the other hand, allow me to glide my fingers easily over the surface, triggering off arpeggios and harmonic riffs that at the moment are a source of delight. Here is a video example of a short live piece "Tenebrae Chorale" in which the Lemur Hexagonal Keyboard (using two different keyboard arrangements) controls the Sunrizer synth playing in an 11 note microtonal MOS scale as developed by Ervin Wilson. This performance takes place using only the iPad.

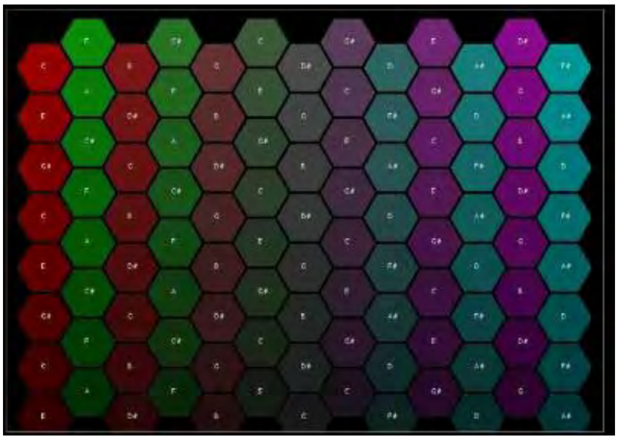

Figure 5: Hexagonal Keyboard Layout in Lemur on iPad http://youtu.be/HM4KstUcr_E (view 0:15-1:15)

3.4 **"Keyboard" layouts that allow live performance and direct control of waveforms themselves, or control of aspects of a process.**

A number of apps allow one to directly perform on a representation of the sound waveform, such as MegaCurtis and Samplr, while others such as TopLap and abcdefgh...allow one control of only semi-predictable processes. For example, Samplr allows you to use up to 6 different samples, each of which can have their waveforms manipulated in real-time with a variety of slicing, looping, and scrubbing techniques. In this piece, "A Degenerate Practice", a quick video demonstration reveals some of the software instrument's capabilities. Six loops are used – two are electronic soundscapes I fingerpainted in Henry Lowengard's iOS based "Tondo" app, another is a loop of music by the late Brazilian accordionist and composer Sivuca. Other sources include a loop from Robert Erickson's "East of the Beach," the opening chords of Beethoven's "Waldstein Sonata," and a quote from American composer Harry Partch.

Figure 6: iPad screen layout for "A Degenerate Practice" http://youtu.be/OtiQQ3jZCVw (view 1:00-2:15)

A number of apps use the accelerometer in the iPad to give a sense of "gravity" to the events on the screen. One of the most clever of these is Jorg Piringer's abcdefghij.....in which samples of German-language phonemes bounce around the screen, being played when they hit a boundary of the screen. The movement of the letter-icons changes as the iPad is tilted. In the performance of my "Lettrist Fantasy" I pick alphabet beads at random from a small box. These provide the phonemes I load into Piringer's program, and the tilting of the iPad changes the patterns of the phonemes. I use the graphics of the program as a score to determine which phonemes I will sing/speak and how I will do that. The use of small belt-mounted loudspeakers in this piece gives me total freedom of movement, so that I'm not even constrained by a connection to a sound system. Also, the use of belt-mounted loudspeakers means that sound from both the electronics and my voice are coming from the same (movable) point in space.

Figure 7: Warren Burt performing Lettrist Fantasy, May 2013http://www.warrenburt.com/journal/2013/5/13/concert-at-box-hill-institute-may-9-2013 html (view 2:30-3:30)

3.5 Multiple Controls

These can be either traditional – like the sliders of a mixer – or they can be more radical – such as setting up a custom performance layout – or they can be a set of controls that simply allow one to navigate through the possibilities of an algorithmic process. For example, I set up a Lemur patch on the iPad to control a patch on a Windows8 computer, using John Dunn's ArtWonk and the Modartt Pianoteq physically modeled virtual piano, which made a canon from the output of one of Julien Sprott's chaotic equations.[8] The Lemur keyboard allows one to turn the patch on and off, select one of 8 tempi, work with 9 different transpositions of the 3 voices of the canon, set 4 different melodic widths for the Sprott equation generated melodies, select 8 different "modes" which use different pitches of the microtonal scale used in the piece, select 3 different

[8] Julien Sprott "Simple Chaotic Systems and Circuits." (Am. J. Phys. 68(8), August 2000, pp. 758-63

ranges for the volume of the initial melody (which is canonically imitated by the other two voices) and have 3 different settings of the pedals of the Pianoteq. With this amount of control, I can move through the space of compositional possibilities of the patch as rapidly or slowly as I wish. That's a lot of different kinds of structures which I can explore with this setup.

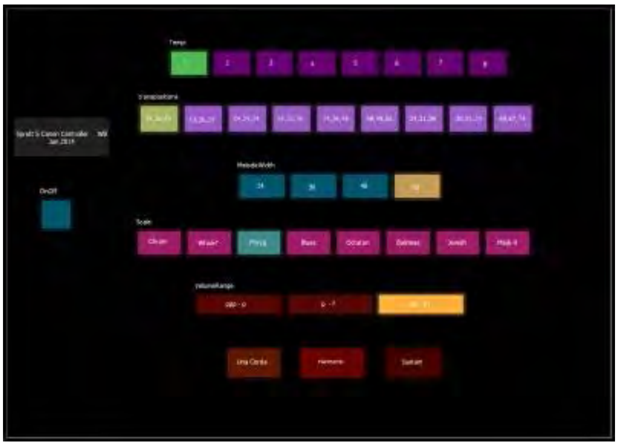

Figure 8: Lemur Keyboard Layout on iPad for "Sprott Canon Controller" 2014 Warren Burt
http://youtu.be/lyTf0J35S2o (view 0:15-1:15)

Another kind of setup is a more direct performance setup – "Four Parisian Mods". In this one, the iPad Lemur patch provides 4 x-y controllers, each of which controls the parameters in 4 different GRM Tools plugins. The MIDI from the iPad is sent to a Windows8 PC, on which the GRM tools are hosted by AudioMulch. Auxiliary sliders in the Lemur patch also control if an effect is heard or not. With a mouse, one can control just one x-y window on one GRM Tool at a time, but with this setup, one can easily control 4 different effects simultaneously. Also, the attempt to meaningfully control 4 different x-y controllers at once on a touch screen leads to some fairly comical physical results.

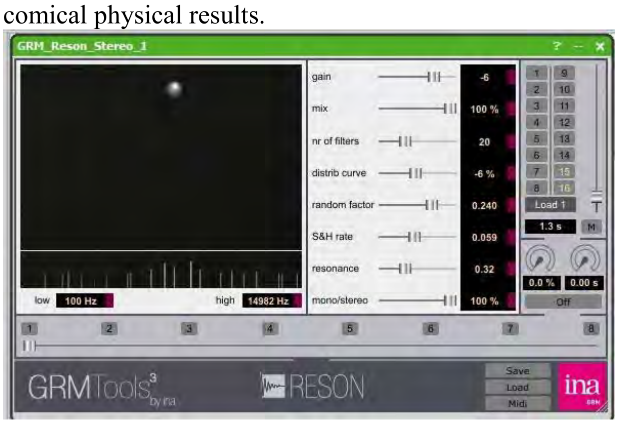

Figure 9: GRM Tools Reson Stereo Filter. The area with the ball in the upper left is the area where parameters of the plugin can be controlled with a mouse or an external controller, such as an iPad based Lemur patch.

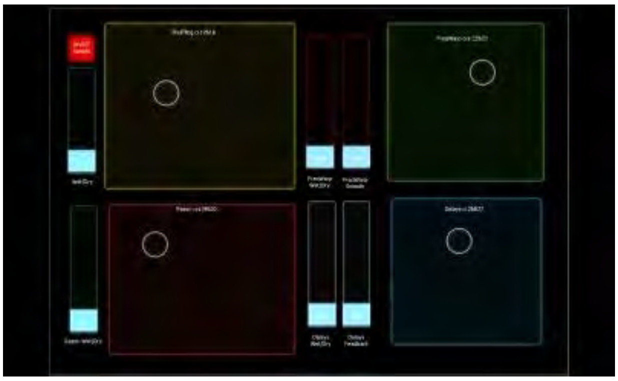

Figure 10: Lemur Control Surface on iPad for "4 Parisian Mods" Warren Burt 2014
http://youtu.be/c93pQzND5dA (view 1:00-2:00)

4. CONCLUSION

One of the essential things about designing a performing interface for a piece is that when you are doing that, you are designing the visual environment that you will be living in while performing. The importance of this cannot be overstressed. Given a visual setup, this will influence the way you perform the piece, and the more deeply you sink into the visual world of the interface, the more deeply you will become enmeshed in the compositional possibilities inherent in your musical material. Further, designing a visual interface, with musical performance possibilities in mind, will create a richer conceptual environment in which to move.

I think it's more interesting to watch someone perform with their full body rather than watching them sitting behind a laptop (even if that only involves someone standing behind a non-obscuring device instead of sitting). (My wife calls this the "Reading One's Email in Public" brand of theatre.) Further, the use of touch-screen technology, either with existing apps or custom made, allows one a greater range of gestures and compositional possibilities than other, more traditional modes of performing. Despite the unfortunate implications of the term, we actually DO want this kind of music making to have some aspects of "the spectacle"[9] about it, so that our performing WILL be an occasion "to be present to."[10] In the future, larger performance platforms might make the performance more physically engaged. For example, I think a 1 meter by 1 meter touch screen could allow for some wonderfully energetic and extravagant full body gestures. But for the moment, concentrating on a single screen, or several screens, will allow performance which both directly shapes individual gestures and also controls aspects of an algorithmic process, and does it quite elegantly.

---

[9] Guy Debord: *Society of the Spectacle* http://www.antiworld.se/project/references/texts/The_Society%20_Of%20_The%20_Spectacle.pdf

[10] Benjamin Boretz: If I Am a Musical Thinker, *Perspectives of New Music*, Vol 20, No. 1 & 2 (Autumn 1981 – Summer 1982) 464-517.

5. REFERENCES

Boretz, Benjamin 1981: “If I Am a Musical Thinker” *Perspectives of New Music*, 20(1/2): 464-517.

Burt, Warren 1999: "Fair Exchanges” *Writings on Dance* 18-19: 51-71.

Burt, Warren 2002 "Computer as Part of Improvisatory Theatrical Performance. Or What I Did with Eva and Bill in July and August 2002." *Chroma* 32: 9-14

Burt, Warren “Picking Up The Threads of a History More Extensive Than Previously Known: Percy Grainger’s Work with Music Technology.” *Proceedings of the 2004 ACMC Conference*”, ACMA, Wellington.

Debord, Guy: *Society of the Spectacle* http://www.antiworld.se/project/references/texts/The_Society%20_Of%20_The%20_Spectacle.pdf (last accessed April 20, 2014

Gaburo, Kenneth 1985 “CONCERNING PHYSICALITY: Harry Partch's BEWITCHED;(aesthetic philosophic discourse)”; *Percussive Arts Magazine (Research Edition)* 23(3)

Gaburo, Kenneth 1987 “ISIT (a philosophical argument with reference to the concept of subject-object;also (synchronously), a text sound-performance composition)” *International Synergy Journal*, 1(2) http://www.angelfire.com/mn/gaburo/isit html (last accessed April 20, 2014)

Gaburo, Kenneth 1987 “LA: (a philosophical argument for Sound As Spirit)” *Perspectives of New Music* 25(1)

Kanach, Sharon 2010 “Music to be seen: Tracing Xenakis’s Creative Process” in *Iannis Xenakis: Architecct, Composer, Visionary*”, New York: The Drawing Center http://issuu.com/drawingcenter/docs/drawingpapers88_xenakis (last accessed April 20, 2014)

Kouvaras, Linda 2013 *Loading the Silence: Australian Sound Art in the Post-Digital Age*” London, Ashgate

Sprott, Julien 2000 "Simple Chaotic Systems and Circuits." *Am. J. Phys*. 68(8): 758-63

Wilson, Ervin 1993 “The Scales of Mt Meru” http://www.anaphoria.com/wilsonintroMERU html (last accessed April 20, 2014)

Xenakis, Makhi, “Meteorites”, in *Iannis Xenakis: Architecct, Composer, Visionary*”, p 130, The Drawing Center, New York, New York: The Drawing Center http://issuu.com/drawingcenter/docs/drawingpapers88_xenakis (last accessed April 20, 2014)

SUBLIME: AN OPEN APPROACH TO COLLABORATIVE PERFORMANCE

Travis Cox

Victorian College of the Arts and Melbourne Conservatorium of Music, Performing Arts

Roger Alsop

Victorian College of the Arts and Melbourne Conservatorium of Music. Performing Arts

ABSTRACT

Over two weeks in March 2014, computer based sound/visual and dance creative practitioners developed new works in the unstructured workshop environment titled *Sublime* in the Newport Substation, a collaborative event auspiced by Ausdance and the Victorian College of the Arts. This presentation maps the development of artworks across that fortnight. The time can roughly be split into: the first week, in which a shared language was developed between the different groups of practitioners; and the second week in which these commonalities were explored and developed in the production of new integrated performance pieces.

This paper discusses the collaborative development of three works that involved each of the above areas in their development and implementation. It outlines the structure of the workshop, the different approaches and processes if the various participants, and the process for discovering commonalities in practice, approach, and language in order to harmonize the various outcomes.

1. INTRODUCTION

Sublime came about through an invitation from Andy Howitt, Ausdance director, to Roger Alsop to develop a collaborative situation in which sound, visual, and dance artists could develop new works in a fully collaborative and improvised way. Jeremy Gaden, director of the Newport Substation, provided the location, allowing access to unused spaces: a large and smaller hall, shown in **Figure 1** and **Figure 2**, and a balcony shown in **Figure 3** below.

The intention was to create an outcome in which all aspects were originated in the workshops. Alsop and Howitt ensured that there was as little preplanning or pre-intentionality as possible, believing that to develop any outcomes from as open a philosophical and practical position as possible would allow all involved to have equal agency, opportunity, and input.

With no ‘top-down’ methodology this approach rendered problems in the first few days as participants anticipated or expected a brief, or guiding hand. However this process proved very effective in creating harmonious outcomes in all respects and across the three art forms, including a strong sense of personal and group agency; interaction in which all parts and participants were considered essential in the development of each works and to any potential outcomes; and a sense that all of the final works were intertwined. This intertwining was facilitated by the sense that all works were being developed concurrently from a neutral base, and that all participants could contribute in areas they wanted to, regardless of previous activities and experience, or relation to a specific art form.

Over the two weeks of development Howitt took an active role as a dancer/choreographer, and Alsop took a more curatorial/ facilitator role. This resulted in Howitt having an involved position, internal to his work, and Alsop becoming a less involved ‘outside eye’.

A goal in the creation of *Sublime* was that the focus be on developing dialogs and interactions between the creators working in areas that they felt most comfortable in. Discussion of various technologies and systems, while not discouraged, was not prevalent while ‘on the floor’. In this respect it was good to see that the conversations mostly transcended the process, instead focusing on the outcomes and keeping these aspects as opaque as possible.

2. PROCESSES FOR DEVELOPING A HARMONIC RELATIONSHIP BETWEEN ART FORMS

2.1. Creating and experiencing harmony

The simultaneous combination of narratives to form a consistent whole[1] of is an essence in the arts. In the performed arts these combinations are represented through sound and vision, such as hearing and seeing actors, and seeing dancers and hearing music. Many ways of creating a seamless harmony between these elements have been used, ranging from the last minute coalescing of choreography, décor and music of Cunningham, Rauschenberg and Cage [1], the multi-faceted approach of Robert Wilson, who works in all roles of performance creation [2], to the continually varying but fully collaborative/interactive approaches of Robert Lepage [3].

Discussing the interrelationships between sound and image in film, Deutsch (4) saw “the sound that accompanies visuals not merely as a combination of disparate disciplines, but as a unified and coherent entity [, where] music, dialogue, effects and atmospheres are heard as interdependent layers.”

Film is an art form in which the interaction between sound (sound effects and music), narrative (the story),

[1] To paraphrase the Apple dictionary definition of ‘harmony’

and visual image (actions, props, locations, costumes, etc.) is required to make a successful whole. In the better iterations of this art form the result is more than the sum of these parts, as is the case for all art forms in which intertwining elements are presented.

2.2. Space and Harmony

Harmony is dependent on empty spaces. This emptiness allows the various resonances of different voices to intertwine and coalesce. These spaces may be analogous to Burrows (5) three fields, a sensorial, observed physical space, mental space of potential options, or diffused spiritual space, which, through an input, coalesce and are subsumed to provide a fulfilled and singular whole: a harmonic outcome. Bachelard (6) saw the varieties of experience within domestic space; the dark, irrational harmonies of the cellar and lighter, intimate, harmonies of walls, furniture, or a chest of drawers. These framings of the term provide a way conceiving harmony as a meta term, in which harmony is a sense derived from the confluences of materials and transitory input or influences.

2.3. Notation and Harmony

Brooks (7) discussed Laban's concept of 'space harmony', drawing together his senses of architecture, movement and music to consider relationships between vertical, horizontal and sagittal spaces in a philosophy that considered the relations of these aspects of the body to each other and the space(s) encompassing them. Laban's notation process was based on the 'quadricosahedron', where three dimensional space is into 27 directions [8], demonstrating that movement into and in relation to space was conceptually paramount.

Musical notation has been used to demonstrate and develop approaches to creating 'harmonic' outcomes. Here the residual resonances of simultaneous sonic events are linked to create a sense of harmony in the perceiver's mind. In early western music these resonances were influenced by the space in which the sounds were propagated, and this space influenced the choices of sounds created.

2.4. The human body and harmony

Bachelard and Burrows offered approaches to considering harmony beyond the sonic or arts based understanding of the word, giving it social, locational, physical, psychological and spiritual dimensions. These are all experienced through and within the human body, and Laban indicated a formalized way to understand and generate strands that can entwine to form harmony possible between the body and its surrounding space.

In the pieces developed in Sublime a process of understanding the unique relationships between the dancer as physical object, the environment and props as physical objects, and sound and projected image as on-physical objects was developed. A conceptual harmony between concept and object was developed and this in turn created an observed harmony between concept and object and finally an intended harmony between concept and object.

3. TIMELINE FOR THE DEVELOPMENT OF SUBLIME

As stated, the Sublime workshop occurred over the course of two weeks at the Substation gallery in Newport and the timeline for the development of works can roughly be split into two parts. The first week, in which participants began to get to know one another and their practices, and the second week in which the established dialogues from the first week were exploited to develop the artworks already in progress. This section offers a break-down of the activities of each discipline, video, sound and dance within each of the two weeks before discussing the (continual) outcomes of Sublime.

3.1. Video

Within the first two days of *Sublime* elements of possible projection were discussed amongst visual artists with an emphasis on silhouette and shadow. These elements were selected for their ability to be either subtle or bold and sit within a variety of environments. This was seen as a way of not 'getting in the way of the dancers' nor 'imposing an idea' but instead being something that could be cajoled to work with the choreography devised by the dancers.

Over the course of the week, through the drawing from code devised for previous projects, a projection element consisting of an array of circular shadows was devised. Mid-way through the week this was also given a rudimentary beat detection algorithm allowing the sound being developed within the space to affect the size and movement of the shadows.

By the second week the projection had gained a discreet options menu that allowed for size, motion, opacity and other properties of the shadows to be controlled in real-time. This, along with an improved beat detection and access to discreet audio rather than relying on ambient sound from microphones improved the quality and responsiveness of the projection to the sound and dance within the space. Further refinements to the program continued for the rest of the week including adding additional real-time options to modify the effects at specific points of the performances.

There was no mapping of the images to the shape of the inflatable being projected on; this decision was both pragmatic and aesthetic. Firstly, the shape of the inflatable was impossible to predict until it was completely inflated, and then the orientation of it in relation to the positions of projectors altered each time it was inflated.

Secondly, the group thought that by allowing the projected images to be influenced by the inflatable being projected on they would reflect a greater relationship to it. The shape of the images morphed between circular blobs and straight lines, and the varying shape and position of the inflatable made the projected images take on new and unique shapes as it was being inflated and interacted with by the dancers.

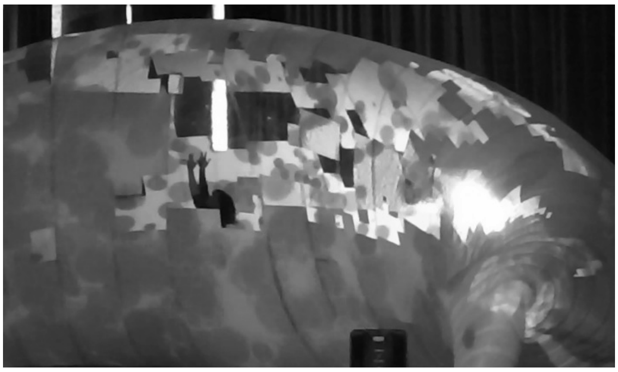

Figure 1. Dancer inside the Inflatable

Figure 2. Inflatable with two dancers.

3.2. Sound

All sound artists involved within the Sublime project drew their inspiration, first, from the environment itself. This would eventuate in a long-form composition by one of the sound artists to accompany the "SpaceWorm" work which was inspired by the sounds of air conditioners and the fan which pumped-up the inflatable during the performance.

The second sound artist working within the project continued to respond in real time through an electric guitar augmented through a layering of effects. This element of sound production was used within the work in the upstairs corridor within the gallery space. This impromptu sound-creation element was well suited for the workshop environment due to its on-the-fly ability to change with the other elements of dance and projection as they developed.

3.3. Dance

Over the course of the first week two separate dance performances developed. The first involved two dancers within the main space this piece being devised around a large inflatable: a second piece was developed by a solo dancer within a corridor of the gallery. For the first week the dancers in both spaces spent their time becoming accustomed to the space and in the case of the duo working with the inflatable, developing a rapport with the way in which it moved and inflated.

By mid-second week the dancers had well-developed performances within their respective spaces and concentration turned to refining their movements.

3.4. Harmony

While it is useful to outline the discreet elements involved in Sublime (video, sound, dance), the harmony within these performances was only achieved through conversations between the creative practitioners.

The development of the projection over the course of the two weeks was defined by a reflexive engagement with the developing performance and sound elements both through active listening and watching by the visual coder as well as conversations with other parties regarding their personal aims for their elements.

Late in the first week, a conversation between the projection artist and the solo dancer in the upstairs corridor (shown in **Figure 3**) lead to a discussion of the story that the dancer had developed for her performance. Part of this story involved the dancer exploring the limitations of her brain's control over her body by simultaneously producing rapid movement with her arm and torso while aiming to move with careful and controlled motion from her hips and legs. This 'split' became a useful visual metaphor and new code was developed to allow for a splitting of the visuals into a fast/slow halves in order to complement these movements.

Similarly, after development of easier real-time modification of the visuals in the second week, a conversation took place between the visual controller and the dancers in the main auditorium space. This conversation allowed the dancers to speak of their narrative intentions for their performance and seek more information on the limitations of the projection element. From this conversation the dancer's structure was solidified to the visual artist who could then respond more readily to these changes through modification of the projection during the performance.

3.5. The 'third work'

If in the second week a harmony was developing between the practitioners, by the Thursday of the second week it had become a song. By this time the two works were well developed and new opportunities could be explored. As the separate practitioners had come to know each other over the course of the two weeks conversations had turned from immediate concerns of the works to instead commonalities within wider theoretical and conceptual concerns.

It was from one of these conversations during downtime that one of the sound artists and one of the dancers began to discuss a mutual interest in social theatre. From this discussion these practitioners, along with a third artist who supplied motion detection technology, developed a new sound/dance performance in the last two days of the workshop.

3.6. Temporal Integration

In their paper, "Temporal Integration of Interactive Technology in Dance: Creative Process Impacts", Latulipe, Wilson (9) discuss the ability of technology to impact on choreographic development within dance productions. Referencing the aspects of time that are involved in the development/choreographing of new dance works they state, "If props are available for use at the beginning of the process, they are more likely to be well-integrated into the choreography." The ability for these 'props', which encompass any element that integrates with performance such as costume, staging or technologies, to affect the choreography is labelled as "temporal integration effects". Through mapping of the integration of elements of the dance production Latilupe et al shows the challenges that can be produced when new elements are added part-way through the creative process. But this can also lead to new possibilities and responsiveness within the performance and therefore should not be maligned.

The performances Latulipe et al developed and interrogate within their article are structured works in which a top-down directive was in place. This can be understood from the references to 'teaching' the choreography to the dancers. However, within the Sublime project there were no such leaders or directors, rather decisions were made through conversation and negotiation, which naturally leads to a less rigid environment of creation.

Within the Sublime project the absence of defined specific outcome or public performance, meant that temporal integration effects were less problematic and additions of new ideas, technology and props were embraced within the workshop environment.

An example of this is the corridor work that added new elements during the course of the workshop. As previously mentioned this work was developed without projection or sound for half of the first week before a projection element was added. Towards the end of the second week sound artist Camille Robinson began to produce sound which responded in real-time to the movement and context.

It is the addition mid-second week, however, of motion tracking technology, which relates most clearly to the temporal integration effect. Mark Pedersen, another of the collaborating artists designed and built a Bluetooth motion-tracking glove using Arduimo hardware, allowing information to be streamed in real time regarding the pitch, roll and yaw of the performer's arm. The artist producing the visuals had not worked with this technology previously and thus a learning curve for integration of the information streamed from the device into the visual projection was limited to just a few days.

While not successful in developing a complex responsiveness based on the movement information a basic modification to the colour of the projection based on the speed of the dancer's arm was introduced to the work which helped to contrast the calm beginnings of the work to the faster, jerkier movements of the climax of the performance. This addition would not have been possible within a more structured environment in which specific outcomes were expected.

The third work, developed over the course of two days at the end of the workshop however, could be situated more readily within Latilupe et al's creative process. The technology and sound were integrated from the outset of the work and developed along with the movement of the dancer. However, this work only came about due to the previous non-structured environment in which the first two works were created; an environment in which each practitioner was free to input new ideas and show 'what they could do' to the larger group, rather than an environment in which specific structures defined the input of individuals.

4. FIRST-HAND EXPERIENCE

In this section the authors consider their personal responses to the development of the works discussed. Their roles evolved to have Cox taking that of a creative agent responding to the developments of the locations and the dancers and their actions, and Alsop having a more facilitator and 'out side eye' role, discussing the experience of the works from an audience point of view.

Cox engaged with the artworks as the producer of the projection element within both the inflatable and corridor works. The projection itself was produced in real-time within the Processing environment with variables modified in response to the changes in the dancer's movements. This code base was shared across both projects, though with different variable settings in order to create different effects.

As stated, the aims when developing the projection were to work in harmony with other elements of Sublime, especially the other visual element, the dancers. Cox produced a simple 'shadow' generator of dark ellipses drifting vertically across a white screen. These spheres played well across surfaces when projected and allowed for an emphasis of the forms that they encountered (including the dancer's bodies) rather then a neutralisation of their ground. As this was produced within the first few days of the workshop it was unknown how this projection would be used as the performances developed.

As the dancers within the main space began to move and perform with the inflatable over the first week Cox began to darken the room and project over their movements, aiming to find the best location and distance to project from. This was also when he began to converse with the dancer performing upstairs regarding the aims for her own performance. From the conversations he began to use the same projection upstairs with her, modifying the size and movement of ellipses to facilitate the more intimate nature of her piece.

While simplistic in appearance, the variability of the visuals generated through real-time modification of variables allowed for a wide variety of effects to be achieved. The larger performance using the inflatable can be characterised as large, uncontrolled and imposing, while the upstairs performance was more intimate and considered. The ease with which I could modify the projection allowed for Cox to align with both the downstairs performance by producing large, brash shapes and plunging the entire space into light or shadow, and also with the upstairs work with it's smaller shapes projected onto the dancer's body creating an intimacy and a highlighting of the dancer's body as she moved.

Alsop viewed each work in various stages of their development. The Balcony work had become a well-developed work with a clear trajectory, intent and outcome, ready for performance in its space, it showed and well considered process, one in which all participants were able to understand the thrust of the work and respond to it with little need for on-going discussion between the participants. It provided an intimate experience for the viewer, focussing attention on the minutia of the dancers movement and the subtle projections provided by Cox. Highly processes guitar music was improvised by Camille Robinson, and this provided a lush, but subtle, backdrop to the visual experience.

Figure 3. Dancer in the corridor space

The Inflatable was a more open piece. While there were intentions to develop a more complete work, the process, and added prop, created a sense of many fertile options. It was decided to allow this, with the sense of developing a series of works that may go to different spaces and contexts.

Viewing the work as it was shown at the end of the workshop, it was clear that this was the best option; but to consider it as a performance the adjective that best describe it are loud, hard, uncontrolled, and intimidating; but enveloping, intriguing and subsuming would be as appropriate. This diversity of response led the group to consider how each of these intellectual and emotional spaces might best be developed in performance and in interaction with the audience and performance space.

5. CONCLUSION

In an environment without directorial control or expected outcomes it can be difficult to recognize whether a workshop can be marked as successful or a failure. If we consider the development of performances and introduction and connection between creative peers Sublime can be marked as a success. Of course this is not to suggest that throwing a handful of creative people into a room will always produce such results. There must be recognition of the structure and support available to the practitioners including, a large amount of space and freedom, access to varied and plentiful technology on-site and the non-structured schedule allowing artists to work within the space at times conducive and convenient for themselves.

The disparate elements of the Sublime project developed over the course of two weeks into a cohesive whole. Sublime shows that a model of non-top-down authority when combined with confident practitioners can develop new relationships and performances that can continue to be leveraged in the future. To this end the individuals involved within Sublime continue to be in contact and are currently exploring opportunities to perform the outcomes to a wider audience as well as produce new works.

6. REFERENCES

1. White B. "As if they didn't hear the music," Or: How I Learned to Stop Worrying and Love Mickey Mouse. The Opera Quarterly. 2006;22(1):65-89.
2. Badham V. Robert Wilson: 'There are schools teaching stage decoration. Burn those schools!' [On Line interview]. Australia: the guardian; 2014 [cited 2014 April 23]. Available from: http://www.theguardian.com/culture/australia-culture-blog/2014/feb/22/robert-wilson-perth-festival-krafts-last-tape.
3. Knapton B. Activating simultaneity in performance: exploring Robert Lepage's working principles in the making of Gaijin 2008.
4. Deutsch S. Editorial. The Soundtrack 2007;1(1):3-13.
5. Burrows DL. Sound, speech, and music. Amherst: University of Massachusetts Press; 1990. 118 p.
6. Bachelard G. The poetics of space: Beacon Press; 1994.
7. Brooks LM. Harmony in space: a perspective on the work of Rudolf Laban. Journal of aesthetic education. 1993:29-41.
8. Herbison-Evans D. Symmetry and Dance Sydney: Basser Department of Computer Science; 2003 [updated 18 October 2003; cited 2007 July 1]. Available from: http://www-staff.it.uts.edu.au/~don/pubs/symmetry.html.
9. Latulipe C, Wilson D, Huskey S, Gonzalez B, Word M, editors. Temporal integration of interactive technology in dance: creative process impacts. Proceedings of the 8th ACM conference on Creativity and cognition; 2011: ACM.

MUSICAL COMPOSITION WITH NAKED EYE AND BINOCULAR ASTRONOMY

Angelo Fraietta
Smart Controller Pty Ltd

ABSTRACT

Multimedia compositions that are inspired by outer space usually display and manipulate photographs from telescopes or utilise fractal images or animations. The performer or listener in these cases does not look into space with their own eyes. This paper describes the progress of a musical interface and compositional paradigms that correlate musical performance with naked eye and binocular astronomy, allowing the performer or listener to manipulate the music based on what is being observed in the night sky in real-time.

1. INTRODUCTION

The night sky has been observed by peoples from all continents for millennia. Naked eye astronomy has been used by civilizations to track seasons and to determine the best time to plant and harvest crops. Many of these civilizations; including the Chinese, Greeks, Egyptians and Sumerians; charted roughly familiar patterns of the stars and named them after mythical gods, legendary characters and imaginary creatures. They did this by using a *join-the-dots* methodology, where the pattern was made by creating imaginary lines between the stars or points of light (Massey and Quirk 2012). Other civilizations, such as the Indigenous peoples of Australia, made patterns from the dark areas of the sky (Norris and Norris 2009). Moreover, the stars were used for navigation by ancient mariners, enabling them to accurately traverse the globe (Chittick 2006).

1.1. Visual limitations to viewers

In our modern industrial world, naked eye astronomy has been severely affected by city lights. Many people in urban areas cannot see the Milky Way at night due to the amount of background light from housing, advertising, security and streetlights (Norris and Norris 2009). Only the brightest stars can be seen at night, and the awe of the night sky is now all but lost by many who live in the city (Bee 2006).

Since the advances in astro-photography and the availability of images from the Hubble Telescope in text books, magazines and the internet, many people tend to view stars vicariously. I believe that our technological advancements have in many ways destroyed the personal relationship between the individual and the night sky. Although there are great software programs available for mobile phones, such as *Google Sky,* that enable users to identify stars and constellations in real-time; this research aims to use the information for an artistic outcome rather than just an informative one.

1.2. Existing musical works

Musical composition inspired by or based on outer space is not a new concept. In many cases, the musical composition is based on the superheterodyne manipulation of signals received from radio telescopes (Harger 2011), music performances with slides of deep space astro-images displayed to the audience (Malin and Wesley-Smith 1989), and fractal generated animations that may or may not have any direct correlation to the work itself (Bright 2012). In these cases, the audience is dictated to by the performers as to what they will see. Ironically, the audience is able to hear the music directly as a live performance, but denied the experience of a real-time visual performance of the night sky.

This paper details a compositional paradigm that attempts to spark the interest and stir the imagination of the audience in the night sky by engaging them in a sound installation or concert, where there is a direct correlation between the music and their view of the night sky. This is accomplished through the use of naked eye and binocular astronomy coupled with algorithmic and improvisational compositional techniques.

2. ARTISTIC PERFORMANCE FACTORS

The factors affecting the artistic performance that are addressed in this paper include: light pollution; determining the viewer's focal point in space at a particular moment in time; and using this information compositionally, either as input to an automaton for sound generation, or as stimulus to live musicians.

2.1. Light pollution

Light pollution is essentially background light that limits or obscures what the observer can see (Norris and Norris 2009). The effect of light pollution is different between binocular and naked eye astronomy, and has both positive and negative effects. Some negative effects can be overcome, while positive effects can be utilised.

2.1.1. Binocular astronomy

The effects of light pollution can be reduced by magnifying the field of view through the use of a telescope or binoculars. This enables the viewer to see stars that are not visible to the naked eye in those conditions (Wallace, Dawes, and Northfield 2012). Although one might consider using a telescope as the interface for the viewer, finding one's way through the night sky for a novice is far easier using binoculars than a telescope, which is "a bit like looking at the sky through a straw" (Wallace, Dawes, and Northfield 2012, p. 10). Binoculars enable the viewer to appreciate open clusters, seeing each star in the cluster in context with the others. A telescope, on the other hand, may provide too much magnification (Bee 2006). The difference between using binoculars and a telescope in the night sky is akin to the difference between observing an insect on a flower with a magnifying glass and observing it with a microscope. The faintest stars visible with the naked eye under dark skies are at magnitude 6; however, a pair of 10x50 binoculars increases this limit to magnitude 9 or 10 (Bee 2006).[1]

2.1.2. Naked eye astronomy

Aside from all the negative effects of light pollution, one positive aspect is that it is easier for a novice to recognise patterns that make up the constellations. This is because there are fewer faint stars in light polluted skies to confuse the viewer (Wallace, Dawes, and Northfield 2012). The area of sky being referenced can be highlighted to the audience through the use of a green laser pointer, which will unambiguously mark the celestial object or point in space. Mounting a green laser to binoculars has two advantages. The first is that the audience can be directed to a point in sky while they are waiting for their turn on the binoculars. The second is that when the audience leaves the performance, they can look up into the night sky and know basically where they were looking, enhancing their memory of the performance (Smith Brindle 1986).

2.2. Mapping the positions of stars in time

Although objects in the night sky move or change position relative to the local horizon throughout the night, it is actually the earth that rotates upon its axis that gives this appearance. For example, a star that appears immediately overhead in Sydney at midnight will not appear overhead in Perth until it is midnight in Perth. Furthermore, due to the Earth's rotation around the sun, the stars each night appear at the positions they did the night before approximately four minutes earlier. The east-west position of a star on the *celestial globe* is called its *right ascension* (RA) and is based on its azimuth at Greenwich Meantime on the vernal equinox. The RA is the *x* axis on star charts and is measured in hours, minutes and seconds (Massey and Quirk 2012). Similarly, the height that a star appears above the horizon is different depending upon the latitude of the viewer. For example, a star that appears immediately overhead in Sydney would appear in the south for a viewer in Cairns. The north-south position of a star on the celestial globe is called its *declination* (Dec.) and is based on its altitude at Greenwich Meantime on the vernal equinox. The Dec. is the *y* axis on star charts and is measured in degrees, minutes and seconds (Massey and Quirk 2012).

It is therefore possible to identify a particular star or point in the in the sky by determining its RA and Dec. and comparing it against entries in a star database. The RA and Dec. can be calculated given the date, time, azimuth and the angle the star appears above the horizon (Duffett-Smith and Zwart 2011).

Many telescopes have a feature called *go-to* that enables users to select a star from an inbuilt database, causing the telescope to automatically slew to the requested star or point in the sky. This is accomplished through the use of gears and rotary encoders (Seronik 2013). A simpler and cost effective alternative, albeit with a lower resolution, is to read the angle and azimuth of the binoculars using triple axis accelerometers and magnetometers through an Arduino microcontroller (Margolis 2011). The azimuth and altitude data is then passed through a Bluetooth serial port to a computer or smart phone to provide further computation.

The output of the sensor unit is converted from horizon to equatorial coordinates by using the date, time and geographic location stored on the computer or smart phone. The magnetic offset between the magnetic and celestial north poles is accomplished by aligning against a bright reference star (Seronik 2013). This resultant RA and Dec. point can now be used as input to the algorithmic music software.

3. COMPOSING WITH RA AND DEC.

Smith Brindle states: "In music, chaos is infinitely tiresome" (Smith Brindle 1986, p. 146). When one considers that people are often in bed sleeping at night when the night sky is at its best, "tiresome" performances should be avoided by composers and performers. Many composers today use algorithmic composition techniques that enable them to map certain physical and conceptual gestures to musical parameters (Winkler 1995). This is possible because excitation and sonification have been separated, allowing for mapping and manipulation, which led to the development of gesture based instruments and responsive environments. This paradigm allows for the input to the instrument to

[1]The magnitude of an astronomical object indicates its brightness. A lower number indicates a brighter object.

come from the output of a series of one or more sensors, which can then be manipulated and mapped to produce sound or music based on those inputs (Fraietta 2006). The RA and Dec. from the sensors are used as inputs to various algorithms that will generate sounds or signals based upon those inputs. There is no *correct* way to compose a work based on this input, nor is there a *correct* style of music. Similarly, there is no algorithm or computational process that could be deemed *correct.* Instead, I suggest different ideas whereby a composition can be based upon creative algorithms that are fed by the sensor input.

3.1. Star map table generation

One method of composition based on a fixed point in space is to create a table of the stars around the celestial point. This table can be created by querying an online database called *VizieR*. Using the RA and Dec. as query parameters, Vizier returns a table of stars with properties such as distance, magnitude, velocity[1] and colour (Ochsenbein, Bauer, and Marcout 2000). For example, if our binoculars were roughly pointed to the star *Acrux,* the RA and Dec. would be approximately 12 hours and -63 degrees. If we queried *VizieR* using these parameters, defining a two degree window, *VizieR* would return a table of over seventy stars with magnitudes[2] ranging from 4.2 to 11. This table can be expressed compositionally or musically in many ways.

Two different works were performed using the *VizieR* star map tables based on the *Crux* constellation. These pieces were experimental compositions to test the viability and practicality of the concept. The first was an improvisation using FM synthesis, while the second was an improvisational stimulus for live musicians. In both cases, the star table map was obtained by inputting the RA and Dec. into the *VizieR* server via a URL such as the following.

```
http://vizier.u-strasbg.fr/viz-bin/asu-txt?-source=I/311/hip2&-c=186.64975588%20-63.09905674,rd=2,eq=J2000&-out=RArad&-out=DErad&-out=pmRA&-out=pmDE&-out=Hpmag&-out=B-V
```

The parameters to this call are as follows:

`-source=I/311/hip2` indicates the astronomical databases to use. In this case, we are querying Hipparcos catalogue.

`-c=186.64975588%20-63.09905674,rd=2,eq=J2000` indicates RA of 186.64975588[3] in degrees and a Dec. of -63.09905674. This happens to be to location of *ACrux.* The `rd=2` parameter indicates it requests all stars within a two degree radius, while `eq=J2000` indicates that it should use the J2000 coordinate system.

The `-out` parameters define which parameters we want to return, which are: RArad — RA in degrees; DErad—Dec. in Degrees; pmRA and pmDE—proper motion in right ascension and proper motion in declination in milliarcseconds per year; Hpmag—magnitude; and B-V—the colour index.

The output is returned as text that includes a table with each star on a separate line. For example, the first line of tabulated information from that URL is:

```
182.43579591 -62.58184518    -5.90
3.09  8.1840  0.042
```

The returned text is parsed and each line is sent as an Open Sound Control message (Wright and Freed 1997) to the Smart Controller workbench application (Fraietta 2005) for algorithmic processing before sending as MIDI for sound generation.

3.2. Solo synthesiser and binocular realisation

Azimuth and altitude information for the binoculars was accomplished through the accelerometer and magnetometer method, described earlier in the paper, by mounting the circuit to the top of the binoculars in a custom built case, shown in Figure 1. The data was converted to RA and Dec. on the computer and sent as OSC data with a start and stop message to indicate beginning and end of the table data.

The Smart Controller workbench stored all the parameters into a ganged table set that enabled a single index to access the data parameters for a particular star. This effectively made the ganged tables behave as a single table with multiple outputs. The star database parameters were scaled and mapped to MIDI parameters as follows: Dec. to note number, RA to pan, magnitude to velocity, and colour to program change. Proper motion outputs were not used in this work. As each star was output, it would have its own loudness, pitch, stereophonic position, and synthesiser voice. The program changes where linked to five different synthesiser voices, depending on what range the colour value was, thus giving five perceivable colour bands. The stars, being defined by the table indexes, could be cycled using a set range *counter,* which gave a cyclic

Figure 1: Sensor unit mounted to binoculars

[1] Velocity in this case means the proper motions in RA and Dec. per year.
[2] Based on the magnitudes, only ten of the stars in this data set would be visible with the naked eye under dark skies.
[3] Converting hours to degrees is done by multiplying the number of hours by fifteen.

motif. Another method employed was using random indexes to select stars, and random durations between these indexes, which gave an indeterminate effect with wide spatial texture. A subset of star map could be easily selected by placing boundaries to the counter and random number generation. Starting with a lower number of selectable indexes and gradually increasing gave the effect that one experiences when star gazing, where more stars become visible over time (Norris and Norris 2009). It was not difficult to improvise the work using sliders on the software.

Although the work was successful as a study, the work could be realised using a Digital Signal Processing (DSP) sound engine. *Granular synthesis* is a method of DSP sound generation whereby the audio generation is based on small samples of sound, known as *grains.* These grains are multiplied, imitated, and layered, which can be used to create effects such as time stretching and pitch shifting (Xenakis 1992). The composer could choose to map individual grains to each of the stars in the data set, processing them according to properties of the star. For example, stars with higher velocities can be pitch shifted over time according to their speed and direction, or moved spatially through loudspeakers. Similarly, amplitude can be mapped to magnitude, while colour can be mapped to placement in the audio spectrum. The different mapping strategies and possibilities are only limited by the composer's imagination.

3.3. Improvisational stimulus for live musicians

Many composers today recognise that "a player can often create spontaneously something much superior to what can be written down with the crude approximation of notation. But he [or she] must take care that what he [or she] creates builds up into a monolithic whole" (Smith Brindle 1986, p. 146). In such a performance, it is imperative that the musicians are not only able to communicate with one another, but that they are able to work within the intended form of the composition. Although certain styles or genres of music base their "improvisation on a ground bass, on a theme, or using a skeletal harmonic framework such as jazz" (Smith Brindle 1986, p. 11), performers of other styles of aleatoric music are sometimes provided with graphic signs or suggested note groups. Smith Brindle states: "It is all too easy to fall into the trap of wandering on, following the whim of the moment, chopping and changing from sombre meditations to joyous paeans. To the listener, such music is incoherent and boring" (Smith Brindle 1986, p. 12). It is possible to use the algorithmic treatment of the data set derived from computations of the sensor to shape the musical form.

The author collaborated with Colin Bright to perform *sp-A-C--E* for Telescope, Binoculars & Improvising Ensemble (Bright 2013). The performance was musically based on an existing work, *Ground Control to Gliese* (Bright 2012), with the difference being that MIDI triggered samples based on the star maps were also output to the audience and fed to the performers for stimulus for improvisation. A large pair of mounted binoculars and a small telescope were employed to enable the audience the view the sky in real time. Although we originally intended to display images on a video screen from an astro-imaging video camera, this proved too difficult to implement. Furthermore, compared to what people already expect to see from astro-images, the display from a video camera on the screen would have been very unimpressive and was therefore abandoned. Instead, a slide show from images from NASA were cycled through as a slide show and displayed on a screen above the performers.

Although many saw the performance as a success, this was based almost entirely on the skill of the ensemble and their music performance, and virtually nothing at all with the astronomy. The most obvious problem with the performance was that the audience had to leave the room where the musicians were performing. The audience was expected to venture through the back door to go outside in order to view the stars through the binoculars. Consequently, few people ventured outside to the astronomical equipment. Furthermore, the audience could not see the sky and the performers at the same time, so a laser pointer was of no use as a guide during the performance. A second performance was made the following week, however, the sky was completely covered with cloud and it rained, so there was nothing to see through the binoculars. The audience, however, enjoyed the ensemble performance with the NASA image slide show with samples fed from stored star tables.

The success of this type of performance requires a completely new composition, rather than just plugging a pair of binoculars and some MIDI samples into an existing work. The performers and the astronomical equipment need to be under the night sky together for the performance to be experienced as single artistic entity.

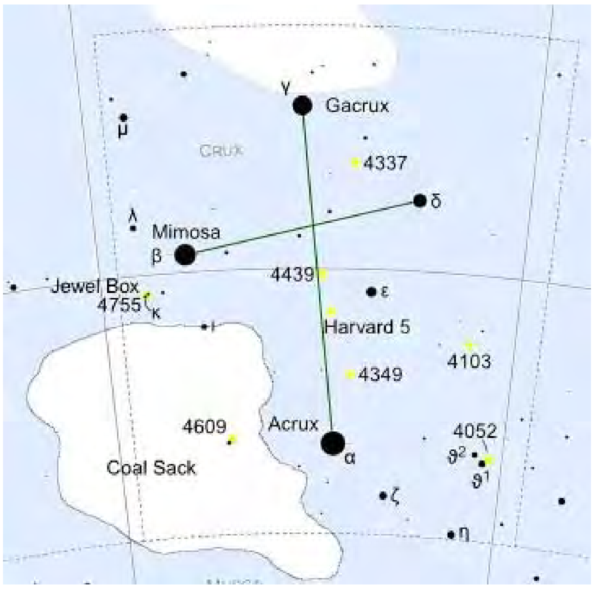

Figure 2: Crux Constellation

Consider a proposed compositional work based on the *Crux* constellation shown in Figure 2, better known as the *Southern Cross* (Massey and Quirk 2012). The focal point is the triple star system *Acrux,* which is used as the initial creative stimulus for the performers. Thematic material is derived from the data set generated from the star map table previously discussed. The data is used to generate MIDI notes, which are fed to the musicians' head phones, similar to how they might use a click track in a recording studio. Alternatively, a live set of motifs can be sent to the performers using a score generating program (Brown 2005). The audience sees the laser is pointing to the southern most star of the Southern Cross, blazing in the midst of the bright *Milky Way,* and when they take their turn viewing through the binoculars, they realise that what they saw as a single star with their naked eye actually appears as two through the binoculars.

The binoculars and laser pointer then move into the large dark nebular that appears to obscure the Milky Way known as the *Coal Sack* (Massey and Quirk 2012, p. 58). "To many Aboriginal groups, the Coalsack [sic] is part of the best known Aboriginal constellation—the Emu in the Sky" (Norris and Norris 2009, p. 5). Instead of providing stimulus from a *VizieR* data set; the performers are fed sounds into their headphones that represent traditional instruments of the indigenous peoples. For example, a drone may represent a didgeridoo, or rhythmic clicks represent clap sticks (Ellis 1985) indicating that they can *free form* with inspiration of the traditional indigenous peoples of that particular area.

Moving out of the *Coal Sack*, the binoculars and laser pointer move into "the pride of the Southern Cross and a favourite of most (southern) amateur astronomers" (Bee 2006 108, p. 108), *NGC 4755,* also "known as the 'Jewel Box' for its varied colours and brightness of the stars" (Massey and Quirk 2012, p. 98). Musically, a complete change in dynamic and style occurs when moving from the *Coal Sack* into to the Jewel Box. Colour spectrum data can be used to generate a MIDI note pitch table, which can be used to create melodic or thematic fragments to both guide and stimulate the musicians.

The work can continue throughout the constellation, moving through to each of the stars and into other *deep-sky objects* such as *NGC 4052;* or possibly the work could include *NGC 4609* in the middle of the *Coal Sack*. The challenge for the composer would be obtaining unity between one section of the work and the next (Smith Brindle 1986), which could be facilitated by an informative narrative that describes the next area of the sky that will be presented top the audience.

3.4. Programmable logic control of musical robots

Some composers today orchestrate for musical robots. These robots generate sound based on the mechanical excitation of a physical object, and are used as alternatives or adjuncts to live musicians or digital synthesis. Control of such robots is accomplished through the use of algorithmic software on personal computers that generate MIDI (Singer et al. 2004). Alternatively, there are also dedicated hardware devices that enable composition and performance based on programmable logic control, the output of which can directly excite musical robots (Fraietta 2006). Although the author has not realised this compositionally in any way using the star data, the concept is quite feasible. For example, star colour from the data set can be used to determine which set of wind chimes or bells are excited. The star count within the data set can also be used to control a loudspeaker loaded with small beads, which, when excited with *control voltages*, create a popping or crackling sound representing a star cluster.

4. CONCLUSION

This paper has described different methods of using algorithmic manipulation to provide compositional stimulus, and meaningful correlation for the audience between what they see and hear. Utilising algorithms that use input from sensors and star charts can provide composers and performers the stimulus needed to create music that has both freedom and form, while simultaneously encouraging audiences to marvel at the wonder of the night sky.

5. ACKNOWLEDGMENTS

The Australian composer Colin Bright provided some of the guiding stimulus for live instrumental improvisation. Frank Thompson of Islington Baptist Church built the case for attaching sensors and laser to Binoculars shown in Figure 1. This was done at as a part of Thompson's volunteer ministry at the church and was at no cost to the author. The constellation chart for Figure 2 was obtained from www freestarcharts.com. Mati Morel from the Newcastle Astronomical Society provided explanations regarding the colour data decoding from star catalogues.

6. REFERENCES

Bee, Robert. 2006. *Heavens above! A binocular guide to the southern skies*: Evalt Graphics.

Bright, Colin. 2012. *Ground Control to Gliese - SYZYGY.* Video. http://www.youtube.com/watch?v=3yco29c6T-Q&feature=youtu.be.

———. 2013. *spA-C--E*. Video. https://www.youtube.com/watch?v=4DZMA5BG6is

Brown, Andrew. 2005. "Generative Music in Live Performances". In proceedings of *Generate and Test: the Australasian Computer Music Conference*, 12 - 14 July, 23-26. Queensland University of Technology, Brisbane.

Chittick, Donald E. 2006. *The Puzzle of Ancient Man*. Newberg: Creation Compass.

Duffett-Smith, Peter, and Zwart, Jonathon. 2011. *Practical astronomy with your calculator or spreadsheet.* 4th

ed. Cambridge [England] ; New York: Cambridge University Press.
Ellis, Catherine J. 1985. *Aboriginal music : education for living, cross-cultural experiences from South Australia*. St. Lucia, Qld.: University of Queensland Press.
Fraietta, Angelo. 2005. "The Smart Controller Workbench". In proceedings of *International Conference on New Musical Interfaces for Music Expression (NIME-2005)*, 26-28 July 2005, 46-49. University of British Columbia, Vancouver.
———. 2006. The Smart Controller: an integrated electronic instrument for real-time performance using programmable logic control. Ph.D., School of Contemporary Arts, University of Western Sydney.
Harger, Honor. 2011. *A history of the universe in sound*. http://www.youtube.com/watch?v=mAAxykQH0Wo
Malin, David, and Wesley-Smith, Martin. 1989. "Star Trails and Quantum Leaping: Astronomy for new Audiences An Audio-Visual production by David Malin and Martin Wesley-Smith." *Newsletter of the Australian Astronomy and Space Liaison Group* (15):51-53.
Margolis, Michael. 2011. *Arduino Cookbook*. 2nd ed. Beijing: O'Reilly.
Massey, Steve, and Quirk, Steve. 2012. *Atlas of the southern night sky*. 3rd ed. Chatswood, N.S.W.: New Holland Publishers (Australia).
Norris, Ray P., and Norris, Cilla. 2009. *Emu dreaming : an introduction to Australian Aboriginal astronomy*. Sydney: Emu Dreaming.
Ochsenbein, F, Bauer, P, and Marcout, J. 2000. "The VizieR database of astronomical catalogues." *Astronomy & Astrophysics Supplement Series* 143 (1):23-32.
Seronik, Gary. 2013. "A Go To Binocular Chair." *Australian Sky & Telescope* 9 (1):89.
Singer, Eric, Feddersen, Jeff, Redmon, Chad, and Bowen, Bil. 2004. "LEMUR's Musical Robots". In proceedings of *International Conference on New Interfaces for Musical Expression (NIME)*, 181-84. Shizuoka University of Art and Culture, Hamamatsu, Japan.
Smith Brindle, Reginald. 1986. *Musical composition*. Oxford: Oxford University Press.
Wallace, Ken, Dawes, Glenn, and Northfield, Peter. 2012. *Astronomy Australia 2013:*
Your Guide to the Night Sky. Georges Hall: Quasar.
Winkler, Todd. 1995. "Making motion musical: Gesture mapping strategies for interactive computer music". In proceedings of *International Computer Music Conference*, 261-64. Banff, AB, Canada.
Wright, Matthew, and Freed, Adrian. 1997. "Open SoundControl: A New Protocol for Communicating with Sound Synthesizers". In proceedings of *International Computer Music Conference*, 101-104. Thessaloniki, Hellas.
Xenakis, Iannis. 1992. *Formalized music : thought and mathematics in composition*. Rev. ed. ed, *Harmonologia series ; no. 6*. Stuyvesant, NY: Pendragon Press.

MOMENTUM. HARMONY. AUDIENCE ENGAGEMENT AND PARTICIPATORY ARTS PRACTICE

Nat Grant
Victorian College of the Arts
Contemporary Music/Production

ABSTRACT

Momentum is a cumulative music composition project conducted throughout 2012. For 366 consecutive days I recorded and collected sounds and musical material, sculpting and layering them utilizing digital and electronic manipulation and processing. I invited contributions to the project, and more than 60 people from around the world collaborated on *Momentum*, providing sounds to be included in the project.

Each day a small audio snapshot of the work appeared online alongside a blog, sharing information on sound sources and compositional approaches. *Momentum* is presented in 12 movements (one for each month of the year 2012), and has since gone on to also encompass a 30-minute album, a 4 day/night live performance, and an ongoing, online community sound art collective. The work is accessible via several online sources[1], and the audio is free to listen to and download.

This presentation will focus on one specific area of my research, that of harmony with audience, encompassing the participatory nature of *Momentum*, and the open and transparent way in which each of the projects was created. The joining together of sounds, sonic environments, experiences, skills, artists and audience throughout each of the *Momentum* projects promoted dialogue, feedback and collaboration.

Audience: Showing Process: Collaboration

Audience played a crucial role in all the stages of *Momentum*. By delivering each of the projects online I was able to instantly share my compositional processes and ongoing outcomes with an increasingly large audience. The online format of *Momentum* also made it very straightforward for audience members to contribute to the project, and to provide feedback on individual tracks as well as on the process as a whole. Throughout 2012 and 2013 I collaborated with dozens of audience members, existing and new friends and artist colleagues, and was also able to facilitate collaborations between these people.

One motivation for presenting *Momentum* daily via a blog and sound hosting site was to share the process of creating a sound work with an audience, *as* it was happening – not only after the fact, in the form of a complete work, cd etc. A deliberate transparency of process combined with free and easy access to the work itself promoted regular conversation, feedback and collaboration with my audience throughout the entire process. Creating *Momentum* involved taking the audience along for the journey, and inviting feedback and engagement with the creative process.

Much of my work draws attention to process as a part of the experience of performing or recording. Like a real time 'making of', my art draws attention to a "selected aspect of experience,"[2] that of using sound as a sculptural medium, and of listening to the world always with the next stage of my project in mind. Similarly in his PhD thesis *Blogging as Art*, performance artist Lucas Ihlein conducted two particular durational art project in Sydney and Western Australia in 2005 and 2006, which were blogged about and also presented in physical hard copy on the day of creation, "so that the moment of publication (the moment, that is, of engaging with the public) was not deferred to some future moment when the artwork was deemed to be complete."[3]

There are countless examples, also, of artists and artisans conducting '365' projects, and blogging or making public each days creation/s. US artist Noah Scalin's *Make Something 365* blog hosts hundreds of these types of projects - *Momentum* was featured on his site in November, 2013.[4]

In the initial stage of *Momentum*, for each day in 2012, I posted on my own blog, describing the kind of sound/s I had recorded that day, how they fit into the overall mix at that time, what kind of digital manipulation had been applied to the recording, and why

[1] . "SoundCloud." 2012 - 2014, from http://soundcloud.com
, Grant, N. "Momentum." 2012-2014, from http://momentumproject.blogspot.com.

[2] Ihlein, L. (2009). Framing Everyday Experience: Blogging as Art, Deakin University.

[3] Ibid.

[4] Scalin, N. "Make Something 365." Retrieved November 30, 2013, from http://makesomething365.blogspot.com.au/.

those sound/s and tools had been chosen on that particular day.

For example, on Day Three Hundred and Four (October 30) the blog post reads:

> *Living in Melbourne means several false starts as far as the seasons go. But it was definitely warm enough to put the fan on and break out my thongs today ... I recorded myself walking on concrete around the side of my house to the washing line, then duplicated the recording so there are two of me, walking out of sync. Plus timpani from yesterday with more EQ experimentation: this time I took out some of the lower frequencies, making them sound more tinny, like they were recorded in a tin can rather than a large carpeted room.*[5]

The daily blog posts were promoted on Facebook and twitter, and gave people the option to subscribe to receive an email with each new blog post. This resulted in audience members 'following' my progress each day, and coming to a greater understanding of the kinds of processes involved in creating my music than from merely hearing finished tracks. I received comments throughout the year describing the personal experiences of different followers, and new ways in which they were appreciating many of the sounds in their own environments.

This also meant that if someone wished to contribute a sound to the piece, or to use any part of *Momentum* in a new or derivative work, they could quickly and easily listen and gauge the general aesthetic of the work, understand the kinds of sounds that had been recorded and contributed already, and access the entire work to date, if they so wished.

From my own experience as an artist and audience member – people are interested in *how* creative work is made. Contrary to Bayles & Orland's views in their book *Art & Fear*, that "to all viewers but yourself, what matters is the product: the finished artwork. To you, and you alone, what matters is the process: the experience of shaping that artwork."[6] Each of the *Momentum* projects suggest otherwise, that it is of interest to my audience how I went about creating *Momentum*, that it provides insight into the artwork itself and my creative process.

I consciously made the decision to blog and post a track each day in each stage of *Momentum*. I was interested to share my compositional processes with my audience; "…a kind of artistic and intellectual intimacy that lets other see *how* they reached a specific point, not simply that they did reach it."[7]

One particular contributor - a first time field recorder - emailed me in May 2013, saying: "Your soundscapes are altering our ways of hearing." She thought it was a "Great idea to have lots of people out 'listening' more intentionally."[8] This woman was amongst others who kept contributing sounds after the initial, yearlong *Momentum* project had concluded, which inspired the idea of an ongoing, online sound collective. This has shown me that my audience is interested in my process, in contributing to the creation of new work and engaging in ongoing dialogue about the work and the process.

Having an audience who is engaged with my work and my process and has even contributed to the work adds value in terms of making a serious engagement with that audience, and was one of the motivations for conduction *Momentum* in plain sight. I am not interested in, as Bayles & Orland put it, my art remaining "the province of genius… something to be pointed to and poked at from a safe analytical distance."[9]

Throughout each stage of *Momentum* I was keen for other people to be involved. If someone has mentioned an interesting sound somewhere in their lives I'd request a recording and photo. If other musicians expressed interest in the project I'd invite contribution. I was keen to have as many different people involved as possible.

From the beginning of the creation of *Momentum* I made it known on the blog, SoundCloud and social media that I was interested in others contributing sounds to the project. On the seventh day I received my first contribution, and five more before the end of the first month. This snowballed throughout the year, and I received more and more external contributions to *Momentum* as the project progressed.

The online nature of the project and the ability to send media files via email and dropbox[10] made these collaborations straightforward and instantaneous. When someone sent me a sound I would usually incorporate it into the work straight away, meaning I could also send them the link to a daily mix with their sound included within days. This made for several repeat contributors, as it was a fairly quick and also fun way to contribute and collaborate and hear a finished product without waiting months for an album to be released. Contributions came from ongoing or previous musical collaborators, from friends with no musical knowledge or experience, from followers on social media sites,

[5] Grant, N. "Momentum." 2012-2014, from http://momentumproject.blogspot.com.

[6] Bayles, D. a. O., Ted (1993). Art & Fear. USA, Image Continuum Press.

[7] Ibid.

[8] Rasmussen, C. (2013). Rain on tin. N. Grant.

[9] Bayles, D. a. O., Ted (1993). Art & Fear. USA, Image Continuum Press.

[10] http://dropbox.com

from people who had learned of the project by word of mouth, and from those who had found the project online or in other ways.

In addition to this there were inadvertent collaborations along the way between various contributors, as on occasion I would receive contributions to the project several days in a row. Sometimes these contributors knew each other, other times not, and at times I was able to introduce artists to each other in this way. For example, on June 10 (Day One Hundred and Sixty Two), I incorporated sound recordings from two different contributors – Melbourne visual artist and my friend Lily Mae Martin, and sound artist Greg Hooper from Brisbane, who I came across through the ABC Pool site.
Greg's recording of him painting his potting shed included birds and other ambient sounds. Lily's recording was of her young daughter Anja, singing and playing with a tin music box.[11]

Conducting *Momentum* gave me the confidence to approach other artists and ask for a contribution. Given the small request, that there were no limits on the type or length sounds to be contributed (it could be as little as a few seconds), I felt able to ask, and most people obliged. New connections were made and new collaborations were borne out of this.

The snowball effect of people contributing to *Momentum* resulted in my decision to have each track in the month of December made up of others' contributions. For the last 31 days I acted as curator, digitally manipulating, blending, building, altering and sculpting with sounds sent in by other artists and members of my audience, and maintaining a style that was coherent and consistent with the rest of the work.

In 2013 I was still receiving contributions periodically, which prompted me to start an online sound art collective on the SoundCloud site where people can contribute source sounds but also where members can download each others' sounds, re-mix them and post back to the group.

Momentum has extended beyond and outside of itself. The initial, yearlong project inspired me to create the three subsequent projects, and the process of disciplined practice and cumulative composition has continued, flowing on to new projects and ideas. Involving my audience in the project was a big drive to continue to create and share my work. Showing where I was up to every day, getting feedback and encouragement to keep going, perhaps even inspiring others to do the same, and receiving so many contributions that turned my audience into collaborators and took the project in new and exciting directions – these were all vital elements in the success and, I believe, completion of the year long *Momentum* project.

Working in this way gave both me and my audience deeper insight into my art and my process, and provided inspiration to continue to work in this way in the future – in tandem with an audience who are both engaged and involved in the creation of my art.

REFERENCES

. "SoundCloud." 2012 - 2014, from http://soundcloud.com

Bayles, D. a. O., Ted (1993). Art & Fear. USA, Image Continuum Press.

Grant, N. "Momentum." 2012-2014, from http://momentumproject.blogspot.com.

Ihlein, L. (2009). Framing Everyday Experience: Blogging as Art, Deakin University.

Rasmussen, C. (2013). Rain on tin. N. Grant.

Scalin, N. "Make Something 365." Retrieved November 30, 2013, from http://makesomething365.blogspot.com.au/.

[11] https://soundcloud.com/natgrant/one-hundred-and-sixty-two

DETERMINING SONIC ACTIVITY IN ELECTROACOUTIC MUSIC

David Hirst
School of Contemporary Music
University of Melbourne
dhirst@unimelb.edu.au

ABSTRACT

This paper seeks to extend the author's SIAM Framework for the analysis of electroacoustic music by posing the following question: What propels an electroacoustic work along from moment to moment, section to section, or scene to scene, given the absence of harmonic cadences associated with traditional tonal music? Taking as its subject matter, Jonty Harrison's piece *Unsound Objects,* the paper provides a brief outline of the SIAM Analysis Framework, describes a previous analysis of *Unsound Objects* using this method, and then extends the analysis to seek a measure for "sonic activity" as an indicator of periods of tension and release within the work. The aim of the current study is to seek an alternative method to the use of "spectral irregularity" for measuring sonic activity in electroacoustic music. This paper combines a number of time-related tools from the program Sonic Visualiser to create Inter-onset Time and Inter-onset Rate plots for *Unsound Objects.* The paper concludes that determining inter-onset time can give us a plot that is useful in showing the main sections within a work. Calculating its reciprocal, inter-onset rate, can generate a graph that provides some measure of the varying activity within an electroacoustic work. The next step in research is to seek to display further subtleties in the organisation of time, eg. Rhythmic patterns.

1. DEFINING THE PROBLEM

While undertaking a recent analysis of Jonty Harrison's electroacoustic musical work, *Unsound Objects* (Hirst 2013) the initial phase involved analysing the acoustic surface to identify sound objects. The next phase required an examination of relationships between sound objects, giving rise to the following question: What propels the work along from moment to moment, section to section, scene to scene? To help answer this question, I observed that an increase in sonic activity seems to elicit expectation in the listener that an important event is about to occur. There is a build up that seems to require a release the longer the build up goes on. But how can we measure something I have called "sonic activity" and, even better, how can we display sonic activity easily within a work? Can some form of signal processing be used and be represented to assist in the interpretation of electroacoustic musical works?

In the analysis of the Harrison piece, I began to address these questions, but there was not sufficient space within the confines of that article to discuss the issues in any great depth. This paper provides an opportunity to take the discussion further, and thereby expand and refine the author's SIAM framework for the analysis of electro-acoustic music. In the following sections, I will briefly describe Harrison's *Unsound Objects,* recap on the definition of the SIAM (Segregation, Integration, Assimilation and Meaning) framework, and then elaborate on what I call horizontal integration, which involves an examination of the temporal dimension of an electroacoustic musical work. This will lead on to a discussion of sonic activity and some possible ways of representing it.

2. HARRISON'S *UNSOUND OBJECTS*

Unsound Objects is a substantial work of some 12 minutes 59 seconds (779 secs). First performed as a commission for the 1995 International Computer Music Conference in Banff, the version used in this analysis appears on the Articles indéfinis CD (Harrison 1996). While often regarded as a traditional "acousmatic" composer, in a 2002 interview Harrison said this about the use of recognisable, recorded sounds in *Unsound Objects:* 'As it would therefore be impossible to expect listeners to be able to exercise "reduced listening" in such a framework I chose instead to exploit this very recognisability in working on the piece. This is an implicit challenge to Schaefferian orthodoxy about the definition of the sound object through reduced listening; it is dangerous and, hence, "unsound".' (Palmer 2002).

An analysis of the whole of *Unsound Objects* was carried out using a modified form of the author's cognition-based SIAM framework for the analysis of electroacoustic music. The detailed analysis of *Unsound Objects* is available in that previous article (Hirst 2013), but here I will briefly recap some of the SIAM concepts in order to provide some sort of context.

3. THE SIAM ANALYSIS FRAMEWORK

The SIAM framework (Hirst 2008) stands for Segregation, Integration, Assimilation and Meaning. It can be briefly summarised as follows:

Segregation: identification of sound events and the factors responsible for identification.

Integration – Vertical: consider vertical integration and segregation as a cause of timbre and texture variance; consider psychoacoustic dissonance and musical dissonance; consider emergent properties relating to pitch (horizontal overlap).

Integration – Horizontal: identify sequential streams and patterns of sonic objects; determine causal linkages, relationships and possible syntaxes; consider organisation in time and the horizontal integration of pitch (where appropriate).

Assimilation and Meaning: consider various types of discourse, e.g. discourse on the source-cause dominant (semantic) to typological-relational dominant (syntactic) continuum; consider global organisation in time and any hierarchical relationships; consider expectation-interruption, arousal and meaning, semiotic relations, etc.

4. ANALYSIS OF *UNSOUND OBJECTS*

Using the SIAM framework as a starting point, the analysis of *Unsound Objects* was adapted according to the directions dictated by the work itself.

The first part of the analysis can be described as **analysing the acoustic surface**. This involved "segmentation". Large scale segmentation into sections, and then small-scale segmentation of sound events from each other. The spectrogram and audio waveform displays were useful for the process. Sound events were annotated on the spectrogram and it was possible to get a time-stamped listing of the annotation layer, using the program Sonic Visualiser (Cannam et al 2010), which was then imported into a spreadsheet program (Microsoft Excel) and printed as a listing of all the annotations. The visual screens and printed time-stamped sound object listings became the data that facilitated detailed identification and specification of sound events within the aurally identified sections of the work.

The next phase of the analysis involved **moving beyond the acoustic surface** to examine structures, functions and motions between sound events. By "zooming out" to look at longer sections of the work, or carrying out "time-span reduction", we can observe changing sonic patterns over the course of the work. We can look at the different sections and ask questions like: What propels the work along from moment to moment, section to section, or scene to scene? To help answer this question, we can observe that an increase in sonic activity seems to elicit expectation in the listener that an important event is about to occur. But how can we measure and, even better, display activity within a work? Well the Sonic Visualiser program provides access to a suite of plugins of signal analysis. In the *Unsound Objects* article, I postulated that the type of analysis that seems to correlate best with sound object activity is a plot of "spectral irregularity" versus time.

There are several different methods for calculating the irregularity present within a spectrum, but essentially they both give a measure of the degree of variation of the successive peaks of the spectrum. Jensen, for example, calculates the sum of the square of the difference in amplitude between adjoining partials (Jensen 1999). What I am postulating here is that where there is a large variation across the spectrum, partial to partial, then this can provide us with a depiction of a high degree of activity. Figure 1 depicts an irregularity plot for the whole of *Unsound Objects*. Within that article I noted that it could be argued that we could simply use a plot of overall amplitude, or one of its variants, but I asserted that simply using amplitude tends to average out "activity", whereas spectral irregularity provides a much more fine-grained representation of activity from instant to instant through time. Subsequent work showed that RMS Amplitude plots give very similar graphs to those of irregularity. I shall return to this point later on.

The analysis of *Unsound Objects* then combined the use of spectral irregularity plots with aurally identified sections, within the work, to provide a detailed analysis of "activity" and to tabulate "sound types" for each section. This table showed "activity amount and type" and "selected sound object types".

The work actually divides into two main halves and after the two halves were compared, a summary of sonic archetypes (in the form of mimetic archetypes and structural archetypes), sound transformations, functional relations, and sonic activity was discussed.

In the final discussion section of the *Unsound Objects* analysis article, I made a number of observations and conclusions, and then ended with this: "Finally, we have used spectral irregularity plots to assist with the identification and isolation of periods of activity within different sections of the work. It seems as though it may be a useful measure for the analysis of other works – a hypothesis remaining to be tested."

This current article documents the follow up work to examine how best to represent and analyse "activity" within an electroacoustic music work.

5. ALTERNATIVE METHODS FOR DETERMINING ACTIVITY

The aim of the current study is to seek an alternative method to the use of "spectral irregularity" for measuring activity in electroacoustic music.

In essence, activity could be defined as the number of sound events in a given time period. Therefore we are interested in the onset time of each sound event, and its duration. Let's start with onset time. What signal analysis tools exist for determining sound event onset time within a musical work?

The program Sonic Visualiser, has a number of tools within it to perform such an analysis. Aubio onset detection (aubio.org) has eight different types which all produce a single list of time "instants" (vertical lines when plotted) of individual start times. This output can be exported to a spreadsheet. Their algorithm can be varied to suit the source material. The Queen Mary, University of London, in-built Sonic Visualiser onset detection algorithm lists three types of onset detector, but these are just the one detector with lots of variables: Program; Onset detection function type; Onset detection sensitivity; Adaptive whitening; Channel options for stereo files; Window size; Window increment; and Window shape. Output is an "onset detection function" which is a probability function of a "note" onset likelihood.

In developing a method for the detection of onsets in *Unsound Objects,* combining several forms of representation was found to provide a more reliable guide to data gathering rather than using any single plot. After some experimentation, the following combination was employed, using the Queen Mary algorithms:

1. RMS Amplitude (or Spectral Irregularity).
2. Smoothed detection function: Time Values (displays probability function of onsets).
3. Note onsets: Time Instants. Program: Soft Onsets; Onset detection function: Complex Domain; Onset detection sensitivity: 60%; Adaptive whitening: Yes.

This resulted in the onsets (#3 above) aligning pretty well with the smoothed detection probability (#2 above), but with some low level noise swells failing to trigger the onset detector (#3 above).

The "time instants" data (#3 above) was exported, then imported into an Excel spreadsheet in order to be able to make further calculations such as "inter-onset times" (the time between onsets). Figure 2 shows a plot of Inter-onset Time versus Time for the whole of *Unsound Objects.* Its peaks show us where there are long breaks in the work, and give a pointer to how the work may be divided up in analysis.

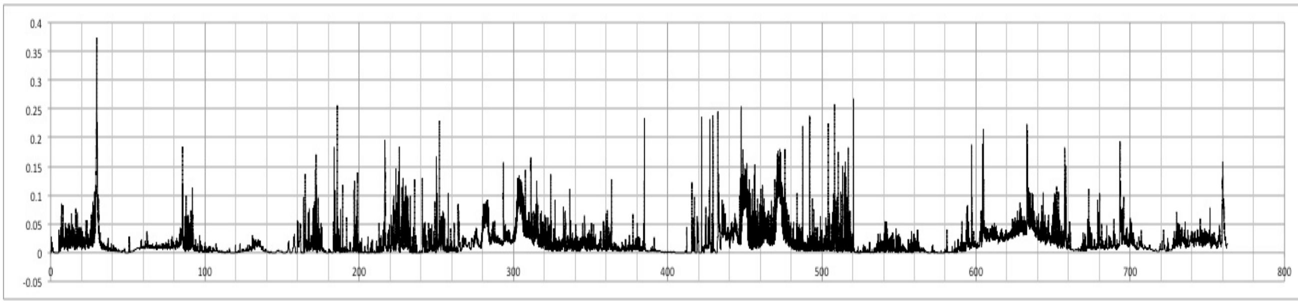

Figure 1. A spectral irregularity plot for the whole of *Unsound Objects*.

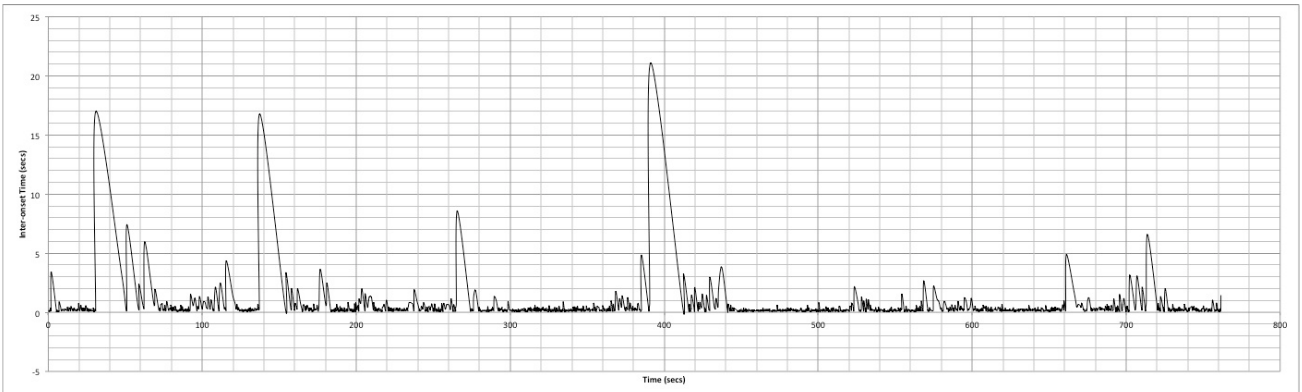

Figure 2. Plot of Inter-onset Time vs Time (secs) for the whole of of *Unsound Objects*.

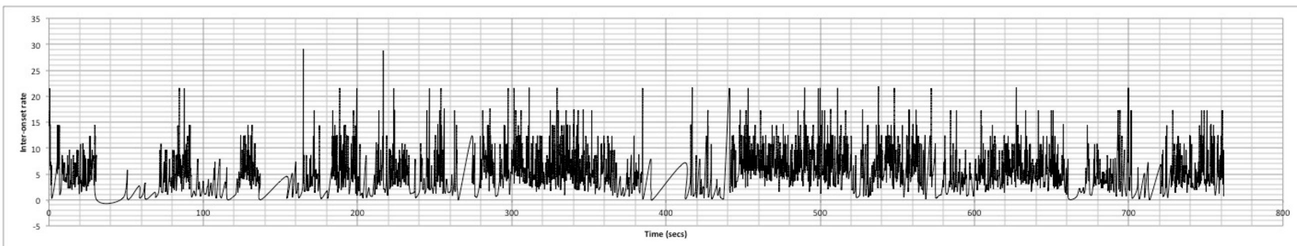

Figure 3. Plot of Inter-onset Rate vs Time (secs) for the whole of *Unsound Objects*.

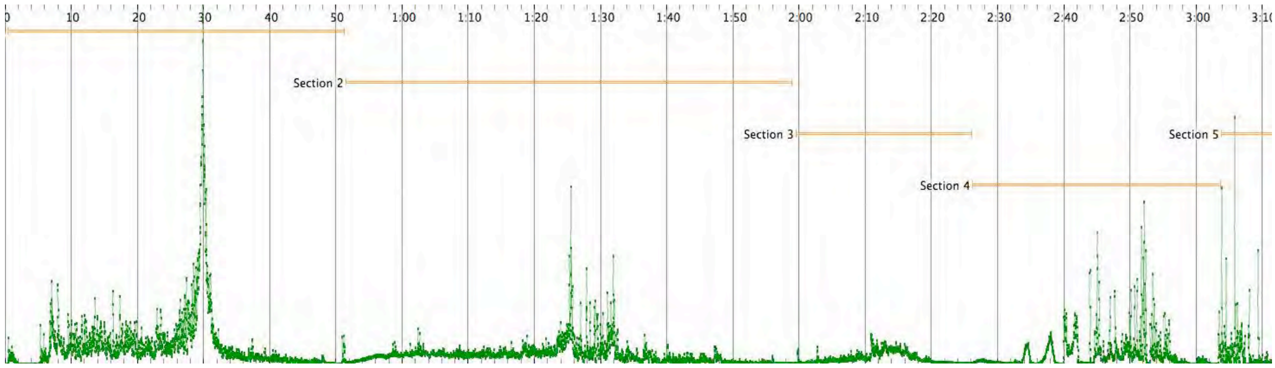

Figure 4. An irregularity plot for the first four sections of *Unsound Objects*. (Note: Time scale is shown in min:sec)

Displaying time instants, however, only progresses us part of the way to obtaining a measure of event "activity". Inter-onset "rate" was then calculated and plotted, as shown in Figure 3. This provides us with a measure of the number of onsets per second, which, in turn, provides a guide to the amount of event initiation activity at a particular time within the work.

6. OBSERVATIONS AND CONCLUSIONS

Determining inter-onset time can give us a plot (Figure 2) that is useful in showing the main sections within a work. Calculating its reciprocal, inter-onset rate can generate a graph that provides some measure of the varying activity within an electroacoustic work (Figure 3). If we had graphed Figure 3 at the beginning of the analysis, we would have observed that the piece does divide into two, with little activity between about 390 and 410 seconds. The first half begins with three bursts of activity, followed by a longer, more active phase of increasing activity until the "mid-break". The second half is more continuously active until around 660 seconds, where the work has several less active periods, perhaps in preparation for the end of the piece.

In the previous analysis of *Unsound Objects,* sections were first determined aurally, then superimposed over the irregularity plot (for an example this for the first four sections of the piece, see Figure 4). Comparing the plot of inter-onset rate (Figure 3) with the irregularity plot for the whole of the work (Figure 1), we can see that the piece appears to be much more active in figure 3 than figure 1, especially in the second half. The question remains as to which is a better measure of "activity"? The inter-onset rate is probably a more accurate method, but it seems exaggerated. This is possibly because it doesn't take into account the loudness of the events. Perhaps if this plot (Figure 3) was modified by the RMS amplitude, then a more useful picture of "effective activity" may emerge. There are also inherent definition problems for "iterative" sound events, such as drum rolls or machine sounds. Is such a sound type one long event or many short events? This phenomenon may skew the events per second data.

In terms of automating analysis, the inter-onset time plot (Figure 2) is very effective in identifying sections in a long musical piece, while the inter-onset rate (Figure 3) does provide a measure of active versus inactive depiction for various passages in a long piece.

The next step in this work is to extend temporal analysis further by looking at finer details within sections (see Hirst 2014).

7. REFERENCES

Cannam, C. Landone, C. & Sandler, M. 2010. "Sonic Visualiser: An Open Source Application for Viewing, Analysing, and Annotating Music Audio Files." *Proceedings of the ACM Multimedia 2010 International Conference*.

Harrison, J. 1996. *Unsound Objects*. On *Articles indéfinis*. IMED 9627. emprintes DIGITALes CD.

Hirst, D. 2008. *A Cognitive Framework for the Analysis of Acousmatic Music: Analysing Wind Chimes by Denis Smalley* VDM Verlag Dr. Muller Aktiengesellschaft & Co. KG. Saarbrücken.

Hirst, D. 2013. "Connecting the Objects in Jonty Harrison's Unsound Objects*." eOREMA Journal, Vol 1*. http://www.orema.dmu.ac.uk/?q=eorema_journal.

Hirst, D. 2014. "The Use of Rhythmograms in the Analysis of Electro-acoustic Music, with Application to Normandeau's Onomatopoeias Cycle". Paper submitted to ICMC 2014.

Jensen, K. 1999. *Timbre Models of Musical Sounds*, Ph.D. dissertation, University of Copenhagen, Rapport Nr. 99/7.

Palmer, J. 2002. In Conversation with Jonty Harrison. *21st Century Music*, Vol. 9/1 (USA, January 2002). Republished in eContact! 10.2 Entrevues Interviews(1). Montréal: Communauté électroacoustique canadienne Canadian Electroacoustic Community, August 2008. http://cec.sonus.ca/econtact/10_2/index html

SOGNO 102: HARMONIOUSLY REVISIONING COMPOSITIONAL TECHNIQUES OF GIACINTO SCELSI

A/Prof Cat Hope *Stuart James* *Dr Lindsay Vickery*

The Western Australian Academy of Performing Arts
Edith Cowan University

ABSTRACT

In 2013, the Western Australian Ensemble Decibel presented a program of late works of the Italian composer, Giacinto Scelsi (1905-1908) at the Goethe Institute in Palermo, Italy, entitled *Inner Space: The Giacinto Scelsi Project*. Amongst the pieces, were two original works by members of the ensemble, as a kind of homage to the composer. One of these compositions was by the director of the ensemble Cat Hope, entitled *Sogno 102,* a piece influenced by Scelsi's composition methodologies and a book of Scelsi's entitled *Sogno 101*, published by Italian publisher Quodlibet, in 2010. The book includes a range of articles derived from autobiographical audio recordings made by Scelsi himself, and transcribed for the book after his death. This article discusses the processes involved in conceptualising and realising the work, in particular the way electronics are used to realise some of the integral elements of Scelsi's own ideas about music into this work.

1. CONCEPTUAL OVERVIEW

Decibel is an ensemble that works with acoustic instruments and electronics. A key element to the Scelsi project was not that the works should all be works that feature electronics, and in fact Scelsi wrote few works that feature electronics, but were conceived using electronics. The majority of Scelsi's compositions were improvisations made on an Ondiola, one of two that he had bought in 1957 and 1958. The Ondiola, also marketed as a Clavioline in other countries, is a small, three octave electronic keyboard instrument attached to its own amplifier, with dials and keys for producing glissandi, quarter tones, vibrato and different timbres [Utti 1985]. It is an early monophonic synthesiser, with a modulator and filters.

Scelsi recorded his improvisations made on the Ondiola onto tape, which were often overdubbed and reorganised to form a final version, or to create different parts, which were then given to a transcriber to notate [Jaecker 1985, Utti 1995]. These transcribers included Vieri Tosatti, Alvin Curran, Sergio Caffaro and Frances-Marie Utti [Utti 1995]. His process at the tape recorder is a kind of compositional editing as can be seen in much electronic music today. However, these recordings were only ever intended as compositional starting points for notation. The original tapes are kept at the Fondazione Isabella Scelsi in Rome, and have only recently been used in live performance, as part of Klangforum Wien's *Giacinto Scelsi REVISTED* project in May 2013, at the Acht Brücken Music Festival in Cologne. The group rebuilt a Clavioline and performed some of Scelsi's original tapes, giving audiences a taste of the sounds of this instrument and an insight to Scelsi's compositional process.

Listening to the Clavioline on Klangforum Wien's introductory video [Klangforum Wien 2013] illustrates clearly the musical timbres Scelsi was able to produce on this instrument, and it is surprising how much it sounds like much of the acoustic realisations. This process of transcription: from electronic improvisation to fixed notation for acoustic instruments, as well as the more recent transcription completed by Klangforum Wien from audio tape back to electronic instrument performance are key conceptual elements engaged in the creation of *Sogno 102*.

Another element of transcription comes into play in the book, *Sogno 101*. The writings (though this hardly seems an appropriate term in this case) in Sogno 101 include anecdotes, reflections and poems. Some of them are funny, others are contemplative. Some quite confusing. When Hope attempted to discuss the book with composer Alvin Curran, a late mentee of Scelsi's, he cut her off abruptly, calling it 'that awful book' [Curran 2013]. He never elaborated, and one can only guess at his reasons for making such a claim. But it left her wondering if Curran thought the book was inaccurate, or contained material of a personal nature that was inappropriate for publication. Or more simply, sometimes, transcriptions do not tell a complete story. Writing is a very different process to that of speaking to tape, just as pre-composed works are very different form free improvisation. This contrast of capturing thoughts and ideas as they pour out in real time onto a tape, as opposed to transcribing them to text after a period of years, was of interest to Hope, and a concept explored in the piece.

Scelsi was interested in exploring the nature and potential of sound in his works. Inspired by musical and philosophical traditions of India and Tibet, he had an interest in the capacity of drone to present the inner energy of sound (Dickson 2012). Drone is another element key to this work.

2. WRITING A HOMAGE

The music of Scelsi had been an important influence on Hope's work as a composer. The relationship of improvising to composition, drone and glissandi had been explored in a range of her compositions, installations, improvisations and text scores. Writing a work as homage is simple in this case, as the influence of Scelsi's work is already in my work. Hope's work employs graphic notation, unlike Scelsi's, who only ever employed traditional music notation in his work, though he did notate some extended techniques.

The obvious elements in *Sogno 102* that can be identified with Scelsi are the use of drones, microtonal steps, glissandi, and over exaggerated vibrato. Scelsi explored all these techniques in his work, and the late string quartets provide good examples. In the first movement of *String Quartet No. 2* (1961), the instruments play long notes that vary in their timbre, attack and vibrato. They remain the same, or drone, for some time, often featuring glissando movements to nearby pitches, and intense tremolo's. Likewise, *Ko-Lho* (1966), a piece for clarinet and flute, which also made part of the Inner Space program, features these delicate shifts around sustained notes.

Scelsi's ideas about the quality and potential of sound are complex and detailed, and beyond the scope of this paper. Scelsi identified and separated two specific ideas about rhythm in his work, and they relate closely to the nature of sound. One is pulsation - the interaction of notes and silence, or of sounds in the air; the other being duration [Jaeker 2013]. This led Scelsi to investigate a time that was not sequential, but rather as a depth in sound [Zenck 1985]. Scelsi undertook what has been called 'Infrasonic' research' [Jaeker 2013], which seems to imply an examination of the vibratory quality of sounds. *Sogno 102* explores this vibratory quality by contrasting pure electronic sine tones with rich timbral qualities of acoustic instruments. This is a process examined in the comprehensive *Still and Moving Lines of Silence in Families of Hyperbola*'s (1973-1974) series of compositions by American composer Alvin Lucier, with pre-recorded tapes of sine tones set against various acoustic instruments.[1] The way this process has been used in *Sogno 102* is described in more detail below.

[1] Decibel performed these sine tone parts live in performance, with Lucier's approval in their performances and recordings of his work for three instruments and two sine tones, *Ever Present* (2002). This was discussed in a paper by the same authors for ACMC in 2010.

3. THE KEY COMPONENTS

The score for *Sogno 102* requires the Decibel ScorePlayer, an Application for the iPad developed by Decibel. This is a mechanism that puts graphic scores in motion on a screen, scrolling past a 'playhead' that is the moment that the music is to be read. It is ideal for music that is pulseless and does not use conventional music notation (Wyatt et al 2013). It also enables the performers to see their own part, only or all the parts together. *Sogno 102* is written for bass flute, bass clarinet, viola, cello, piano and electronics. Each instrument has its own colour on the score, to make them easy to identify. Pitch choices are free, but the relationships between the pitches is proportional on the score. Decibel always uses the same colour for the same instrument, wherever colour is employed in a score. This began with the development of Lindsay Vickery and Cat Hope's collaborative composition *The Talking Board* (2011) that uses coloured circles that performers follow. Though Hope has been using colour in scores since 2009, the system was not standardised for Decibel until the composition of *The Talking Board*. The choice of colours for each part aims to maximize the distinctness of each part using schemes explored for similar purposes in the visualization of data (Tufte 1990) and transport maps (Green-Armytage 2001).

The electronic part is notated with timed events that indicate points at which the sound of a specific acoustic instrument will be captured by a microphone, sampled and analysed. The resulting value, representing the instrument's strongest sinusoidal peak at that moment, is used to control a `cycle~` object, producing a sine tone at that frequency. The sinetones are spatially distributed between five speakers, one for each instrument. As the work proceeds the frequency of each sinetone is gradually altered, microtonally ascending or descending toward the conclusion, as notated in the score. The amplitude of each sinetone is also increased very gradually over the final 60 seconds of the work.

Each instrument is sampled only once, on its second iteration, so there are as many electronic voices as acoustic ones. In each performance, the sample pitch for each instrument will be different. This is because the acoustic instrument parts are not pitched on the score, and performers make different choices each time, but also because the computer operator will never sample them in exactly the same place every performance. Scelsi liked to say he was looking for the right sound, rather than the right note (Scelsi, 1981) - and this is the intention in *Sogno 102*.

Figure 1. One example of sample points for the electronics performer in the score for *Sogno 102*. This figure also shows the clustering of tones, the varying of a tone, the sine note appearing from the sampled instrument, and the central (dashed) reference line.

The score is proportional in its description of pitch, so each instrument makes a choice of pitch in relationship to the other instruments in the group, and to a central reference line, that runs throughout the piece. This dashed, opaque line shows where the original pitch chosen by the first player, the cellist in this piece, lies in relation to the other parts throughout the piece. It provides a development on the 'single note' style for which Scelsi became known his later works. This involved a single focal pitch or cluster of pitches which are subjected to a range of treatments [Drott 2006]. In *Sogno 102*, a range of pitches cluster together, but move as a whole. The reference to the original pitch is only evident to the performers, and they do not return to it.

The idea of sampling on the second 'note' of each instrument is a direct reference to the nature of transcription, described earlier. Often a statement is made, it can be re-iterated, but it wont ever is exactly the same. This references the audio process of dubbing on tape, where each pass disintegrates the quality of the recording and the tape, as well as the process of transcription in the book *Sogno 101*. Scelsi spoke to tape, someone listened and transcribed to the best of their ability. The clarity of the source is never discussed. Similarly, Scelsi's composition process seemed to have a large frame, two-step process: the improvisations to Ondiola, then the transcription to notation, which enabled the work to be performed by acoustic instruments.

This process of sampling to sine tones is one that Hope had used before, in the work *The Lowest Drawer* (2013). This work is for bass flute, bass clarinet and cello, but here the instruments are sampled thirteen times, in many different places of the piece, and the samples, whilst also transcribed to sine tones, do not change pitch as they do in *Sogno 102*.

4. ELECTRONICS IN SOGNO 102

The Lowest Drawer saw an early inception of the Max patch and score type for *Sogno 102*. The patch was developed by Decibel member Lindsay Vickery, and designed to track the fundamental frequencies analyzed from three microphones, one for each instrument, and allocates them to sine tones of identical pitch at the instant they are specified in the graphic score, played back in a quadraphonic speaker set up. The sine tone is really only audible when the instrument varies from the pitch in even the most subtle way, and once the instrument concludes the note sampled. Timed cues were originally triggered automatically from the score, which ran in Max, rather than the Decibel ScorePlayer, and projected in the performance space. A subsequent performance using eight powered loudspeakers, two bass guitar amplifiers, and one powered subwoofer saw further development of the patch by Decibel member Stuart James. For ease of signal redirection James added a configurable matrix that enabled customized speaker routings on the fly to the eleven individual speakers. The variations in the instrumentalists' distance and dynamic often resulted in tones 'popping out' loudly, limiting the effect of the acoustic instrument masking the tone. This effect could be exaggerated according to different perceptions of the tone in different areas of the audience space. As a means of managing the perceived loudness of different events, taking into consideration the distance of various loudspeakers, and their differing frequency responses, it was necessary to add facility for controlling the relative loudness for each individual sine tone, tuning their relative loudness for the performance space, and a global level control for the electronics performer to manage during the performance.

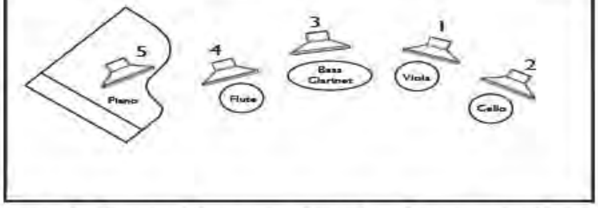

Figure 2. One possible set up of speaker placement according to channel number. The speakers can be placed directly beside or behind the performers as no feedback can occur in this microphone/speaker electronics scenario as recording and playback never occur together.

Pressure gradient, or more specifically cardioid, dynamic microphones are best suited in order to avoid spill from both the other instrumentalists on stage and the tones in the speakers. This is beneficial for tracking the fundamental frequency generated, particularly to compensate for the natural roll-off in frequency response that exists with smaller diaphragm dynamic microphones such as the Shure SM57. *Sogno 102* requires the sampled tones to change pitch as notated on the score, and the patch was constructed to automate this, only requiring the electronics performer to turn the audio on, trigger the cues as per his/her own copy of the score on the Decibel ScorePlayer, networked with the other performers, and facilitate the crescendo at the end. Input and output gain structures can be sound-checked and preset, and internal parameter changes are generated and managed by the patch automatically. Trigger cues are color matched with the score, and the current frequency after it is detected and modulated is displayed.

Figure 3. The interface window of the Max/MSP patch for *Sogno 102*

The *sigmund~* object was used to track the fundamental frequency of five instrumentalists. A low pass filter was inserted just before this signal analysis stage, performed by the *sigmund~* object, to avoid anomalies in fundamental frequency detection that occasionally occurred in *The Lowest Drawer*. James used four numerical accumulation functions generate the change in pitch determined by an initial value describing the rate of change, or gradient as shown in Figure 4. Some of accumulators within the patch were biased to generate in upward (positive) or downward (negative) direction. In theory this raises some questions in relation to the pitch scale, as to whether it is logarithmic or linear. Due to the fact that pitch change is determined linearly in frequency, this means they are perceived by the listener to be logarithmic in nature, meaning lower frequencies shift at an increasingly larger gradient than higher frequencies. The Max patch for the piece is available from http://www.cathope.com/sogno-102-2013.html(.)

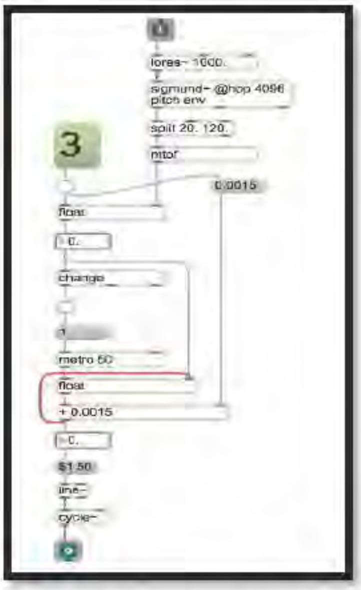

Figure 4. The accumulator function of the Max/MSP patch for *Sogno 102*.

The interaction of the electronics and acoustic instruments is subtle. The gradual pitch changes in the sine tones is almost imperceptible, and the different playing techniques applied to single sustained notes highlight these very slight differences in pitch, as well as the timbral difference of the sine tone and acoustic instrumental tone. Beating often eventuates between the instrument and tone, and between the tones and instruments themselves. When the sine tones crescendo, they appear to change timbre, as if they take on some characteristics of the speaker that they play through.

Figure 5. The end of *Sogno 102*, showing the crescendo in the electronic parts.

5. CONCLUSION

The development of *Sogno 102* explores the notion of a harmonious homage from conceptual and craft aspects. The practical development of the electronics for the work has grown from a process of trying ideas in different settings, and finding practical solutions for different issues that arise at each performance and rehearsal. *Sogno 102* does not sound like a work by Giacinto Scelsi. But in the context of the concert, Inner Space, it adds a rich dimension to the impact and reach of Scelsi's work beyond his lifetime.

6. REFERENCES

Curran, A. Personal communication, August 2013.

Dickson, I. (2012). Towards a Grammatical Analysis of Scelsi's Late Music. Music Analysis, 3/11. p 217

Drott, E. 2006. *Class, Ideology and il caso Scelsi.* The Musical Quarterly, 89 (1) p 81.

Green- Armytage, P. (2001). COlour Zones: connecting colour in everyday language. *SPIE Proceedings* 4421, 9th Congress of the International Colour Association, 976 June 6, 2002.

Klangforum Wein. 2013. *Giacinto Scelsi REVISITED* http://youtu.be/USk6UrE8cuI, Accessed April 14, 2014.

Hope, C. (2013). *The Lowest Drawer.*

Hope, C. (2013). *Sogno 102.*

Jaeker, F. 2011. "Funziona? Or Non Funziona? An excursion through the Scelsi archives". In Sciannameo, F. and Pellegrini A. C. (eds). 2013. *Music As Dream Essays on Giacinto Scelsi.* London: The Scarecrow Press. Pp 148-149.

Lucier, A. (1974). *Still and Moving Lines in Silence of Families of Hyperbola's*. Frankfurt: Material Press.

Montali, A. 1998. "Conceptual Reflections and Literary Evolution in Giacinto Scelsi's Works". In Sciannameo, F. and Pellegrini, A. C. (eds). 2013. *Music As Dream Essays on Giacinto Scelsi.* London: The Scarecrow Press. p 55.

Scelsi, G. 1981. *Son e musique*. Rome: Le Parole Gelate, p12.

Scelsi, G.1961. *String Quartet No. 2.* Paris: Editions Salabert.

Scelsi, G. 1966. *Ko-Lho.* Paris: Editions Salabert.

Tufte, E. R. (1990) Envisioning Information. Cheshire CT: Graphics Press.

Utti, F.M. 1995. "Preserving the Scelsi Improvisations". *Tempo,* 194. October, p. 12.

Vickery, L. and Hope, C. (2011) *The Talking Board.*

Wyatt, A., Hope, C., Vickery, L. and James, S. 2013. "Animated Notation on the iPad: or Music Stands Just Weren't Designed to Support Laptops". *Proceedings of the International Computer Music Conference,* Perth: The International Computer Music Association. p 203.

Zenck, M. 1985. "L'irriducibilita come criterio d'avaguardia: Riflessioni sui quattro quartette per archi di Giacinto Scelsi." In Cremonese, A. (ed). *Giacinto Scelsi.* Rome: Nouva Consananza and Le Parole Gelate. p 76.

PROTOCOL AND MODULARITY IN NEW DIFFUSION SYSTEMS

Bridget Johnson
New Zealand School of Music
Victoria University of Wellington

Michael Norris
New Zealand School of Music
Victoria University of Wellington

Ajay Kapur
School of Engineering and Computer Science
Victoria University of Wellington

ABSTRACT

This paper proposes a move to further modularity in spatialisation systems. The focus is specifically on systems that are driven by the design of new diffusion user interfaces. The authors recognise a need for system specificity based on the communication needs of different user interface designs. As examples the paper presents two case studies of new user interfaces and how they are each integrated into fully modular real-time spatialisation systems. The paper also recognises a need for intuitive and shared protocol formats in order to achieve modularity, some examples of this are proposed.

1. INTRODUCTION

In the past 15 years a wide range of new spatialisation systems have emerged, many of which exhibit the potential to be segmented. Full spatialisation systems are made up of many parts from the user interface through the spatialisation algorithms and finally the audio drivers and speaker systems. One common trend amongst new systems is to have them be made up of a series of modules that allow integration between modules of varying systems. In order for this to be achieved all systems must implement the same communication protocols. This paper will discuss two options for implemented modularity and protocols that are currently under development at the author's institution. The solutions proposed in this paper are aimed at systems that place a strong importance of user interface design, and the incorporation of physical user interfaces into wider spatialisation systems.

The paper first discusses movements and systems in the paradigm of diffusion performance that have led to a move towards modularity in spatialisation systems. It then goes on to discuss new practice driven research that is currently being conducted by the authors, and how the desire for modularity has influenced the design of two new spatialisation projects; tactile.motion and Chronus_2.0, in different ways. The fourth section continues to use these two new systems as examples for discussion about the need for universally accepted protocol in order to ensure the modular aspects of diffusion systems are viable for the future. The paper is concluded with a look to the future of spatialisation systems for performance practice.

2. RELATED WORK

For the most part developments in spatialisation performance (other than algorithm developments) have remained specific to the institutions developing them. This is largely due to the fact that each spatialisation performance system depends on the hardware available.

A number of the first speaker orchestras including the GRM Acousmonium (Desantos, Roads, and Bayle 1997), BEAST (Harrison 1999), and the Gmebaphone (Clozier 2001) were able to travel and therefore had to be easily configurable for varying performance venues. Traveling with these systems involved moving the entire system, from computers to amplifiers and mixing desks as well as any number of loud speakers. The physicality and cost of moving such immense amounts of hardware may have limited the potential distances these orchestras were able to travel. With increasing technological advancements in the 1990s, artists began to work on the development of customized software that would allow advanced spatial trajectories to be performed, as well as decreasing the amount of hardware necessary.

In an attempt to counter some of the issues with traveling, the M2 diffusion system (Moore, Moore, and Mooney 2004) was developed in the mid 2000s with a focus on the capability of being compactable and robust enough to take on a plane. The Motu24i/o interface and relatively small fader console can be packed into a flight case for traveling. The flight case then opens up at one side allowing all required connections to be made without removing the hardware from the flight case.

One of the first significant attempts at making certain parts of a diffusion system more easily compatible with others, was to release the software that drives a system. As mentioned for the most part spatialisation systems, including their software, are custom-built for an institution. Once the systems were being driven by custom-built software it became very easy to release and share the tools they had created with the wider international diffusion community. Birmingham released BEASTMulch, a system which allows control of multi-channel sound diffusion with a range of spatialisation algorithms and an ability to communicate with standard MIDI controllers as well as custom OSC fader boards (Harrison and Wilson 2010). This system and others like it meant that institutions were able to incorporate the software into use with their own speaker configurations.

Significant developments in spatialisation algorithms such as Vector Base Amplitude Panning (Pulkki 1997), Ambisonics (Gerzon 1977) and Wave Field Synthesis have also encouraged an increased modularity in diffusion systems. These techniques for spatialisation are featured in the more traditional diffusion systems previously mentioned, as well as range of new diffusion systems including the SoundScape Renderer (Bredies et al. 2006), ReSound (Mooney and Moore 2008), and The Allosphere (Hollerer, Kuchera-Morin, and Amatriain 2007).

A number of new spatialisation systems have recognized a need for published open protocol across spatialisation systems. One common movement is towards the spatial scene descriptors. The SpatDIF (Spatial Sound Description Interchange Format) (Peters 2008) implemented in the HOLO-Edit system (Peters et al. 2009) proposes a format for scene description and also outlines a modular approach to design of a full spatialisation system.

The SoundScapeRenderer, an example of a user interface driven spatialisation system, also proposes a scene description format the Audio Scene Description Format (ASDF) (Geier et al. 2007).

The 2008 International Computer Music Conference held a panel discussion about the need for a recognized and implemented spatial scene description format proposed and lead by authors of the systems mentioned above (Kendall, Peters, and Geier 2008).

3. MODULARITY

The vast majority of the first author's research has been based around the design and development of new performance interfaces for diffusion practice. In designing new interfaces, one major consideration is how they will be integrated in to the larger diffusion system. One of the design goals for these interfaces is to have them not be system specific so that performers may choose to use the interfaces with any spatialisation systems they have access to. However, even so there is often a need for full systems to be developed for prototypes of new interfaces, observations made whilst developing these systems have influenced the majority of research presented in this paper.

Two user interfaces are currently under development, each of them require different support from the wider system and as such are not fully modular, however the intention is to make the two interfaces as interchangeable as possible whilst still adhering to their specific design goals.

3.1.Case Study 1: Tactile.motion

Figure 1. The tactile.motion graphical user interface. Each coloured circle represents a draggable audio object.

tactile.motion (shown in Figure 1), is an iPad application designed to replace the mixing desk as a user interface for diffusion performance. The original concept was to port the authors touch table application tactile.space (Johnson and Kapur 2013) to iPad, however the advantageous features of the iPad as a user interface have allowed significant development of new features. The feature most relevant to this paper is the incorporation of Apple's Zero Configuration Protocol (known as Bonjour). This allows a host (the computer) to broadcast itself as an available OSC service via Bonjour, and the iPad can present the user with a list of all available services so they only need select the appropriate host from the list.

The main advantage of this type of communication system is that inexperienced performers can connect to the network without any knowledge of the set up procedure. There is no need for input of an IP address or port number, the system handles the passing of data when it resolves the broadcasted service. This communication is outlined diagrammatically in Figure 2.

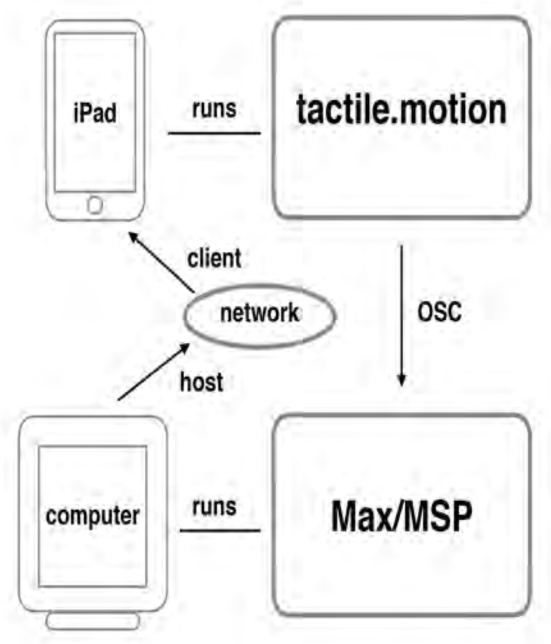

Figure 2. The tactile.motion networking and communication system overview.

On start-up the application links itself to the desired host therefore there is no permanent connection to any specific max Patch or audio software or even any specific computer. This means the application could be used by any spatialisation system.

Other advantages of the system are its stability and remarkably short set-up time. In a concert setting, the computer hosts an ad-hoc network and the performer can select said network from the system preferences on the iPad as they would any Wifi service. This action is very familiar to the majority of performers.

One of the main advantages of the system could also be seen as its disadvantage. The tactile motion application only searches for services broadcast via as available to receive OSC over Bonjour, thus if one wanted to use the application with other software they would need to implement ZeroConf. For most custom built audio software the implementation of a broadcasting service is very simple and can be done with only minor adjustments to examples. In the Max Patch that drives the audio for tactile motion we use the OSCBonjour externals and example patch[1]. For tactile.motion to communicate with another audio driver an understanding of the OSC protocol would also be necessary. This will be discussed in section 4.

3.2. Case Study 2: Chronus_2.0

The Chronus series has slightly different design specifications and as such requires modified communication protocols. The current version, Chronus_2.0 (shown in Figure 3) features a rotating disc with a potentiometer on top. The rotation of the disc provides an angle for phantom source positioning and the fader provides the distance. There are significant technical design features which allow this rotation whilst keeping the potentiometer oriented intuitively, they are outlined in (Johnson, Norris, and Kapur 2014).

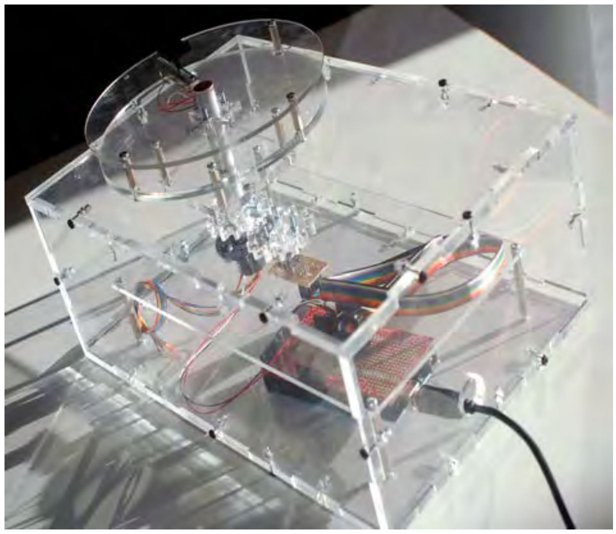

Figure 3. The Chronus_2.0 User Interface

The Arduino Uno microcontroller connects the Chronus_2.0 physical controller to the computer via a USB cable. Therefore there is no need for any wireless communication to occur, thus no broadcasting of Bonjour services. This represents the fundamental difference in communication requirements between the two interfaces. Once the computer receives the data there is an attempt for the interfaces communication streams to remain as similar as possible.

The microcontroller sends serial data with the radial and distance positions. This data is processed in a custom built Processing sketch, the main job of which is to convert the serial to OSC messages. At this point there are a few different options. The OSC messages can be read by a second Processing sketch implementing Pulkki's VBAP principals (Pulkki 1997), this sketch then sends the processed data as weighted gain factors to be unpacked and implemented by a very simple Max patch that drives the audio. This system is outlined in Figure 4.

---

[1]http://recherche.ircam fr/equipes/temps-reel/movement/muller/index.php?entry=entry060616-173626

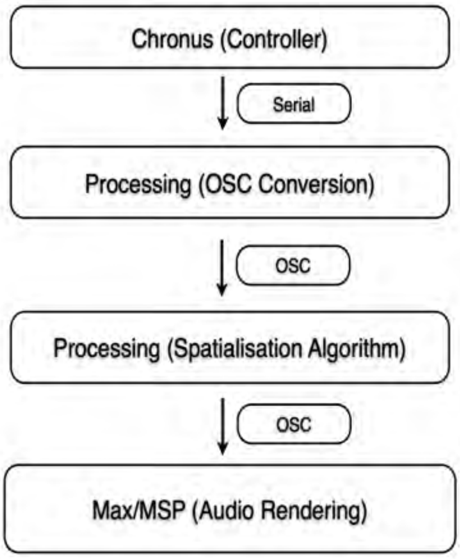

Figure 4. The Chronus spatialisation system separated into modules.

This process best outlines the intended modularity. The system is divided into a controller, a conversion process, an implementation of a real-time spatialisation algorithm and finally the audio rendering. There is a lot of flexibility within this system. For example it is very common for the spatialisation algorithm to be incorporated within the audio rendering application, this can easily be achieved by using a VBAP (or other algorithm) external for Max. Likewise at any point the OSC messages are not confined to Max, but can be received in by any application. The separation of each stage to implement a single modification allows the system to remain extremely modular and means that any single part could be easily incorporated into another spatialisation system. This flexibility is one advantage of such a modular system.

Again the advantage could be seen as the disadvantage. It could be argued that this separation over complicates the systems, and it is true that the same could be achieved with only two modules; the controller and the audio driver. One of the design goals for Chronus_2.0 was to build an interface that would encourage live spatialisation performance outside the diffusion paradigm. It is hoped that performers from the wider live electronic performance community will embrace the Chronus controller and integrate it into their established live set up. The current modularity of the system ensures that this integration could be a smooth one even for performers without experience with spatialisation algorithms. A simple Max for Live patch would also allow communication with Ableton Live as is a common choice of DAW for many live electronic performers.

4. PROTOCOL

The Open Sound Control (OSC) protocol (Wright, Freed, and Momeni 2003) was chosen because of its ease of implementation especially across different platforms and devices, and its flexibility and range. As mentioned, the configuration of the OSC messages sent from one module to the next are extremely important for ensuring the modularity of the system, and allowing for communication between systems. In designing this protocol there has also been an attempt to remain within the most standard input configurations for popular spatialisation algorithm externals so these may be easily incorporated. The messages sent from both interfaces are configured as polar coordinates oriented around the centre of the speaker array. Polar coordinates are a commonly required input data for spatialisation rendering tools including the custom built Processing sketch discussed in section 3.2 and the VBAP external used in the current Max/MSP patch for tactile motion. As both interfaces are intended for multi-channel diffusion they also send an identification tag that directs messages to control the designated audio stem. An example of this protocol is shown in Figure 5.

Protocol
```
object i/f/f
object number/distance/theta
```
Example Message
```
object 3/3.25/2.91
```

Figure 5. Example of protocol output from spatialisation controllers.

Currently there is no 3-dimensional speaker array available for use with these spatialisation systems and thus has not been a focus of this research, however this may be implemented in the future and thus the potential need for a fourth data stream, the azimuth was also considered in designing the OSC protocol. The adaptation to the third dimension would only require that message to be added to the current data, rather than requiring any new protocol design.

In Chronus_2.0 the second module merely converts serial data messages into the OSC messages. If a separate spatialiatision algorithm module is used before the final audio driver, the most common form of messages for it to output is weighted gain factors. For earlier versions of tactile motion this was the case. The OSC messages for weighted gain factors are configured with the stem (or object) number, the speakers number, and the gain as a floating point value. In a similar way to the positional coordinates the gain factors are formatted with the intention of being as universally accepted as possible. By using a custom built VBAP implementation the system can be incorporated into

DAW's without sophisticated multichannel implementations.

Implementing intuitive and universally recognized protocol is an important step to the modularization of spatialisation systems. All systems need to format messages in the same way in order to have modules be interchangeable between systems.

5. CONCLUSIONS AND FUTURE WORK

Both the systems presented in this paper have proved to be intuitive and have been used in performance making use of various modular configurations. The tactile.motion performance interface was featured in a concert at the Wellington City Gallery in September 2013. For this concert the system was configured as is mentioned for Chronus_2.0, without the need for the OSC conversion. The modularity incorporated into both systems meant that this transition was a fast and easy process to make. Interfaces from the Chronus family have also being used in performance with live audio input rather than standard diffusion models in a system configuration much more similar to the proposed tactile.motion set-up. Again the adaptability and modular separation built in to all aspects of the spatialisation system have proved easy to implement and afford a performative expressive range.

At the time of writing there are a number of concerts scheduled to take place featuring both the presented interfaces. A user study is also scheduled to evaluate the tactile.motion user interfaces expressivity as a new performance interfaces for more traditional diffusion concert settings.

Extensive new features are proposed for both interfaces presented and protocol specifications will be constantly reassessed, as new features are developed in order to assure the most intuitive protocol specifications are always in use. Adaptations to the system beyond the user interface are also intended in order to more closely sync the systems and encourage their use from a wider electronic performance community.

This research has exhibited a focus on the development of new performance interfaces for real – time spatialisation. In order to encourage wider use of new interfaces the authors have designed and implemented modularity throughout their spatialisation systems in order to make these user interfaces available to the wider diffusion performance community. It is hoped that the intuitive nature of the implemented OSC protocol will allow other performers to integrate the use of these new performance interfaces into their own spatialisation systems.

Acknowledgements

The authors would like to thank the New Zealand School of Music and the School of Engineering and Computer Science, Victoria University of Wellington for their on-going support of this research.

6. REFERENCES

Bredies, Katharina, Nick Alexander Mann, Jens Ahrens, Matthias Geier, Sascha Spors, and Michael Nischet. 2006. "The Multi-Touch SoundScape Renderer." *In Proceedings of the Working Conference on Advanced Visual Interfaces*. New York, USA.

Clozier, Christian. 2001. "The Gmebaphone Concept and Cybernephone Instrument." *Computer Music Journal MIT Press*. 25 (4). Winter: 81–90.

Desantos,, Sandra, Curtis Roads, and François Bayle. 1997. "Acousmatic Morphology: An Interview with François Bayle." *Computer Music Journal MIT Press* 21 (3): 11–19.

Geier, Matthias, Jens Ahrens, A Mohl, Sascha Spors, J Loh, and Katharina Bredies. 2007. "The SoundScape Renderer: A Versatile Software Framework for Spatial Audio Reproduction." *In Proceedings of WFS Workshop*. Ilmenau, Germany.

Gerzon, Michael. 1977. "Design of Ambisonic Decoders for Multispeaker Surround Sound". *Audio Engineering Society Convention Paper.* New York, USA.

Harrison, Jonty. 1999. "Sound, Space, Sculpture: Some Thoughts on the 'What', 'How' and 'Why' of Sound Diffusion." *Organised Sound, Cambridge University Press* 3 (2): 117–27.

Harrison, Jonty, and Scott Wilson. 2010. "Rethinking the BEAST: Recent Developments in Multichannel Composition at Birmingham ElectroAcoustic Sound Theatre." *Organised Sound, Cambridge University Press* 15 (3): 239–50.

Hollerer, T, J Kuchera-Morin, and X Amatriain. 2007. "The Allosphere: A Large-Scale Immersive Surround-View Instrument." *In Proceedings of the 2007 Workshop on Emerging Displays Technologies.* New York, United States of America: ACM Press.

Johnson, Bridget, and Ajay Kapur. 2013. "Multi-Touch Interfaces For Phantom Source Positioning In Live Sound Diffusion." *In Proceedings of New Interfaces For Musical Expression.* Kaaist, South Korea.

Johnson, Bridget, Michael Norris, and Ajay Kapur. 2014. "The Development of Physical Spatial Controllers." *In Proceedings of New Interfaces For Musical Expression.* London, England.

Kendall, Gary, Nils Peters, and Matthias Geier. 2008. "Towards An Interchange Format For Spatial Audio Scenes." *In Proceedings of International Computer Music Conference*. Belfast, Ireland.

Mooney, James, and D Moore. 2008. "Resound: Open-Source Live Sound Spatialisation." *In Proceedings of International Computer Music Conference*. Belfast, Ireland.

Moore, A, D Moore, and James Mooney. 2004. “M2 Diffusion - The Live Diffusion of Sound in Space.” *In Proceedings of International Computer Music Conference.* Miami, USA.

Peters, Nils. 2008. “Proposing SpatDif - The Spatial Sound Discription Interchange Format.” *In Proceedings of International Computer Music Conference.* Belfast, Ireland.

Peters, Nils, Trond Lossius, Jan Schacher, Charles Bascou, and Timothy Place. 2009. “A Stratified Approach For Sound Spatialization.” *In Proceedings of 6th Sound and Music Computing Conference*. Porto, Portugal.

Pulkki, Ville. 1997. “Virtual Source Positioning Using Vector Base Amplitude Panning.” *Journal of Audio Engineering Society* 45 (6): 456–66.

Wright, Matt, Adrian Freed, and Ali Momeni. 2003. “Open Sound Control: State of the Art 2003.” *In Proceedings of New Interfaces For Musical Expression,* 153–59. Montreal, Canada.

EXPRESSIVE PARAMETERS FOR MUSICAL ROBOTS

Jim Murphy
Victoria University of Wellington
NZSM / School of ECS

Dale A. Carnegie
Victoria University of Wellington
School of Eng. & Comp Sci

Ajay Kapur
Victoria University of Wellington
NZSM / School of ECS

ABSTRACT

Expressive musical robots are those that, through carefully-chosen parameters, afford composers a relatively wide range of compositional options. This paper introduces a number of robots designed with parametric expressivity in mind: robotic chordophones, percussion instruments, and keyboard instruments are presented, and the mechatronic means by which their expressive parameters are realised are discussed.

Such parametrically-dense instruments can pose a challenge to use: a musical robot-specific suite of networking software is presented, intended to allow for easier-to-use complicated musical robots. In addition to the presentation of new musical robots, a history of preceding and contemporary expressive musical robotic systems is provided.

1. INTRODUCTION:

Performance-oriented musical robots are instruments, conceived of as tools for extending the compositional terrain able to be explored by composers, artists, and performing musicians. As musical instruments, they must be capable of interfacing between the artist's musical wishes and the physical world. One way to do this is to present the artist with a vast array of limited-capability devices. A second approach is to create a smaller number of robotic instruments, each able to convey a relatively wide range of musical ideas. The second approach is what is focused upon in this paper: the design, building, and application of expressive musical robots.

Why is expressivity a goal worth striving for in the design of musical robotics? Robotic instruments bring computer music into physical space. Their physicality is what makes them compelling, and the reasons for this are twofold: sonically, the robot shares the space with the listener; visually, the contours and shifts present within the structure of electronically-composed music can be visually coupled with an object. By creating objects with many parameters, each capable of predictable and precise manipulation of the instrument's audio output, those two reasons are accentuated: the visual and aural appeal of musical robots increases along with their mechatronic expressivity.

Expressivity is a word that appears many times in this paper. As a potentially subjective word, a definition of expressivity is in order. In this context, expressivity refers to the ability of a mechatronic musical system to affect a wide range of musical parameters. In essence, according to synthetic instrument builders Xavier Rodet and Christophe Vergez, an instrument capable of "subtle nuances or accentuated features which are under the control of the musician allows for musical expressivity and phrasing" (Rodet, 1999).

As little has been written specifically concerning expressive robotics, a roboticist may turn to other music writing to learn more about the subject. Patrik N. Juslin et al., in "Feedback Learning of Musical Expressivity" (Juslin, 2004), engage in a demystification of musical expressivity in all aspects of music, from compositionally-based to instrument-derived sources. Their work serves to provide useful criteria for any instrument builder concerned with providing musicians and composers with affordances for expressive use. Juslin and his co-authors subdivide musical expression into five categories: piece-related, performer-related, listener-related, context-related, and instrument-related. Juslin states that the factors affecting instrument-related expressivity include "acoustic parameters available, Instrument-specific aspects of timbre, pitch, etc., [and] technical difficulties of the instrument" (Juslin, 2004). The work presented in this paper aims to work from those three parameters (as well as those described by Cornelius Poepel in (Poepel, 2005)), focusing on the creation of musical robots endowed with arrays of acoustic and timbral parameters while remaining usable by human composers, musicians, and artists.

After an overview of existent expressive musical robotic chordophones, keyboard instruments, and percussion systems, four recently completed instruments are presented. The instruments' mechatronic subsystems are focused upon: details are provided concerning the means by which expressive sounds are created through mechanical and electronic systems. These instruments are described, with a focus on their expressive parameters. Finally, the ways in which human composers may interface with the new instruments are discussed, and opportunities for subsequent work are presented.

2. RELATED WORK

While many of the challenges of early mechanical musical instruments (such as those described in (Ord-Hume, 1983)) were likely centred upon a need to allow for the repeatable playback of musical events (with parametrically-rich expressivity taking a back seat to this goal), the latter half of the 19th Century saw a maturation of the technology. With maturing mechanical music systems, inventors began to add additional parameters to their instruments: for example, (*Science*, 1888) describes fully mechanical chordophones capable of tremolo effects.

Figure 1. Trimpin with sound sculpture *Conloninpurple*

The advent of the loudspeaker largely precipitated the end of rapid innovations in mechanical music systems, and it was not until the 1970's that artist/engineers began to re-explore mechanical music, integrating electronic components into their new instruments. The decline of automatic music and the subsequent ascendency of mechatronic music is described in (Murphy, 2012), which introduces sound artists Trimpin (Leitman, 2011) (see Figure 1) and Godfried-Willem Raes as early innovators in expressive musical mechatronics (Maes, 2011). The following subsections detail other recent innovations in expressive musical robotic systems.

2.1. Recent Advances in Robotic Chordophones

Robotic guitars, bass guitars, violins, and other chordophones have ancestors in pre-loudspeaker automatic instruments such as the Banjorchestra and the varieties of Orchestrions popular in the 19th century. These instruments typically augmented existing human-playable instruments by placing them beneath pneumatic plucking mechanisms and string fretting systems. As with other automatic instruments, these devices largely fell out of favour with the advent of the phonograph and, subsequently, the loudspeaker.

Modern innovations in robotic chordophones fall into a two main categories: expressive pitch-shifting techniques and expressive string actuation techniques.

There are two main methods of pitch shifting (also called fretting) on musical robots: arrays of solenoids and linear motion-based approaches. Fretters consisting of arrays of solenoids (such as those employed by Trimpin on his 2007 *Jackbox* or Godfried-Willem Raes on his 2011 Synchochord instrument) function by pinching the string against the guitar's neck; such approaches require a fretboard and are of relatively low resolution and latency. To allow for higher resolution (and more expressive control over an instrument's pitch output), a number of artist/engineers have developed linear motion-based fretting systems. Nicholas Anatol Baginsky's Aglaopheme robot (1992) is an early example of such a system, featuring a linear-motion slide capable of clamping against the string at various points along its travel, changing the string's pitch with a much higher resolution than can discrete solenoid arrays. (Murphy, 2013) describes Baginsky's Aglaopheme in greater detail. A second similar system is Eric Singer's GuitarBot (Singer, 2003), a modular guitar robot capable of even higher pitch expressivity than the Aglaopheme: as shown in Figure 2, each string has a separate linear motion slide, allowing for individual control over each string's pitch. The new works presented below in this paper build upon the continuous linear positioning systems of Baginsky and Singer.

Figure 2. Eric Singer's GuitarBot

Just as two main methods are used to fret the strings, there are two main techniques for picking the strings of robotic chordophones: reciprocating solenoid pickers and motorised pickwheels. Trimpin's JackBox uses both: reciprocating solenoids alternately drag a pick back and forth across a string; a pickwheel consists of a rotary motor (attached to a DC motor or stepper motor) with picks axially mounted to its shaft. As the pickwheel spins, the picks are dragged across the string, actuating it. While neither method allows for a high degree of precise expressive control over the string actuation events, the rotary pickwheel system is built upon in this work because of its inherent high speed and ease of implementation.

(Huang, 2012) and (Ferguson, 2013) present other methods of string actuation: the bowing mechanism shown in (Huang, 2012) is mechatronically complicated and, while potentially quite expressive, outside the scope of this research. The feedback-based system described in (Ferguson, 2013) could be useful to integrate into existing systems, allowing for an additional method of string actuation.

While many existent mechatronic chordophones are variations on those presented here, the works described above are notable in that they strove to increase the expressivity of robotic guitars and basses. The new works presented below are inspired by these existent works, largely focusing on the application of more actuators to the mechatronic subsystems, allowing composers more degrees of expressive freedom.

2.2. Recent Advances in Robotic Keyboard Players

Two parallel trends are evident in studying recent automatically-actuated keyboard instruments: that of assistive augmented pianos and that of fully-automated instruments. Assistive augmented pianos consist of instruments created to be played by a human performer; the performer's real-time playing is then altered through mechatronic means. Conversely, fully-automated instruments are capable of audio playback with no direct human interaction. Occupying a middle ground between assistive and automated keyboard instruments are the Yamaha Disklavier and similar instruments, playable both interactively and autonomously.

While not new (Ord-Hume, 1985), recent mechatronic keyboard playing aids are capable of high degrees of expressivity. Edgar Berdahl et al. in (Berdahl, 2005) present a piano augmented with electromechanical actuators. Subsequently, Michael Fabio created a similar purpose-built instrument designed to be human-played (Fabio, 2007). Most recently, Andrew McPherson's Resonator Piano adds the ability to perform pitch-bend, tremolo, and other untraditional playing techniques on augmented pianos (McPherson, 2010).

A common theme of the above instruments is that they are designed to be played in real time by a human performer. A second avenue of mechatronic instrument development focuses on the creation of keyboard instruments designed to self-actuate. Trimpin's Vorsetzer (Focke, 2011) is an early example of such a mechatronic instrument; Godfried-Willem Raes uses a modified version of the Vorsetzer. Further, the Waseda Team developed an anthropomorphic self-playing robot. While not capable of such complicated augmentations as the aforementioned assistive playing devices, these automated augmentations allow composers a high degree of compositional flexibility over a piece's amplitude, melody, and harmonic content.

A widespread mechatronic instrument capable of a high degree of precision expressivity in the Yamaha Disklavier (and similar instruments), a device that can be viewed as a hybrid between assistive and automated keyboard players. The Disklaiver has a rich history of expressive and experimental compositions, with composers such as Horacio Vaggione (Budon, 2000), David Rosenboom, Terry Riley, Michael Pascal, and Morton Subotnick creating pieces for it. More recently, composer Jeff Aaron Bryant has explored the integration of new musical interface-driven performance with Disklaviers (Bryant, 2012).

New modes of human/instrument interaction in works such as those by Bryant highlight the creative affordances of modern mechatronic keyboard instruments. The new work presented below in this paper seeks to create a similar level of expressivity for a reed-based keyboard instrument.

2.3. Recent Advances in Robotic Percussion

As the most widely-built variety of musical robots, robotic percussion systems are used in most of the discipline's sub-fields. The abundance of robotic percussion instruments is likely due to their relative ease of implementation: potentially mechatronically-simple, robotic drummers can be built with few moving parts, allowing for workers to focus on compositional rather than engineering goals.

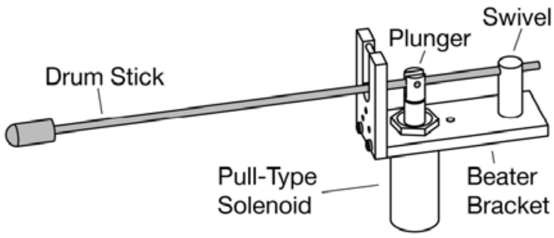

Figure 3. An example of a solenoid-based mechatronic drum beater.

(Murphy, 2012) and (Kapur, 2005) provide a detailed history of robotic percussion systems; this paper will focus briefly on those systems that attempt to further the expressive potential of mechatronic percussion devices.

Many existent robotic percussion systems use solenoid actuators with a single degree of freedom, configured to bring a drumstick or other effector into contact with a drumhead upon actuation of the solenoid. An exemplar device is illustrated in Figure 3. Such drumming systems are used in many contexts, including the authors' robotic ensembles (Kapur, 2011).

While existent solenoid-based drumming systems allow for easy implementation and simple driver electronics, they lack the ability to consistently vary their striking force (Murphy, June 2012) and position above the drum head. Roboticist Gil Weinberg has built a system that addresses these shortcomings (Weinberg, 2006). Other approaches involve using many discrete solenoids (Kapur, 2011b) or (in the case of Trimpin and Godfried-Willem Raes) more complicated driver electronics.

Based upon the state of the art of mechatronic percussion, a need is perceived for an easy-to-build, inexpensive, compact, and easily-interfaced mechatronic drum system. This system is introduced in the following section.

3. NEW EXPRESSIVE MUSICAL ROBOTS

This section presents an overview of the current ensemble of expressive musical robots built by the authors through Victoria University of Wellington's Sonic Engineering program. While consisting of a diverse array of instruments, this ensemble was constructed with three common goals: to build robotic instruments that are more expressive than typical existent instruments, to provide composers with relatively means of interfacing with the robots, and to create them using inexpensive manufacturing techniques. This section presents five recently-completed instruments: two chordophones, a harmonium, and a mechatronic drum beater all designed to fulfil the above goals. The means by which the robots fulfil the goals are described, and brief overviews of the robots are provided.

3.1. Expressive Robotic Chordophones

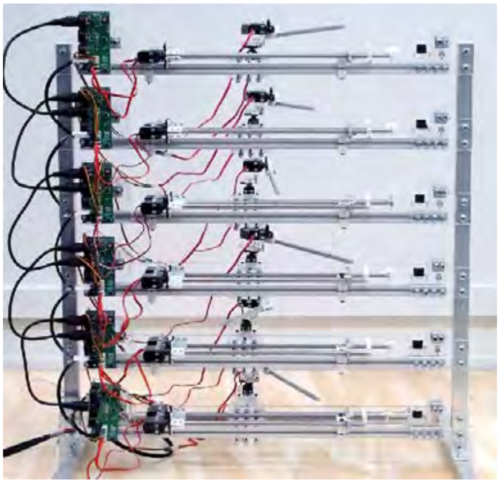

Figure 4. Swivel, a six-stringed robotic slide guitar.

In an effort to build expressive chordophones, a robotic slide guitar and bass guitar have been built. Both the slide guitar, dubbed Swivel, and the bass guitar, dubbed MechBass, are modular instruments: as shown in Figures 4 and 5, MechBass and Swivel consist of arrays of string modules. These modules are independently-addressable and allow each string to be composed for individually.

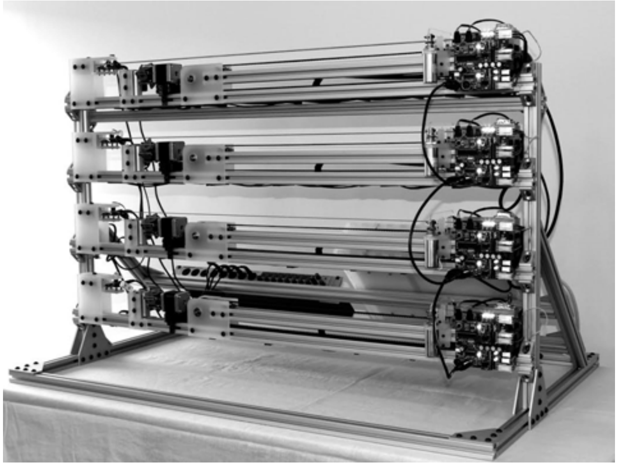

Figure 5. MechBass, a mechatronic bass guitar.

Both MechBass and Swivel 2 are assemblies of subsystems; each subsystem is designed to allow for a greater amount of compositional expressivity than is normal in most existent mechatronic chordophones.

To alter the pitch of a module's string, Swivel 2 and MechBass employ fretter mechanisms. After examining existing mechatronic chordophones, solenoid array fretting systems (described above in Section 2.1) were rejected in favour of motorised positioning systems. These positioning systems were chosen due to their ability to be used for vibrato, pitch bend, and alternative tuning schemes: such techniques extend the compositional expressivity of the instruments.

MechBass uses a fretter system similar to Eric Singer's GuitarBot (Singer, 2003), but with small changes implemented to increase the compositional flexibility and performance of the fretter. Where GuitarBot uses a linear motion slide permanently in contact with the string, the fretter system on MechBass can be engaged and disengaged from the string with the aid of solenoid-actuated clamps: by disengaging from the string, the fretter allows composers to choose whether a pitchshift event is a pitchbend or discrete step.

After the development of MechBass, it was noted that the use of a positioning device for pitch shift events introduced some degree of latency between fretting events: in the worst case, the fretter trolley must move the full length of the guitar's neck, introducing some delay into the fretting action. A fretter capable of very fast repositioning is therefore advantageous. It was a goal in the development of Swivel 2 to build such a fretter. Swivel 2's fretter, shown in Figure 6, is a rotary arm mounted on each module in such a manner that it can reach each position on the string at different motor shaft angles. Upon reaching the desired angle, the arm is pressed against the string by a second motor, changing the string's vibratory length.

The fretting mechanisms on both Swivel 2 and MechBass allow for further parametric control over fretting events than is possible with prior mechatronic chordophones. As such, they fulfil the initial goals of building mechatronic instruments endowed with a larger number of expressive parameters than are present in previous mechatronic instruments.

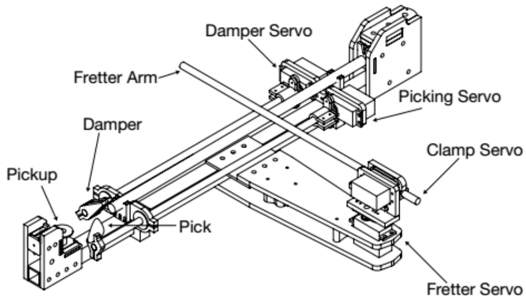

Figure 6. A drawing of a Swivel 2 module.

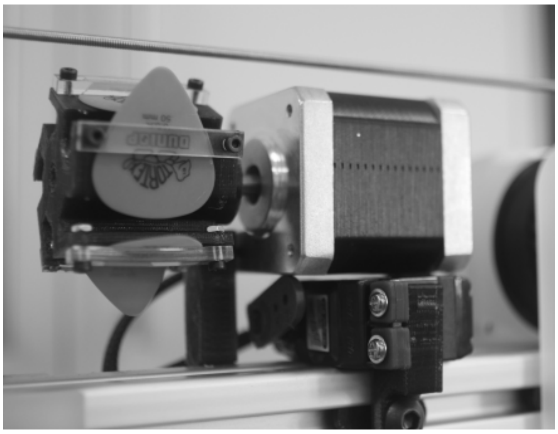

Figure 7. The adjustable-height pickwheel used on each MechBass module.

The second subsystem designed with an intention of enhanced expressivity for MechBass and Swivel 2 is the robots' picking mechanism. While Swivel 2 uses a reciprocating pickwheel similar to GuitarBot and other pickwheel-based systems, MechBass uses an enhanced pickwheel, designed to afford composers an additional parameter of control.

The pickwheel system on each module of MechBass can be raised and lowered in response to a composer's commands, changing the picks' proximity to the strings. By playing closer to the string, the pick's power increases, resulting in a louder, more powerful pick event. Softer, less powerful pick events can be achieved by lowering the pickwheel away from the string. To raise and lower the pickwheel, the stepper motor is mounted on a pivoting cam. The cam is connected to a miniature servomotor's shaft: as the servomotor rotates, the pickwheel, shown in Figure 7, pivots about its hinge.

Similarly, the damper mechanisms of both Swivel 2 and MechBass allow for more control than is typical of robotic chordophones. Where most dampers are on/off solenoids, MechBass and Swivel use rotary servomotors, able to vary the amount of force with which they press the string.

To retain compatibility with the authors' other musical robots, Swivel 2 and MechBass respond to MIDI messages from a host device. Each of the MechBass and Swivel 2 modules is assigned a distinct MIDI channel, and each actuator subassembly responds to distinct MIDI commands. The commands' parameters are converted by the modules' microcontrollers to output voltages to interface with PCB assembly-mounted motor drivers.

3.2. An Expressive Robotic Harmonium

Figure 8. Kritaanjli, a mechatronically-augmented harmonium.

In building a new mechatronic keyboard-playing instrument, a portable desktop instrument endowed with the most expressive parameters of the existent surveyed mechatronic keyboard instruments was desired. As discussed in Section 2.2, contemporary augmented pianos are capable of tremolo (volume modulation) and extended polyphony. These two parameters were chosen to be added to a harmonium, an instrument used abundantly in Indian folk and devotional music (a genre explored by the second author).

Like MechBass and Swivel 2, Kritaanjli is best viewed as an assemblage of mechatronic systems, positioned around a musical instrument to allow for computer-controlled actuation events. Kritaanjli is made up of two such assemblies: a keyboard playing mechanism and a bellows-pumping system. The remainder of this subsection discusses the means by which both subsections are constructed to afford compositional expressivity.

To allow for the realisation of harmonically-complicated compositions, a parallel approach is used by Kritaanjli's keyboard playing system: above each key is a linear solenoid actuator, switched by a power MOSFET. A sufficiently large DC power supply is used to allow for all 44 keys to be depressed simultaneously. Further, the solenoids may be cycled quite quickly, allowing for the playback of new extended technique compositions.

Harmoniums produce sound whose loudness corresponds to the rate at which the instrument's bellows are pumped. Kritaanjli is capable of variable-loudness playing by varying the speed with which its bellows are automatically pumped. The automatic pumping mechanism consists of a DC gearmotor driving a crankshaft. The shaft's reciprocal motion is connected to a hinge on the instrument's bellows, allowing it to be moved back and forth in response to the crankshaft's displacement. To achieve expressive volume changes, the motor's speed can be varied.

Like Swivel 2 and MechBass, Kritaanjli responds to MIDI commands sent from a host computer. The keyboard-playing solenoids are switched on in response to NoteOn events and disabled upon the arrival of NoteOff commands. The bellows-pumping speed is adjusted by varying the value of a MIDI Control Change (CC) 7 message.

The combination of full-keyboard polyphony and adjustable output dynamics allows for Kritaanjli to be used in a compositionally-expressive manner: by providing composers with a relatively large number of adjustable parameters, Kritaanjli, like the other mechatronic instruments described in this paper, allows composers increased freedom over their compositional decisions.

3.3. A Parametrically-Rich Mechatronic Drum Player

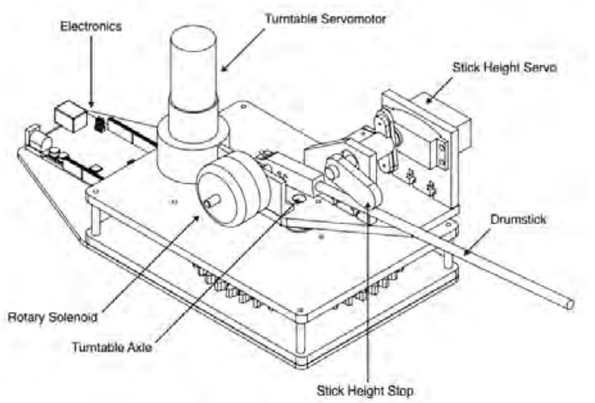

Figure 9. A drawing of Nudge, illustrating its rotary drumstick-beating mechanism and adjust height stop.

In creating an ensemble of musical robotics designed to allow for increased compositional expressivity, it was deemed important to include in the ensemble a mechanism capable of striking drums or other percussive objects. The resulting mechatronic instrument, Nudge (a drawing of which is shown in Figure 9), is "parametrically-rich," allowing composers more control over the percussion system's parameters than is normal for existent mechatronic drum beaters.

Like the majority of existent drum beaters, Nudge uses a solenoid to rapidly bring a drumstick into contact with a drum or other percussive object. Other actuator types were rejected due to low speed or high actuation noise. While most pre-existing mechatronic drum beaters are limited to the single solenoid actuator pre-positioned above a drum head, Nudge provides two more parameters to allow for increased compositional expressivity: Nudge's drumstick may be rotated above one or more percussive objects and its drumstick's at-rest height can be adjusted on the fly.

By being able to rotate its attached drumstick above the drum head, Nudge allows composers to select which part of an object will be hit: for example, Nudge can engage in rimshots at the start of a bar, before repositioning and hitting the centre of the drum at the end of the bar. While (Weinberg, 2006) allows for repositioning of the drumstick, Nudge accomplishes such positioning with a greatly simplified mechanical system, requiring no linear motion hardware.

A recurring problem with many mechatronic drum systems (particularly those based upon solenoid actuators) is that of drumstick positioning in relation to the drum head. If the drumstick's at-rest position is too far from the drum head, rapid, soft percussive events will not move the drumstick's solenoid sufficiently to bring it into contact with the drum. Conversely, if mounted too close to the drumhead, the drumstick will be unable to gain sufficient inertia to allow for powerful, loud strike events. To address this problem, Nudge's drumstick's at-rest position can be varied by a servo-attached cam mechanism. As the cam rotates, the servo pushes the at-rest drumstick from a position high above the drum head (suitable for slower, louder strikes) to one quite close to the drum head (suitable for rapid rolls and quiet strike events).

Like Kritaanjli, Swivel, and MechBass, Nudge communicates with a host device with MIDI. NoteOn and NoteOff commands enable and disable the drum-beater solenoid. A closed-loop control scheme is used to rotate the turntable: a MIDI pitch bend command serves as a setpoint, and an internal PI controller uses readings from the turntable's DC servomotor to reach the setpoint. Drumstick stop height is adjusted in response to MIDI CC 7 commands.

As with the other robots described in this paper, Nudge allows composers to expressively realise compositional ideas that would have been difficult with traditional mechatronic drum beaters (such as those described in (Kapur, 2005)). By allowing the drum-striking solenoid to be flexibly positioned above the drum head, a single Nudge drum player takes the place of a large number of static drum beaters.

Figure 10. Nudge, positioned to strike a frame drum.

4. INTERFACING WITH EXPRESSIVE MUSICAL ROBOTS

The large number of user-controllable parameters made available on MechBass, Swivel 2, Kritaanjli, and Nudge present a potential problem: as the number of parameters increases, the complexity of interfacing with an instrument increases (a topic further explored in (Dobrian, 2006)). To facilitate not only the robots' parametric density but also their expressive potential, Tangle, a robot networking software suite, was designed and built with the aid of VUW student Paul Mathews.

Tangle, implemented in the ChucK programming language (Wang, 2008), allows for a client computer to connect to a server to which are attached any number of musical robotic devices. While the robots may have a plethora of interfacing options, the server maps them to a device-agnostic interface, presenting the client with an array of similar-looking instruments: where two robots might require different low-level messages, Tangle allows Clients to interact with them at a higher level, allowing composers unfamiliar with the robots to more rapidly begin exploring their expressive parameters.

Further, Tangle allows instrument-specific "macros" to be programmed, allowing for a single message from a client computer to instigate a potentially complicated sequence of actions on the server.

While a detailed discussion of Tangle's architecture lies outside the scope of this overview, the following paragraphs detail three exemplar features of Tangle, deemed representative of the open-ended nature of software-based augmentation able to be performed on Tangle-connected robots.

Further information about Tangle, including detailed descriptive documentation and source code, may be found at https://github.com/PFCM/Tangle.

5. CONCLUSIONS

A fast-growing field, mechatronic and robotic music systems are in need of increased expressivity. This expressivity is best achieved through two means: enhanced parametric control of the mechatronic systems and improved means of interfacing with them.

The work presented in this paper endeavours to improve both: new systems with increased degrees of actuator freedom are presented, alongside a software suite intended to allow for increased ease of interfacing with them.

Current work on the mechatronic systems focuses on valuating composers' experiences with them: user studies are in progress wherein composers are asked to write pieces for the robots that explore their expressive parameters; the composers are asked to evaluate their experiences with the new systems.

While significant work still lies ahead in both the advancement of parametric complexity of the robots and enhanced ease-of-interface between the robots and users, the work presented in this paper advances the state of the art, pointing toward a future of expressive and intuitive mechatronic instruments.

6. CITATIONS

E. Berdahl, S. Backer, and J. O. S. III, “If I had a hammer: Design and theory of an electromagnetically-prepared piano,” in Proceedings of the 2005 International Computer Music Conference, 2005.

J. A. Bryant, “Non-affirming strategies in digital correlation,” Master’s thesis, California Institute of the Arts, Valencia, California, 2012.

O. Budon and H. Vaggione, “Composing with objects, networks, and time scales: An interview with horacio vaggione,” Computer Music Journal, vol. 24, no. 3, pp. 9–22, 2000.

C. Dobrian and D. Koppelman, “The e in nime: Musical expression with new computer interfaces,” in Proceedings of the 2006 Conference of New Interfaces for Musical Expression, (Paris).

M. A. Fabio, “The Chandelier : an exploration in robotic musical instrument design,” Master’s thesis, Massachusetts Institute of Technology, 2007.

S. Ferguson, et al. “A Corpus-Based Method for Controlling Guitar Feedback” in Proceedings of the 2013 Conference on New Interfaces for Musical Expression (NIME), (Daejeon, Korea), 2013.

A. Focke, Trimpin: Contraptions for Art and Sound. Seattle, Washington: Marquand Books, 1st ed., 2011.

H. Huang, et al. “Automatic Violin Player” in Proceedings of the 2012 World Congress on Intelligent Control and Automation. Beijing, 2012.

P. Juslin et al., *Musical Excellence*, Chapter 13. Oxford University Press, Oxford. 2004.

A. Kapur, "A history of robotic musical instruments," in Proceedings of the 2005 International Computer Music Conference, (Barcelona, Spain), 2005.

A. Kapur, M. Darling, D. Diakopoulos, J. Murphy, J. Hochenbaum, O. Vallis, and C. Bahn, "The machine orchestra: An ensemble of human laptop performers and robotic musical instruments," Computer Music Journal, vol. 35, no. 4, pp. 1–15, 2011.

A. Kapur, M. Darling, J. Murphy, Jordan, D. Diakopoulos, and Trimpin, "The karmetik notomoton: A new breed of musical robot for teaching and performance," in Proceedings of the 2011 Conference on New Interfaces for Musical Expression, (Oslo, Norway), 2011.

S. Leitman, "Trimpin: An interview," Computer Music Journal, vol. 35, no. 4, pp. 12–27, 2011.

T. R. Laura Maes, Godfried-Willem Raes, "The man and machine robot orchestra at logos," Computer Music Journal, vol. 35, no. 4, pp. 28–48, 2011.

A. McPherson and Y. Kim, "Augmenting the acoustic piano with electromagnetic string actuation and continuous key position sensing," in Proceedings of the 2010 Conference on New Interfaces for Musical Expression, (Sydney, Australia), 2010.

J. Murphy, A. Kapur, and D. Carnegie, "Musical robotics in a loudspeaker world: Developments in alternative approaches to localization and spatialization," Leonardo Music Journal, vol. 22, pp. 41–48, December 2012.

J. Murphy, A. Kapur, and D. Carnegie, "Better drumming through calibration: Techniques for pre-performance robotic percussion optimization," in Proceedings of the 2012 Conference on New Interfaces for Musical Expression, (Ann Arbor, Michigan), NIME, June 2012.

J. Murphy, J. McVay, D. A. Carnegie, and A. Kapur, "Designing and building expressive robotic guitars," in Proceedings of the 2013 Conference on New Interfaces for Musical Expression (NIME), (Daejeon, Korea), 2013.

A. W. J. G. Ord-Hume, "Cogs and crotchets: A view of mechanical music," Early Music, vol. 11, pp. 167–171, April 1983.

A. W. Ord-Hume, *Pianola: The History of the Self Playing Piano*. Unwin-Hyman, 1985.

C. Poepel, "On interface expressivity: A player-based study," in Proceedings of the 2005 Conference on New Interfaces for Musical Expression, (Vancouver), 2005.

X. Rodet and C. Vergez, "Nonlinear dynamics in physical models: From basic models to true musical-instrument models," Computer Music Journal, vol. 23, pp. 35–49, Autumn 1999.

"Musical boxes," *Science*, vol. XII, December 1888.

E. Singer, J. Fedderson, and D. Bianciardi, "Lemur guitarbot: Midi robotic string instrument," in Proceedings of the 2003 International Conference on New Interfaces for Musical Expression, (Montreal, Canada), 2003.

G. Wang, The ChucK Audio Programming Language: A Strongly-timed and On-the-fly Envi- ron/mentality. PhD thesis, Princeton University, 2008.

G. Weinberg and S. Driscoll, "Toward robotic musicianship," Computer Music Journal, vol. 30, pp. 28–45, Winter 2006.

FIELD EFFECT: MOVEMENT, SOUND AND SEMBLANCE IN AUDIO/VISUAL COMPOSITION

Dr. Mark Pedersen

Victorian College of the Arts and
Melbourne Conservatorium of Music

University of Melbourne

ABSTRACT

This article takes Brian Massumi's examination of semblance and event as the starting point for building an audio/visual composition framework. *Field Effect,* a specific work arising from this investigation, is discussed and reflections on the composition process are offered. In particular, the problem of mapping from movement to sound in interactive works is considered, and strategies for examining the mapping process through content reduction and the use of non-local linkages are presented.

1. INTRODUCTION

This paper examines the process of creating an audio/visual installation based on human movement, using Brian Massumi's concepts of semblance and event.

Central to Massumi's philosophical framework is Benjamin's concept of *non-sensuous similarity* – the semblance of perception, which is to say, the experience of perceiving without having seen or heard. One example is the experience of having "seen" the extended trajectory of a fleeing mouse, even though you could not have actually seen where the mouse went (Massumi 2011, p. 17). A more sophisticated example draws upon Michel Chion's concept of *audiovision* in film, where, for example, a punch is thrown, and a victim staggers, and yet the film does not show the punch connecting, but only includes the sound of the hit. The audience will claim they saw the punch connect, even though it was never shown – the sound effect is experienced as a semblance of vision (Chion 1994).

Exploring this concept of non-local linkage, Massumi examines the way in which activation contour*s,* or vitality affects, principally experienced in movement, detach themselves from their originating medium, and are experienced in other media, other forms (Massumi 2011, 106-108). The transfer of movement from one billiard ball to another has the same quality as a rear-end car collision, and the same quality can be perceived purely in sound, without the visual component. Despite the medium, the pattern, the rhythm of the movement is what remains.

In postulating the independence of such vitality affects, there are similarities with Scrimshaw's writing on the independence of affect in general (Scrimshaw 2013), and, drawing upon the rhythmanalysis of Lefevbre, a deep connection with sound itself (Lefevbre 2004).

All of which may appear to be purely speculative if not for the neuro-musicological evidence behind embodied sound cognition, which demonstrates the deep linkages that exist between movement, visual perception, and audition (Leman 2008).

Within the embodied sound cognition framework, there is an increasing body of work examining movement, gesture and sound (Godøy 2010; Altavilla et. Al. 2013) which in some respects can be understood as an investigation of the harmony between sound and movement.

It is important at this point to clarify that such a harmony is not based upon representational parallels between symbolic gestures, either in movement or sound, but at the deeper level of a shared activation contour, or more abstractly, a shared affective tonality. A rain dance does not look like or sound like rain, in any literal sense, and only over time does it become symbolically representational of rain – it originates in expressing the body's feeling of rain.

Creating new work, in whatever modality, frequently revolves around breaking through the accumulated symbolism in order to re-engage with the immediacy of direct experience. For example, Merce Cunningham's approach to choreography strips dance of "all representative and emotional elements that might drive movement… to focus on pure movement" (Gill 2011, p. 121).

"The dance is the conversion of the body's movement in space into the Time of its alteration: its speculative translation into a universe of pure bodily becoming. The life of the body unlimited, in a pure experience of its becoming. Semblance of life. Embodied life: pure expression of the body's aliveness. Animatedness as such." (Massumi 2011, p. 141).

Taking this concept of "absolute movement" into music, Massumi suggests that absolute music is "music that fusionally mutually includes in an as its own pure event of expression all factors entering into its relational field, so that upon entering that field they *become* musical forces" (Massumi 2011, p. 145).

The concepts of "absolute movement" and "absolute music" are not to be understood as modes of expression which exist in isolation from all other forms. Conversely, they are are deeply relational – fusional – to the point that they "fold in" all other forms *in the moment of their own unfolding*, rather than rely upon prior knowledge, symbols or supporting text. In this sense, the impulse to create cross-modal works which share a common affective tonality is in itself the quest to express that which can only be expressed in the intersection of those modes, much in the same way that Chion's audiovisual punch does not exist without the specific interplay of vision and sound.

My own work within this framework of cross-modal expression encompasses gestural synthesis and interactive installation as well as visual-music composition (Pedersen 2012, 2013). The challenge in this cross-modal work frequently lies is exposing enough of the relationship between the different forms that a sense of the fusional nature of the work is apparent, while at the same time concealing enough of this same relationship that the work retains interest.

This is precisely the *semblance* at the heart of Massumi's investigation of Benjamin's concept of non-sensuous similarity, where the notion of concealing and revealing truth is central to the process of art making. Benjamin acknowledges the traditional link between beauty and truth, commenting that "what makes these truth-revealing sensuous forms evocative is their beauty, defined in a specific way: as a *harmony* of the sensuous forms involved" (Benjamin 1996, p. 224).

I have previously explored Benjamin's aesthetics with respect to the concepts of beauty and the sublime, and how this may apply to sound design, and found that an element of noise or dissonance is important in opening up a space in which to experience the sublime, as opposed to attempting to represent the sublime through perfection (Pedersen 2013). Likewise, Massumi is quick to distinguish between semblance which is symbolic (gesturing to some abstract truth), and semblance which is "quivering with life", "pure activity of life, abstractly appearing with all the reality of an intensive event," (Massumi 2011, p. 178).

What is most interesting in the crux of Massumi's argument is that these two aspects of semblance are polarities on a spectrum, not discrete dichotomies. Thus harmony still has a role to play. "Quivering life is never symbolic, but it can be forcibly contained in harmony and its symbolic associations. When it is, the harmony trembles with life, but does not release it… The expression of the truth hesitates, unable to complete the evocative retreat into the depths of reference. The *a priori* principle of resemblance waivers. Harmony is no longer guaranteed and authorized by it… Symbolism and metaphor are composed-away… Life quivers all the more intensely perceptually-felt, in a compelling openness of indetermination regained." (Massumi 2011, p. 178 – 179).

In applying Massumi's concepts of affective tonality, and in particular, the concepts of non-local linkage, field effect, semblance, harmony and openness to a new work, I've sought to develop an approach which addresses some of the challenges inherent in creating mappings between movement and sound, seeking the polarity of semblance that is still "quivering with life".

2. SOUND, MOVEMENT AND AFFECTIVE TONALITY

2.1. The composition framework

The works developed in this investigation centre around video recordings of movement by a dancer and the creation of an audio/visual composition derived from this recording[1].

In adopting framework for creating interactive sound and movement works through Massumi's concept of affective tonality, I've sought to use the following principles:

- Enable the activation contour jumping across media in order to create a shared affective tonality
- Employ indirect mapping which embodies the concept of a field effect, where many elements contribute to a whole
- Employ independent audio processes aimed at expressing only their own "occurence" yet containing the trace of external movement

The composition framework consists of three parts:

- A video analysis stage
- An intermediate "field effect" stage
- A surround sound rendering stage

The video analysis stage breaks the original recording into a 3x3 grid and extracts motion centroid and overall quantity of motion data from each cell of the grid.

[1] I gratefully acknowledge Ashlee Bye for permission to use a recording of her original choreography and performance, and the support of AusDance and the Newport Substation in providing the opportunity and the space in which to conduct this work.

The use of video analysis is not new in this kind of sound work, but personally the approach differs from earlier work I have done using the Kinect depth-sensing camera, where the movement analysis and sound synthesis made use of more direct, and also more symbolic, linkages (e.g. left-left hand position controls spectral content, right hand position controls spatialisation). Stepping away from tracking and labeling specific joints and specific gestures to focus purely on the quality and quantity of motion within the frame aligns the technical approach more strongly with the concept of bare activity underpinning the philosophical approach.

The intermediate field effect stage consists of a simulation of a 10 x 10 grid of linked particles which are anchored in a gravity-free 3D world by two central points, but otherwise are free to move. Individual particles are affected by forces generated by kinetic bodies moving through the plane of the grid, guided by the motion parameters extracted from the video analysis stage.

Individual particles collide with the kinetic bodies as they move, and they also collide with each other. The linkage constraints between the particles cause ripples to flow through the grid as a result of the original forces.

The audio rendering stage consists of an ambisonic encoder and decoder, with a 10x10 grid of source points in a horizontal plane. The target speaker system for rendering the work can be any number of speakers, but ideally uses no less than 16 channels in equidistant configuration around the circumference of the listening space. In this work, height information is not directly used, but overhead channels (even if irregular with respect to the horizontal configuration) have, in my experience, proven to aid the sense of immersion experienced by the listeners. A schematic of the overall composition framework is shown in Figure 1.

2.2. The composition

Collision data generated by the intermediate field effect stage is used to drive simple impulse generators connected to the 10x10 ambisonic source grid. This produces a spatialised field of clicks surrounding the listener, acting as an audible trace of the bare activity of the dancer in the original work.

Visualisation of the intermediate field effect stage forms the visual element of the composition. Where a multi-channel visualization is possible, a motiongram of the original movement is also presented.

This sparse approach to the sonic content of the work is a deliberate strategy of “composing away” any material which may be interpreted as symbolic or in any way extraneous to the bare activity of the dancer. In relation to “composing away”, Massumi references the visual artwork of Robert Irwin, who continually stripped away elements of his work in order to find a point where the work became a “non-sensuous intensity of effect” (Massumi 2011, 158-159).

By reducing the sonic content to a mere impulse (a click) there is no consideration of timbre, pitch or envelope. This reduction assists in the investigation of the work as a fusion of movement and sound, in as much as the sound stimulus now only “makes sense” when experienced as audible spatialisation presented in conjunction with the visual trace of movement in field effect stage.

Just as Irwin gradually added additional elements back in to his visual work, timbre, pitch and duration elements can be added once the sense of a common affective tonality – a sense of fusion – has emerged.

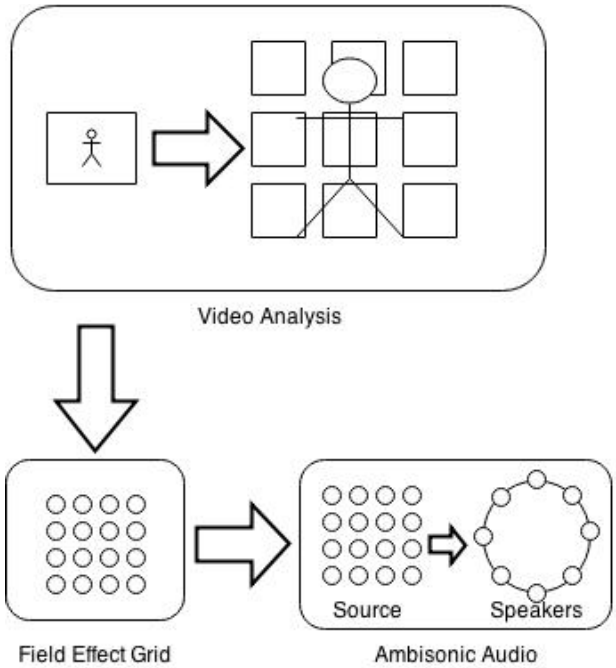

Figure 1. Schematic of the composition framework.

3. REFLECTIONS

In reflecting upon the compositional process, two elements stand out:

- focusing, if that is the right word, on the concept of the non-local linkage, or in other words, the indirect mapping
- reduction of content in order to foreground process

In my previous work with gesturally controlled synthesis, the vast array of possible mappings into complex synthesis space made the selection of only a few parameters seem like an impoverished option on the one hand, while the complexity of developing a more nuanced mapping that sonically reflected the intention of the gesturer was overwhelming. It was difficult to find a productive middle ground.

By stripping away the majority of the available sonic parameters, I was able to focus more clearly on the raw quality of the original movement as it was traced in the audio layer. Reducing the dimensionality of the problem allowed me to evaluate the degree of semblance, of non-sensuous similarity, between the two sensuous modes.

One notable aspect of this reduced dimensionality was the way in which both the granularity of the visual analysis and the degree of spatialisation became important to maintaining the sense of linkage between the visual and audio modes. If the visual analysis is reduced to only a single centroid of motion, and the audio reduced to a mono channel, then the linkage was very weak. The perception is almost one of just a random clicking sound juxtaposed with a random jostling of the field effect grid.

As the granularity in both these dimensions increased, the sense of linkage increased[1]. With more nuanced analysis of the original video, movement qualities in different parts of the visual field translated across into multiple sites of activation within the field effect stage, and consequently as polyphony in the audio stage which matched the intensity envelope of the original choreography. Moving from mono to stereo to ambisonic surround sound, the sense of a shared affective tonality increased. In particular, my subjective experience of the moments of intensity in the choreography began to be matched by my experience of the immersive polyphony of clicks, even without watching the original movement recording. At this point the shared affective tonality that was missing with only a single motion centroid and a single channel of audio, began to emerge.

3.1. Further Work

The work done on this piece is only the beginning of a series of investigations into affective tonality in audio/visual works. In particular, examining how simple additions to the sound synthesis stage impact the composition process and the audience experience is an immediate next step. Using simple processes of subtractive synthesis (filtered white noise, with amplitude envelopes) will begin to slowly introduce elements of timbre, pitch and duration.

Alternate motion capture methodologies are also planned, including muscle activation sensors, inertial motion sensors, and Kinect point-cloud data, in order to see how the fidelity of the motion capture process impacts the overall composition framework.

Finally, further experimentation with the intermediate field effect stage is needed in order to see how an increase in dimensionality (e.g. a 3D cube of connected particles, rather than just a 2D grid) affects the overall process, particularly as the dimensionality of the synthesis space increases.

In addition to these technical aspects, the philosophical aspects of affective tonality and non-sensuous similarity warrant further consideration in the context of audio/visual art work, particularly where, as Robert Irwin does with his visual installation work, attention can be drawn no just to what is being perceived, but also to the process of perceiving. It is in this space that the harmony of our bodies and the sensuous forms of the work can lead to experiences of “liveness” which transcend any individual element, and from which new world-lines upon which to travel can emerge.

[1] This is my own purely subjective experience and not based on any kind of listener study.

4. REFERENCES

Altavilla, A., Caramiaux, B., and Tanaka, A., 2013. 'Towards Gestural Sonic Affordances', in *NIME’13*, May 27-30, 2013, KAIST, Daejeon, Korea.

Benjamin, W. 1996. “On Semblance.” *Selected Writings. Vol 1, 1913-1926,* 223-225. Harvard University Press, Cambridge.

Chion, M. 1994. *Audio-Vision: Sound on Screen*, translated by Claudia Gorbman. Columbia University Press, New York.

Gil, J. 2002. “The Dancer’s Body.” *A Shock to Thought: Expression after Deleuze and Guatari,* Massumi, B. (ed), Routelge, London.

Godøy, R. I., Leman, M., 210. *Music, Gesture, and the Formation of Embodied Meaning*, Routlege, New York.

Lefebvre, H. 2004. *Rhythmanalysis: Space, time and everyday life*, Continuum International Publishing Group.

Leman, M., 2008. *Embodied music cognition and mediation technology*. The MIT Press.

Massumi, B. 2011. *Semblance and Event: Activist Philosophy and the Occurrent Arts*. The MIT Press.

Pedersen, M. and Alsop, R. 2012. “An approach to feature extraction of human movement qualities and its application to sound synthesis,” in *Interactive: Proceedings of the 2012 Australiasian Computer Music Conference*, Brisbane, Australia.

Pedersen, M., 2013. `Sound Labyrinth: Exploration Of The Embodied Sublime Through An Immersive Audio/Visual Installation’, in *Proceedings of the 2013 International Computer Music Conference*, Perth.

Scrimshaw, W. 2013, “Non-cochlear sound: On affect and exteriority”, in M. Thompson & I. Biddle (eds) ‘*Sound, Music, Affect: Theorizing Sonic Experience*’, Bloomsbury Academic.

MEMORY WALK

Camille Robinson
Victorian College of the Arts, University of Melbourne
Faculty of VCA & MCM, 234 St Kilda Rd, Southbank
School of Performing Arts VIC 3006

ABSTRACT

Memory Walk is a sound and video piece that aims to demonstrate and shape the counterpoint and harmony between perception, memory, and imagination in listening. It is one of a group of pieces that encourage the listener to critically engage with listening in the process of listening to them, made as part of the ongoing PhD project *Listening Music*. This paper discusses the aims of the piece, the project of which it forms a part, and the research process used to test its success at achieving its aims.

1. MEMORY WALK

Memory Walk is a video and sound piece that aims to demonstrate and shape the counterpoint and harmony between perception, memory, and imagination in listening, using the framework of a sound walk as the basis for a critique of listening. *Memory Walk* forms part of the PhD project *Listening Music*.

2. LISTENING MUSIC

The project *Listening Music* is about making sound works that encourage the listener to critically engage with listening in the process of listening to them. It stems from the observation that listening as a concept is extremely difficult to define, and that there is a necessity for the sonic arts to participate in the investigation of listening currently underway in the research community, through the medium of sonic artworks themselves. Summed up as a question, the project asks: How can sound works be made that form a critique of listening?

3. RESEARCH METHOD

To answer this question the project takes sound works that aim to form critiques of listening in the process of being listened to, and subjects them to testing with listeners, using an empirical phenomenological approach to gather data on listeners' experiences(Moustakas, 1994). Before listening to the draft work participants are asked to give a definition of listening; they then listen to the work; and after are asked to give their definition of what listening is, and to describe their experience of the work and what they believed it meant.

Analysis of participants' descriptions of their experiences and definitions of listening demonstrate the effect the piece is having on their understanding of listening, and gives evidence of communication of the piece's intent and other meanings that may arise. Following collection and analysis of listeners' responses, I use my observations to make the finished piece. Through the aforementioned process, the hypothesis that a sound work can form a critique of listening is tested, and the broad variation in possible definitions of listening as a concept demonstrated, further justifying the necessity for the work.

4. DEFINITION OF LISTENING

In the day-to-day context, the word "listening" would appear to have a common sense meaning or set of meanings: it refers to the process of experiencing sound. However, in the field of listening studies, there is ongoing controversy as to listening's precise definition, from Glenn's *A Content Analysis of Fifty Definitions of Listening*(1989) to the present day in Wolvin's *Understanding the Listening Process: Rethinking the "One Size Fits All" Model*(2013). Articles like *Development and Validation of the Imhof-Janusik Listening Concepts Inventory*(Imhof and Janusik, 2006) suggest that listening differs as concept and experience from person to person.

Listening studies researchers are experts in understanding listening as a communication process, artists working in sound are experts in understanding listening as an aesthetic process, both have valid contributions to make in the investigation and definition of listening. Sonic artists already have an extensive history of exploring listening through the invention of modes of aesthetic listening.

5. AESTHETIC LISTENINGS

The most well known mode of aesthetic listening is music. The sounds that are accepted as constituting music seem to have been under revision for as long as music has existed, culminating in Cage's assertion that any sound can be listened to *as music*, identifying music as a mode of listening(Cage, 1961, Small, 1998). However, the expectations that make up musical listening consist of more than an acceptable set of sounds, and cannot describe the qualities of all experiences of aesthetic listening.

The limitations of the musical frame on possible sound behaviour and meaning have led artists to invent other modes of aesthetic listening. These include: Pierre

Schaeffer's acousmatic listening(2004), R. Murray Schafer's acoustic ecology(1994), Pauline Oliveros' deep listening(2005), Trevor Wishart's sonic art(1996), and the currently evolving sound art(Voegelin, 2010, LaBelle, 2006, Licht and O'Rourke, 2007, Kim-Cohen, 2009). In *Memory Walk*, the source material relates most closely to acoustic ecology.

6. ACOUSTIC ECOLOGY

Acoustic ecology is the aesthetic study, preservation, and curating of the soundscape: the overall sound of a particular place and time. A sound walk is a curated walk through a soundscape, in which the walker listens observantly to sounds that under other circumstances may go unnoticed. *Memory Walk* uses the sound walk's mode of listening to critique itself, and other more general features of listening. The term used to describe this process of self-critique is immanent critique.

7. IMMANENT CRITIQUE

To do an immanent critique means to judge a concept using its own method of judgement(Sonderegger and Boer, 2011, Merleau-Ponty, 1962, Kant, 1899). In the context of listening, this means to judge listening in the way that listening judges sound. Listening judges sound events constantly; it judges their identity, and judges the need to give them attention and to act upon them. The theoretical term often used to talk about these processes of listening-as-judgement is schemata.

8. LISTENING-AS-JUDGEMENT: SCHEMATA

Schema, or schemata for the plural, is a term used in philosophy, psychology, and AI to describe the sets of rules by which a thing or an idea is understood to be that thing or idea, its meaning, and actions relevant to it in a given context(Piaget, 1977, Rumelhart, 1980, Leman, 1995, Kant, 1899). I understand a clap to be a clap because I recognise its short sharp attack, rapid decay, its source as a set of hands, and contexts I can expect it to apply to such as a concert.

What is particularly useful about the concept of schemata in doing immanent critique is that schemata not only judge their object, they also judge and refine their own rules of understanding that object. Theoretically this process of refinement of the rules of a schema should be able to be used to force listening to judge itself according to the rules by which it relates to the object of its attention, by presenting an object of listening whose attributes come to make listening subvert the schema for listening in a given context.

9. AIMS OF MEMORY WALK: LISTENING IS PERCEIVING *AND* REMEMBERING *AND* IMAGINATION

Memory Walk aims to question the boundaries between perceiving, remembering, and imagination of future events, to argue that these are co-present elements of the experience of listening; to demonstrate and shape the counterpoint and harmony that takes place between perception, memory, and imagination in listening.

The listener begins in their "normal" listening state with an unremarkable stimulus – the video and audio of walking up the stairs. The second audio recording of the walk, with the video of sitting at the table follows. At this point, there should be recognition of similarity with the walk just experienced, and a comparison of details of these recordings, setting memory against perception. A period of relative silence follows where the comparison of memory can peter out. With the recordings of walking down the stairs, imagination of the walk reversed joins remembering and perceiving, completing the counterpoint between the three modes of intentionality.

In terms of schemata, the listener begins with their normal schema for listening to environmental sound; the piece challenges the rules of that schema by altering the relationship of sound to context and memory, creating a sense of confusion. This confusion can then be resolved either by the listener adjusting the schema's rules to account for what is going on, or by the listener coming up with an alternate explanation for their experience.

10. OUTCOMES

Outcomes of the project cannot be finalised yet as only half the sample group has tested *Memory Walk* at time of writing. However, preliminary data shows evidence that it is successful in its aims. Listeners report disorientation in time during the piece, ongoing comparison and analysis of remembered and heard versions of the sound, and when asked for a definition of listening before and after experiencing it, listeners report a heightened awareness of the role of memory and imagination in listening following experiencing the piece. Although I cannot yet make a final judgement of the piece's success, at this stage it appears major revisions will be unnecessary in the final version.

11. CONCLUSIONS

I can draw the preliminary conclusions that this piece successfully demonstrates and shapes, and produces a heightened awareness of the counterpoint and harmony between perception, memory, and imagination in listening. It also gives credence to the hypothesis that a sound work can form a critique of listening.

Participants' comments suggest that I could make some changes to the piece. For example: two participants who listened to the piece at the site of filming reported experiencing extra layers of memory – of their actual walk up the stairs – and suggested making the piece site specific at other showings for a "richer" experience. One participant reported the poor fidelity and portrait format of the video was distracting, while another reported that this heightened their attention to sound while grounding their experience with some extra sense of reality, so the video element needs further consideration. One participant also reported a claustrophobic emotional response to the feeling of wearing layers of headphones; this effect may also warrant further investigation.

This has been a fascinating process through which to make an artwork, and as a research method, it appears fruitful and offers many avenues for future exploration.

12. REFERENCES

Cage, John. 1961. Silence: lectures and writings. Middletown, Conn.: Wesleyan University Press.

Glenn, Ethel C. 1989. "A Content Analysis of Fifty Definitions of Listening." International Listening Association. Journal 3 (1):21.

Imhof, Margarete, and Laura Ann Janusik. 2006. "Development and Validation of the Imhof-Janusik Listening Concepts Inventory to Measure Listening Conceptualization Differences between Cultures." Journal of Intercultural Communication Research 35 (2):79-98.

Kant, Immanuel. 1899. Critique of pure reason. Translated by J. M. D. Meiklejohn. New York, N.Y.: Willey Book Co. Dictionaries

Kim-Cohen, Seth. 2009. In the blink of an ear : towards a non-cochlear sonic art. New York: Continuum.

LaBelle, Brandon. 2006. Background noise : perspectives on sound art. New York: Continuum International.

Leman, Marc. 1995. Music and schema theory : cognitive foundations of systematic musicology, Springer series in information sciences: 31. Berlin; New York: Springer.

Licht, Alan, and Jim O'Rourke. 2007. Sound art : beyond music, between categories. New York, N.Y.: Rizzoli International Publications.

Merleau-Ponty, Maurice. 1962. Phenomenology of perception. Translated by Colin Smith, International library of philosophy and scientific method. New York: Humanities Press.

Moustakas, Clark E. 1994. Phenomenological research methods. Thousand Oaks, Calif.: Sage.

Oliveros, Pauline. 2005. Deep listening : a composer's sound practice. New York: iUniverse.

Piaget, Jean. 1977. The origin of intelligence in the child. Translated by Margaret Cook, Penguin education. Harmondsworth: Penguin.

Rumelhart, David E. 1980. "Schemata: the building blocks of cognition." In Theoretical issues in reading comprehension : perspectives from cognitive psychology, linguistics, artificial intelligence, and education, edited by Rand J. Spiro, Bertram C. Bruce and William F. Brewer. Hillsdale, N.J.: L. Erlbaum Associates.

Schaeffer, Pierre. 2004. "Acousmatics." In Audio culture : readings in modern music, edited by Christoph Cox and Daniel Warner. New York: Continuum.

Schafer, R. Murray. 1994. The soundscape : our sonic environment and the tuning of the world. Rochester, Vt.: Destiny Books ; [United States] : Distributed to the book trade in the United States by American International Distribution Corp.

Small, Christopher. 1998. Musicking : the meanings of performing and listening, Music/culture. Hanover: University Press of New England.

Sonderegger, Ruth, and Karin de Boer, eds. 2011. Conceptions of Critique in Modern and Contemporary Philosophy. Basingstoke: Palgrave MacMillan.

Voegelin, Salomé. 2010. Listening to noise and silence : towards a philosophy of sound art. New York: Continuum.

Wishart, Trevor. 1996. On sonic art. Edited by Simon Emmerson, Contemporary music studies: v. 12. Amsterdam: Harwood Academic Publishers.

Wolvin, Andrew. 2013. "Understanding the Listening Process: Rethinking the "One Size Fits All" Model." International Journal of Listening 27 (2):104-106.

SOPRANO SAXOPHONE AND iPSi - TOWARDS A REFINED PERFORMANCE PRACTICE

James Savage
Faculty of VCA and MCM
University of Melbourne
jsava@student.unimelb.edu.au

ABSTRACT

This paper explores the downstream process of developing the techniques and repertoire required to carry a new extended instrument, the soprano saxophone and iPSi, beyond its initial "novel" status and towards the goal of a truly refined performance practice.
The purpose of iPSi is to facilitate chromatic control of a polyphonic pitch shifting plugin to produce real-time harmony with a single-line instrument. Challenges around setup and the reinforcement of acoustic and processed, amplified sound in a performance or rehearsal environment are addressed with specific reference to the soprano saxophone, although solutions could equally be applied to other instruments in the woodwind family. The paper also traces the development of additional mechanisms of control and refinement of the physical hardware of the interface.

1. INTRODUCTION

New examples of hybrid, meta-, hyper- and extended instruments based around acoustic brass and woodwind instruments surface each year at conferences such as NIME and ICMC. A feature demonstration and discussion of potential musical applications is often given, but there is little evidence that many of these instruments subsequently go on to be fully explored through new bodies of musical work.

In 2012 I designed and built the *harmonic table pitch shifter*, now known as the *isomorphic pitch-shifting interface* and referred to as *iPSi*. A seated instrumental performer's feet play the isomorphic keyboard surface. When the buttons, representing chromatic intervals over a transposable 2.5 octave range are pressed, MIDI output is produced to control pitch-shifting software. The shifted pitches blend with the live instrumental performance to produce harmonies in real time.

Examples of writing on the process of developing material for a new extended or hybrid instrument are rare in comparison to technical papers regarding their development, construction and features. The overarching purpose of my practice-based research is to report, from the perspective of an improvising performer/developer, on the downstream process of taking my extended instrument, the soprano saxophone and iPSi, beyond 'novel' status and developing refined pieces for solo performance. This paper discusses approaches taken to refine the electronics, control and sound production of a new extended instrument to address issues emerging during exploration of the instrument and development of early repertoire.

1.1. Brass, Woodwind and Electronics

Brass and woodwind practitioners have established several distinct approaches to extending their sound palette with technology, the first being the use of conventional effects processors or digital audio plugins to loop and process sound or to add harmonies. Jon Hassell employs fixed-voicing pitch shifting (Hassell 2009) to add single intervals such as perfect fifths or triadic voicings which move in parallel with his live playing, while Arve Henriksen (Henriksen 2008) and Peter Knight (Knight 2012) apply less linear approaches; sampling, looping or applying granular processing to their instrument to create harmonic or textural backgrounds for their melodic playing. Håkon Kornstad creates functional harmony in real time using impressive control over multiphonics and split tones on tenor and bass saxophones; these harmonies are layered and added to with a hardware looper (Kornstad 2009).

Augmented or hybridised acoustic instruments represent a more integrated approach, where the instrument itself is modified to include an array of sensors, switches, potentiometers and other devices in the pursuit of additional avenues of expressivity or intuitive means to control digital audio processing parameters. Hans Leeuw's *Electrumpet* (Leeuw 2009) and Sébastien and Traube's *Electronically-Augmented Saxophone* (2006) follow this train of thought, with both instruments including a means of tracking performer gestures via accelerometer control. Matthew Burtner's *Metasaxophone* (2002) and *SABRe* (Scheisser & Schacher 2012) deeply couple the mechanisms of the acoustic instrument and electronics; the former incorporating force-sensitive resistors on every saxophone key touch, the latter utilising Hall effect sensors to detect the position of bass clarinet keys. Both approaches allow expressive continuous control data to be derived from key states that were previously only "open" or "closed".

Further innovators discard the conventional acoustic instrument altogether and rebuild from the ground up to create still more radical interfaces for electronically enhanced performance. Melbourne's "Bent Leather Band" perform with several such instruments, including the Serpentine-bassoon and Contra-monster (Favilla, Stuart, and Cannon 2006). They create otherworldly free music sonics using a mixture of processed acoustic sound and live 3D spatialisation.

A common element between the three above-mentioned approaches is the individuality and specificity of each design. The extended instruments are usually designed and implemented by the performers themselves, and each setup features a bespoke feature set catering to the desires and tastes of the individual. Systems are rarely transferrable from one instrument to another, for example it is not a simple process to transpose the mechanisms found on the Electrumpet in order to make them suitable for a saxophone-based system. Each instrument or system delivers unique opportunities for musical expression which may be explored and exploited through the development of a refined performance practice and body of musical works.

1.2. The MIDI Footpedal and iPSi

The MIDI footpedal represents a useful interface for live control of audio processing software or hardware for instrumental performers whose hands are tied up in the very action of playing their instrument. More often than not, MIDI footpedals are designed for switching of presets and manipulation of a limited number of continuous control parameters, but in themselves do not represent intuitive devices for actually "playing" musical material. For example, trumpeter and laptop artist Peter Knight often manipulates a pair of foot treadles to control X/Y plugin parameters (Knight 2012). Organ pedal MIDI foot controllers featuring a linear keyboard scale could be more conducive to musical playing; however, the pitch range is limited and there is little scope to produce more than one or two notes at a given time.

Peter Davies' Opal[1] and the Axis 49 by C-thru Music[2] are both hand-played MIDI interfaces that leverage hexagonal isomorphic layouts to provide an easily navigable, musically intuitive key array with benefits such as single-digit triads, transpositionally invariant scale/chord forms and closely-located consonant pitches such as perfect fifths and major or minor thirds. By applying a similar layout to a MIDI footpedal prototype, I created a MIDI footpedal that was better suited to controlling MIDI tone engines and, more importantly, parameters for live digital harmonisation of instrumental performance (Savage 2013).

1.3. Precipitation of Additional Principles

In his framework for the evaluation of digital musical instruments, O'Modhrain (2011 p. 32) defines principles within digital music instrument design as representing "general design goals, expressed in terms which provide little explicit direction as to how they should be implemented". When I first began designing an interface for real-time harmonisation I had certain top-level principles in mind. The interface would allow a single-line instrumentalist to produce chromatic and diatonic harmonies with up to four voices. The device would not have a static pitch reference, but would instead transpose by fixed chromatic intervals with respect to to the pitch of the live instrumental performance. Harmonised voices would have an ADSR envelope which could be adjusted to facilitate a range of musical contexts.

As development continued and solutions for these core challenges were found, certain lower-level principles also became apparent, biased by my lived experience using technology in music performance and pragmatic concerns that emerged as the extended instrument was put into practice. Concerns around the amount of equipment involved, time taken to setup and repeatability of the setup arose. Ergonomics in practice and performance and the ability to communicate on stage while playing in a physically expressive manner were factors that were initially easy to overlook, but in fact would be crucial to developing a refined performance practice. Compatibility with different computer systems became desirable, as did a simplified and straightforward build for future iterations of the interface.

Thumbswitches placed on the acoustic instrument itself had opened up new musical and technical options such as triggering audio *freezing* for sustained accompaniment and a *damper* function to facilitate smooth footpedal transitions. Expanding upon these simple controls could yield additional functionality, but the drive to explore this avenue was tempered by a desire to minimise impact on the acoustic instrument itself and to keep to a protocol that would allow creation or adaptation of controls to other single-line instruments.

In summary, the new low-level principles were: minimal equipment requirements, ease and repeatability of setup, improved ergonomics and freedom of movement, compatibility across computer systems, simplified physical build and additional, adaptable controls.

Section 2.1.1 contains information on a simplified approach to the wiring scheme for the interface and the transition to presenting the device as a class-compliant MIDI interface for enhanced compatibility across

[1] "History | the Shape of Music | Opal Keyboards," Peter Davies, http //www.theshapeofmusic.com/overview.php.
[2] "C-Thru Music Miscellaneous," C-Thru Music, http //www.c-thru-music.com/cgi/?page=info_misc.

various computer systems and audio plugins. The development of a replacement saxophone thumbrest with additional electronic controls is also discussed. Section 2.1.2 describes how the approach taken to improve amplified sound quality lead to improved ergonomics and a lighter equipment loadout.

2. TOWARDS A REFINED PERFORMANCE PRACTICE

Throughout 2012 I developed a number of interface prototypes - then called the harmonic table pitch shifter, *HTPS*. I had composed a limited amount of material for saxophone and HTPS and given a handful of performances. While I was satisfied with the chromatic pitch-shifting harmonisation mechanic and was growing more comfortable with the interface and accompanying mental bandwidth challenge of coordinating hands, feet, breath and balance at once, I did not consider the system to be complete or the musical outcomes refined.

2.1. Technical experimentation

From a technical perspective, there were issues with hardware reliability, and compatibility across different computer systems. I also felt the rudimentary thumbrest-mounted switches I had incorporated could be expanded upon for enhanced control. Finally, miking, monitoring and amplification were variables that could be difficult to control across varying performance venues. A streamlined system that would be more productive in day-to-day use while still offering enhanced options for sound reinforcement was called for.

2.1.1 MIDI Hardware and Control Mechanisms

Early iPSi iterations incorporated a disassembled USB-MIDI interface[1] to transfer MIDI messages from iPSi's Arduino Mega 2560 to the computer system. One concern was that the device required installation of drivers, which might not be updated/supported into the future. The soldered connections to the board were also fiddly and mechanically weak, again a failure risk.

Class-compliant HIDUINO MIDI firmware (Diakopoulos, Dimitri and Kapur 2011) installed on the Arduino's ATmega16U2 presented an elegant solution to this problem, and in testing has proven reliable across a variety of computer systems. A secondary advantage is the possibility of recompiling the firmware from edited source code to assign an appropriate name to the interface for easy identification.

Early prototypes also used custom stripboard PCBs to consolidate pull-down resistors and other components for connection to the 29 buttons. These were eliminated in the third iteration of the interface, which uses a bare minimum of components (see Table 1) and no additional PCBs, instead using Arduino's internal pullup resistors with button logic inverted at the programming stage. This allows an extremely reliable device to be manufactured quickly and for a low cost.

iPSi footpedal interface	
Arduino Mega 2560 R3	1
Sanwa arcade buttons	29
32-way ribbon cable w/ male connector on one end	1
10K "B" rotary potentiometers	8
Female USB "B" jack with male cable (out to Arduino)	1
RJ45 keystone jack w/ short length of Ethernet cable	1
Wire to distribute +5V and ground connections	

Table 1. List of electronic components, latest iPSi build.

Figure 1. Thumbrests; a. lever microswitches only b. RJ45 connector and DIN switches added c. Forza-inspired enlarged rest.

It was desirable to expand the number of available thumb/hand-operated controls, which also required a connecting cable with multiple cores for different signals. For build simplicity direct connections were preferred to solutions involving multiplexing/charlieplexing methods. The saxophone thumb rest is comparatively small, so a small connector was preferable. While DIN connections are sometimes used in these circumstances, connectors and cable are hard to come by and a faulty bespoke cable could make or break a performance. Common RJ45 connectors and cable were ultimately selected, for several attractive qualities: eight cores meant +5V, ground and six control signals could be accommodated, Ethernet patch cables are cheap and easy to come by – plus they have a locking connector and the female receptacle would take up minimal space on a thumbrest-sized PCB. As shown in Figure 1, several thumbrest prototypes were created, the latest based on similar principles to the Ton Kooiman Forza rest for saxophone, but incorporating two rotary potentiometers, two momentary lever microswitches and two DIN toggles. The levers are assigned to audio "freeze" and "damper pedal" functions, while the

[1] Roland UM-1X

potentiometers are mapped to multiple parameters in effects chains and the DIN toggles are mode selectors, not usually operated during active playing.

2.1.2 Microphone techniques and approaches to amplification

During 2012 I explored variations of a typical two-microphone approach for my instrument, the straight soprano saxophone (Savage 2013). One microphone was placed at the bell of the instrument, and another over the keywork in the vicinity of my top hand. The brightness of the bell sound and warmth of the body sound were blended to produce a reasonable representation of the instrument sound in the room. In most cases, small diaphragm condenser microphones were used, although when spill and feedback became an issue in louder environments a dynamic microphone was substituted at the bell. Such spill and feedback was extremely undesirable as unwanted harmonies were created when pitch-shifted material was looped back through the system.

Amplification was initially provided by a 100-watt full range speaker placed behind the playing position, in the manner of a guitar amplifier. I began experimenting with stereo placement of the pitch-shifted signals and moved to a small 2.1 monitoring system designed for digital drums. These speakers were still placed behind the performance position to provide adequate monitoring, but I would setup perpendicular to these speakers - the microphones picked up less of the amplified sound as they were now off-axis with respect to the speakers, as shown in Figure 2.

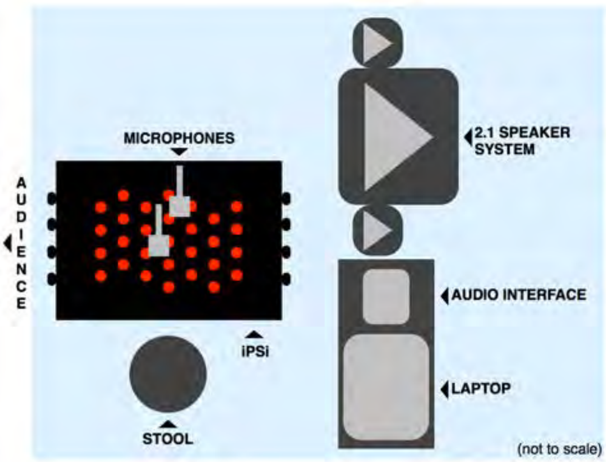

Figure 2. Setup with fixed microphones and 2.1 channel speaker system

The experimentation with microphones and amplification had improved the overall sound and monitoring quality from a performer's perspective but still hadn't eradicated spill and feedback issues. The amount of equipment required had also escalated beyond a point that was practical for day-to-day practice and rehearsals. The combination of the straight soprano saxophone with a fixed two-microphone technique was becoming problematic due to the poor ergonomics of holding this rigid position for long periods of time and the lack of ability to move during performances to communicate with associate performers or simply produce physical expression.

A move to a fully curved soprano saxophone delivered several benefits over the more common straight soprano—an important one being that a single DPA 4099S[1] supercardiod clip-on condenser microphone was effective at capturing the full tone. Freer movement and physical expressivity became possible as a fixed microphone stand was no longer necessary, and I was able to reorientate my performance position to face the audience directly due to the focussed pickup pattern. A woodwind pickup[2] mounted on a spare neck provided a higher level of isolation and this signal could also be used to sidechain gate the clip-on microphone in loud environments, reducing spill and allowing a louder amplified output. There was now sufficient isolation to experiment with highly saturated audio effects on the pitch-shifted signal. Combinations of overdrive and resonant low-pass filtering could be adjusted to blend or contrast with the acoustic saxophone tone and the effects were responsive to live dynamic and tonal variations.

These processed timbres were better reproduced through a guitar amplifier as opposed to a full-range speaker, as the high-frequency tweeters lent an unpleasantly brittle quality to overdriven tones. The lowest note the iPSi system produces with soprano saxophone is Gb_0, which is the instrument's lowest note, Ab_3, transposed down 3 octaves and a major second. While the fundamental frequency of such a note (≅23Hz) sits well outside the optimal range for a guitar amplifier, in practice this unusually low range is not utilised and indeed most playing occurs at pitches of 80Hz and up - above the approximate pitch of the lowest note of a guitar in standard tuning.

These advances in the production of the amplified sound lead to an enhanced feeling of connectedness between the acoustic instrument and the pitch shifted, harmonised output. Placing the amplifier close behind the playing position improved monitoring and could improve intelligibility from an audience perspective, as the acoustic and amplified sounds would come directly from the performance position (see figure 3 below). The pitch shifted signal is necessarily summed to mono before reaching the line-in of the guitar amplifier, but this small compromise has reduced the weight of equipment and setup time dramatically. In a larger performance situation a full stereo signal and dry split of the clip-on microphone signal can still be sent to the

[1] http //www.dpamicrophones.com/en/products.aspx?c=item&category=118&item=24344
[2] http //tjbassoon.com/introducing-the-little-jake-electric-bassoon-pickup/

house PA for sound reinforcement and expanded monitoring possibilities.

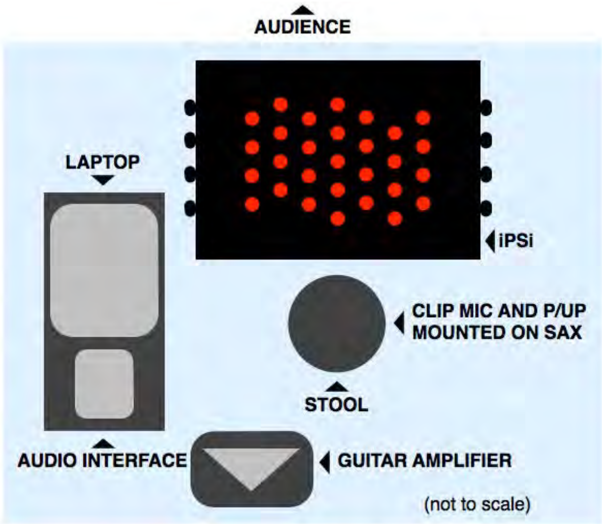

Figure 3. Setup with clip microphone, pickup and guitar amplifier.

2.2. Musical experimentation

An important goal of my practice-based research is the development of refined solo performance pieces for soprano saxophone and the isomorphic pitch shifting interface. Improvisation and the creation of pre-composed material propel the experimental process and cycles of evaluation aim to direct future progress and highlight which approaches deliver successful (and less-successful) musical outcomes.

To date, methods towards developing refined solo performance pieces could be grouped into improvisational, compositional and generative or procedural clusters. Crossover between methods is certainly allowed and there have been cases where a snippet of generative serial material, for example, has been edited and composed and the resultant material used as cells or motifs within an improvisation.

Inspired by a story on John Zorn's approach to composing his prolific "Masada" series, I have lately challenged myself to write a "composition a day" in a small 4-system notebook yielding snippets of 8-12 bars length. This procedure has been carried out in alternating fortnights - one on, one off.

Most commonly, though, video material of improvisation sessions is reviewed and promising motifs or techniques are noted for incorporation into composed segments which may stand alone, or go on to inform or bookend a free improvisation[1]. Therefore improvisation feeds composition and composition feeds improvisation.

While this yields a satisfactory-sounding couplet, reality and lived experience came to reveal that such an approach has a tendency towards stagnation – a closed system. The inclusion of generative or procedural processes seeds composition and challenges technique. I am forced to play the footpedal interface in less natural-feeling and sometimes literally uncomfortable ways. One such algorithm was developed to produce serial 2-part inventions based on stochastic variables. As might be expected, the resultant music varies wildly and is sometimes unplayable without alteration, but nevertheless it succeeds in refreshing my concept of the system's capabilities.

3. CONCLUSION

This paper presents steps towards a refined performance practice for soprano saxophone and iPSi. The process is by no means complete, and further cycles of experimentation and reflection will inevitably lead to further changes to equipment, technique and approach – driving towards the goal of a highly refined extended instrument system and performance pieces exemplifying such a level of refinement.

4. BIBLIOGRAPHY

Arve Henriksen. *Cartography*. ECM 2086, 2008. Compact Disc.

Burtner, Matthew. "Noisegate 67 for Metasaxophone: Composition and Performance Considerations of a New Computer Music Controller." In *Proceedings of the International Conference on New Interfaces for Musical Expression*, 24-29. 2002.

Diakopoulos, Dimitri and Ajay Kapur. "HIDUINO: A firmware for building driverless USB-MIDI devices using the Arduino microcontroller." In *Proceedings of the International Conference on New Interfaces for Musical Expression*, 405-408. 2011.

Favilla, Stuart, and Joanne Cannon. "Children of Grainger: Leather instruments for free music." In *Proceedings of the 2006 conference on New interfaces for musical expression*, pp. 370-375. IRCAM—Centre Pompidou, 2006.

Jon Hassell. *Last night the moon came dropping its clothes in the street*. ECM 2077, 2009. Compact Disc.

Knight, Peter Harold John. "Allotrope- Works for Trumpet, Laptop, Pedals and Guitar Amplifier." In *Proceedings of the International Computer Music Conference*, 567-572. 2012.

Kornstad, Håkon. *Dwell Time*. Jazzland 270-971-0, 2009. Compact Disc.

---

[1] Video documentation of certain iPSi performances, pieces and improvisations can be viewed online at http://www.youtube.com/playlist?list=PLvMDtEPaBCihRZb1zv0X81KbaT5wJAEsz

Leeuw, Hans. “The Electrumpet, a hybrid electro-acoustic instrument.” In *Proceedings of the International Conference on New Interfaces for Musical Expression*, 193-198. 2009.

O'Modhrain, Sile. "A framework for the evaluation of digital musical instruments." *Computer Music Journa*l 35, no. 1 (2011): 28-42.

Peter Knight. *Allotrope.* Listen/Hear Collective, 2012. Digital Audio Recording.

Savage, James. “Developing a Chromatic Interface for Live Digital Harmonisation of Saxophone Performance” In *Proceedings of the International Computer Music Conference*, 321-326. 2013.

Schiesser, Sébastien, and Caroline Traube. "On Making and Playing an Electronically-Augmented Saxophone." In *Proceedings of the International Conference on New interfaces for Musical Expression*, 308-313. 2006.

Schiesser, Sébastien, and Jan Schacher. "SABRe: The Augmented Bass Clarinet." In *Proceedings of the International Conference on New interfaces for Musical Expression*. 2012.

DESIGNING AN INTERACTIVE SONIC MEDIA ENVIRONMENT FOR STROKE REHABILITATION

Diana Siwiak
Victoria University of Wellington
New Zealand School of Music

ABSTRACT

Immersive media environments assist in solving the real-world issues of presenting multimodal information and of responding to user reactions, such as with patients during stroke rehabilitation therapy. Embedding media in the experience creates a new perspective for both the therapist and the patient. It is in this perspective that creative implementation of the arts within technology flourishes. A sonic media environment designed for stroke therapy provides the patient with the ability to produce an accurate trajectory. Based on self-assessment, focus is on the linearity of movement and speed is not crucial. Gradually reducing the amount of real-time media information and promoting a more constructivist learning approach towards therapy, is a cornerstone of this paper.

1. INTRODUCTION

Media arts is an effective medium to assist users, such as school children or rehabilitation patients, in devising individual learning strategies for themselves using constructivist learning techniques, see (Duff *et al*. 2010; Sundaram & Rikakis 2006). These are two of the many approaches for communicating this concept. In an adaptive, mixed reality rehabilitation (AMRR) system, each level of feedback is designed to communicate the appropriate type of movement performance information, while maintaining engagement and understanding through an unfolding, interactive narrative. This knowledge blossoms by synthesizing perception, modeling, interaction, feedback, and sensing into a precariously balanced, mutually beneficial relationship.

This paper discusses the iterative, rapid prototype design of a real-time, customizable generative sound engine and toolkit, including the extensive prior work that fueled this research. As part of a three-year project, the research team set out to prove the meaning, which arises from the interrelationships, connects action to music and to propose changes to a developed system. The generative sound engine, as part of a more complex, multi-modal system architecture, provides an interactive, multi-temporal sonic media environment (SME) for patients recovering from movement restrictions present after stroke. However, only the audio component of the system is addressed. The toolkit includes adaptable features to make adjustments to the soundscape in real-time, thus affording possibilities for a plethora of intricate sound designs.

2. PRIOR WORK

Current research within the field of rehabilitation investigates the best therapeutic strategies for stroke rehabilitation. This includes the use of virtual and augmented reality in order to leverage sensory-motor integration for interactive learning. Embedding media in the experience creates a whole new perspective for both the therapist and the patient. The Mixed Reality Rehabilitation (MRR) Research Lab has been in the process of developing an adaptive, mixed reality rehabilitation (AMRR) system for stroke survivors recovering from right-handed hemi-paresis since 2006. The introductory system is embedded in a hospital. Comparative trials of 11 patients have already been completed by the lab and are currently under assessment by graduate students in the lab. Preliminary results prove the expected outcome, which is an improvement in movement performance through mediated physical therapy versus traditional physical therapy.

Experiential media systems explore the relationship between digital technologies, which can sense human action, and cultural processes, whereby human actions transform media information through the use of these digital technologies. The MRR Research Lab has shown that experiential media systems, or interactive environments, can be used to provide effective feedback and feed-forward information and to promote positive motor learning skills, including retaining and applying these skills. (Duff *et al*. 2010) The immersive SME (in this paper, feed-back and feed-forward information with respect towards audio) is connected to both observable and unobservable movement activities. These encourage active physical and cognitive participation, with the end goal being the learning of generalizable movement strategies. The movement activities and media environments created here are adaptable to a patient's individual ability and progress. This process allows the patient to be challenged physically and cognitively without frustration. The following two sub-sections have been heavily adapted, edited, summarized, or directly quoted from seminal manuscripts written by the MRR Research Lab, in order to lay substantial groundwork for this research.

2.1. Real-Time Sonic Media Environment

An interactive SME (designed to imitate a form of structured music) is a powerful tool for movement training, particularly utilizing its temporal aspects to assist in pacing the speed of the movement activity. (Jaques-Dalcroze 2007) discusses how music helps the brain connect body, space, and action in an intuitive manner. (Dourish 2004) notes that multimodal media compositions, such as films, rely on music to communicate implicit messages and emotional states. The same methodologies are employed in this system in order to encourage natural interaction with the patient. The harmonic progressions commonly found in popular tonal music can encourage a feeling of forward movement, giving a natural affordance for use within an interactive training environment or rehabilitation system to convey positive feedback.

The first version of the interactive sound engine, embedded within the hospital scenario, capitalizes on the inherent powers of music, such as harmonic structure and rhythm, by mapping movement activities onto the features of an abstract sonic environment. "The resulting immersive environment has two main goals: a) to encourage the patient to perform the required movement activity and b) to offer to the patient an intuitive way to self-assess their movement performance, understand the cause of error, and develop an improvement strategy," (Duff *et al*. 2010). Spatial and temporal information are not mutually exclusive. However, spatial thresholds hold the key to the harmonic framework of chord progressions. Temporal thresholds predominantly drive the rhythmic framework of note densities.

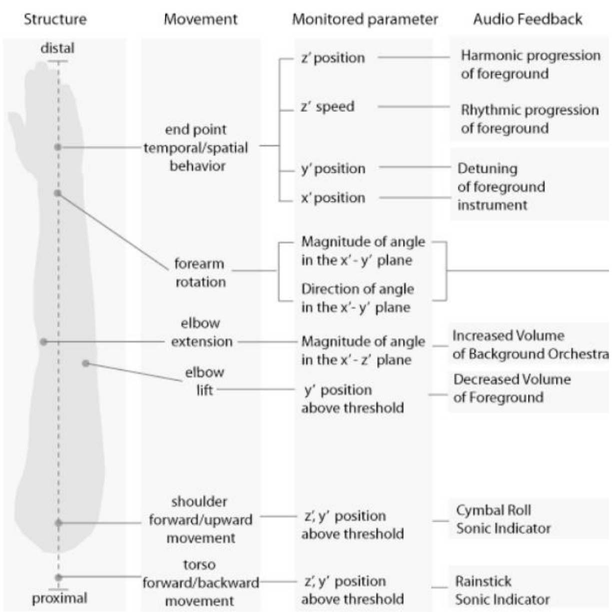

Figure 1. Parallel movement components and feedback mapping

Figure 1, depicts how various movements, monitored parameters, and media environments (both audio and visual) co-exist with one another in the mixed reality rehabilitation system. The patient's arm is laden with several passive motion capture rigid bodies.

Eleven infrared cameras capture these rigid bodies in real-time. The data from the movement of these rigid bodies within the capture space is sent to a processing program. The processing program then provides information to determine various parameters, such as position data and spatial information. The parameters are managed and distributed to the sonic media environments, which create an immersive and interactive media space. The real-time media environment provides explicit and implicit information from which the patient can self-assess his or her movement strategies. It is through this self-assessment that the patient can plan his or her next movement activity and make necessary improvements, creating a cohesive media environment.

This real-time sonic media environment provides prescriptive knowledge about the movement performance. In order to accomplish a quick learning curve towards interpreting the media environments, the audio scene does not change instruments nor chordal progressions, but does sustain a structured composition. This rigid sonic media environment provides an opportunity to expand the possibilities of sound designs, such as in the manipulation of an ocean soundscape, where the patient is affecting the density of waves or bubbles, as opposed to note densities.

2.1.1. Harmonic Content

In the hospital system, harmonic progressions sonify real-time movement activities through both temporal and spatial information. A key signature is stochastically chosen from either a current chord or from one of the adjacent chords within the circle of fifths. This creates sequences of related progressions with smooth key changes between movement activities. At the same time, one of three possible chord progressions, all of which are common harmonic cadences, is randomly chosen. Each progression has four chords and the chords are mapped to thresholds in space, as shown in **Table 1**. Each note is selected by pulling pitches from the chosen chord. The intent is that the patient be intuitively motivated to complete the reaching movement activity so he or she can hear the entire musical phrase performed to completion. It is through this initial basis, set forth by the hospital-based system, that the home-based system (Siwiak *et al*. 2011) was provided with the opportunity to build upon using harmonic content to sonify temporal and spatial information.

Table 1. Harmonic progressions & spatial thresholds

State (normalized velocity)	Prog 1	Prog 2	Prog 3
Resting (0.0-0.1)	I	I	I
Reach (0.1-0.9)	V7/IV (I7)	vi	ii
Grasp (0.9-1.0)	IV	ii	V7
Return (0.9-0.1)	V7	V7	IV

2.1.1. Rhythmic Content

Musical rhythm sonifies the speed of the movement activity performed by the patient. The notes per beat progress from top to bottom through the durations depicted in **Table 2**. As the patient's movement becomes faster, the notes per beat increase. The mapping of spatial information to notes per beat is scaled by a desired maximum movement speed, predetermined by the therapist. If this maximum speed is reached, the notes per beat peak, with individual notes merging into a continuous stream of seemingly obscurely defined notes.

This change from individual notes to a streaming series provides an implicit notification for the desired maximum speed. Through these mappings, the patient is encouraged to perform a smooth acceleration and deceleration pattern. The peak speed occurs in the middle of the reach, so that a smooth rhythmic pattern is heard. The rhythmic information conveyed in the hospital-based system laid further groundwork for iterative design within the home-based system. (Siwiak *et al.* 2011) The patient focuses on successfully completing the movement activity while the music implicitly trains the timing structure.

Table 2. Rhythmic patterns & temporal thresholds

% of velocity with respect to desired velocity peak	Foreground note subdivision (notes per beat)
0-20% (resting)	Quarter notes (1.0)
20-40%	Triplet eighth notes (1.5)
40-60%	Eighth notes (2.0)
60-80%	Quintuplet eighth notes (2.5)
80-100% (streaming)	Triplet sixteenth notes (3.0)

2.1.1. Musical Context

The musical progressions present in the hospital system function to assist in learning a timing structure to help the patient achieve ideal results (i.e. proper and consistent movement performance). The interconnection between training musical rhythm and training movement in time and space has been effectively utilized under the Dalcroze method and other music pedagogy methods (e.g. Kodaly, Orff). It is important to note that, within these pedagogies, the learning of movement and music is subconscious. The complex synchronies and phrasing structures are learned intuitively. Because of this intuitive connection of music to movement activity training, SMEs have been used for several therapeutic purposes, (Ghez *et al.* 2000; Headlee, Koziupa, & Siwiak 2010; Siwiak, Berger & Yang 2009).

2.1.2. Emotional and Cognitive Connections

A patient's emotions play an important role in the success of his or her stroke recovery. Feelings of tension can exacerbate other physiological symptoms and the side effects of stroke, such as tremor or compensation techniques. For this reason, these MRR systems are designed to provide a calming SME and a continued motivation throughout the movement activity. (Gabrielsson & Lindstrom 2001; Seitz 2005)

As described in (Ghez *et al.* 2000), the brain maintains a strong memory for musical constructs. The patient connects successful movement activities with pleasing musical phrases. The memory of the pleasant musical phrase is then used to plan future successful movements. The majority of this type of music-assisted movement activity learning happens subconsciously, similar to the intuitive learning that happens during dance. (Prinz 1990, 1997)

2.2. AMRR: Integrative, Interactive Learning

Abstract feedback mapped to movement activity re-contextualizes a performance of the movement activity into a performance of an interactive SME The patient receives the knowledge of his or her performance and the results of his or her movement activities in terms of his or her interaction with the SME Within this context, the patient is able to intuitively develop evaluative measures of his or her performance on key aspects of movement activities in the SME While the patient is learning to control the SME, he or she is intuitively developing more efficient movements.

Another effective aspect of using abstract feedback is the promotion of active learning. Abstract feedback requires the patient to determine the causality between components of movement activities and individual attributes of the SME The patient must connect how the overall form of the SME reflects the overall form of the movement activities. The metaphorical reconstruction between movement activity and media environment requires active engagement, and eventually results in parallel cognitive and motor learning by the patient. This approach to learning is at the core of constructivist learning theories. (Cirstea & Levin 2007; Gardner 1999; Papert 1980)

By taking a constructivist approach to the system design, these methods decrease the likelihood that the patient develops dependencies characteristic of prescriptive learning, which may not allow for generalization or maintenance past the therapy session. (Schmidt 1991) Therefore, abstract media mappings are consistently applied to key aspects of movement across a variety of movement activities.

3. CURRENT WORK

The adaptive, home-based, interactive sonic media environment (SME) evolved based on the prior work that was completed and assessed in a more comprehensive version of the software, the hospital-based system. As the home-based system strives to be the next generation of an interactive multimodal rehabilitation system, its design, goals, and implementation begin to grow. The goals of this new SME are to 1) reduce the amount of detailed information given in real-time as the movement activity is happening, 2) provide summaries of performance on a per movement activity level and across groups of movement activities, 3) provide a media environment that correlates to a reconfigurable space, and 4) begin to integrate media environments into the patient's physical space. In other words, this new, simpler rehabilitation system endeavors to focus on economy, scalability, and a more long-term engagement. The innovative approach towards gradually reducing the amount of real-time media information, therefore promoting a more constructivist learning approach towards therapy, is the cornerstone of the home-based system when compared to the hospital-based system. (Siwiak *et al.* 2011)

The ecological coupling (Sundaram & Rikakis 2006) between a patient's movement speed and the SME relies on rhythm, so therefore the system maps hand speed to density of notes. Burkholder clarifies this in [**Figure 1**]. During the reaching task, the harmonic intervals and attributes of the sound design are altered accordingly, based on movement performance. The movement activity becomes more consistent in speed as a consequence of the patient retaining motor learning skills by practicing movements. The soundscape displays encouraging and peaceful music. The new sounds designed for the home-based system have been created to allow for changing soundscapes without changing the semantic meaning they convey. The patient does not have to repetitively listen to the same droning sound throughout the training period, but can, instead, interact with waves in an ocean scene or with fallen leaves in a wooded scene. Sound suites such as these each convey similar semantics and can thus relieve the monotony for the patient during therapy.

The research on audio design for the home-based system works towards a more predictive (or feed-forward) model, rather than a reactive (or feedback) model. This is because the reaction to a triggered media stream, such as an audio cue, has inherent delay. This could cause confusion during real-time movement activities if the delay happens to be substantial. This research uses a probabilistic way of analyzing and predicting action to iterate through a predictive model for feed-forward information, thus creating an interactive SME containing both feed-forward and feedback information about movement performance.

3.1. Generative Sonic Content

In order to transition to a constructivist learning approach, the first version of the sound engine (from the hospital-based system) needed a redesign. The motion capture and analysis engines in the home-based system create a large amount of processing overhead. In order to reduce this processing, significant improvements within the interactive audio environment were made to provide a more efficient system. The processing now occurs at the level of the operating system, which has reduced the necessity for proprietary (and processor-heavy) software. It also opened the door to create custom software that is more efficient and effective. This redesign also created a more integrative interaction between the various system components.

With the background, training, and experience in digital audio signal processing and sound design, the proposed changes to the original sound engine included developing a real-time customizable, physically modeled sound design toolkit. The C++ library used to develop this toolkit is The Synthesis ToolKit in C++ (STK). "STK is a set of open source audio signal processing and algorithmic synthesis classes written in the C++ programming language. STK was designed to facilitate rapid development of music synthesis and audio processing software, with an emphasis on cross-platform functionality, real-time control, ease of use, and educational example code."[1] It is with this library through which a custom software wrapper could be implemented within the larger system architecture.

Using synthesized sounds in the home-based system allows for detailed manipulation of each sound. This action opens the auditory landscape while significantly reducing the overhead of computer processing and storage requirements when compared to the hospital-based system, which uses a dedicated sample-based software synthesizer application, Kontakt. Some popular audio processing programs such as Max/MSP, Pd, and ChucK host proprietary wrappers for STK. However these applications are all processor-heavy and have the innate characteristic of significant delay from an action to a reaction. Since the infrastructure of the home-based system is a proprietary Mac-based program, it afforded the opportunity to create an Audio Unit wrapper for the STK and a custom integrated Audio Unit host.

In order to ease the transition from the hospital-based scenario (where the primary instrument is a marimba), the home-based scenario utilized the banded waveguide modeling class included in the STK. Essl and Cook describe in their documentation that "this class uses banded waveguide techniques to model a variety of sounds, including bowed bars, glasses, and bowls." (Essl & Cook 1999) The sonic content of a plucked or bowed wooden bar is not unlike the sonic content of a marimba. The parameters and presets within the banded waveguide class are easily manipulated to create a marimba sound.

3.2. Multi-Layered Sonic Media Environment

The research team employed a different approach for designing compositions consistent with some standard movements for a stroke survivor. Once the compositions have been determined and designed, the proximate data streams are analyzed and a mapping function within this interaction space is then determined. This afforded the opportunity to be creative with the music compositions and to not be constrained by mapping functions or streaming data. A higher-level narrative emerged from this unbounded design scheme, where the various interaction layers (real-time, post-activity, and post-set) could be simultaneously individual, yet connected.

[1] Synthesis Toolkit. CCRMA. https://ccrma.stanford.edu/software/stk

In the hospital-based system, there are two layers of interactive media environments: real-time and post-activity. In the home-based system, there are several layers of interactive media environments, including: real-time, post-activity, and post-set (where a set is a group of movement activities without a real-time media environment).

In this paper, the concept of dovetailed layers describes the transitions that occur between prescriptive (real-time movement performance) information and reflective (summary movement performance) information. The information contained within each layer of the SME is dovetailed in order to facilitate the patient's understanding as the system transitions between layers. Because of the flexibility of the methodology, different scenarios within these layers may be developed to suit the needs and preferences of different stroke rehabilitation patients. However, the form and content of each scenario must be able to adapt and accommodate the appropriate type of movement performance information being communicated.

3.2.1. Real-Time Activity Media Environment

The basic concept from the hospital-based system carries over to the home-based system where note density maps to movement activity speed and harmonic changes map to spatial information. The parameters in **Figure 1** have been reduced in the home-based system, especially since the sensing capabilities have been reduced (to encourage and economically feasible system), which means the SME contains a fraction of the original sonic content. The sound still encourages the patient to move with a smooth, natural speed, but less information is given regarding body function movements (such as elbow openness, which in **Figure 1**, is signified by orchestra volume). The main differences in the sound design are the back-end system architecture, the communication protocols between the audio engine and the system engine, and, most importantly, the soundscapes now possible with the implementation of Audio Unit wrappers for STK. To date, a handful of wrappers (one for each class in the STK) have already been designed and implemented for use within the SME for the home-based system.

3.2.2. Post-Set Activity Media Environment

The sound design concepts for the post-set layer build upon the real-time SME The post-set layer creates an affective summary, or sonic images. Compositional constraints based on well-known paradigms associate speed to rhythm, as previously mentioned in Section 2. The design concept applied herein ventures to progress from musical composition to data mapping to motion capture data stream, rather than the opposite approach. The scheme includes individual movement activities, where one-to-one mappings of data to sonic information is displayed, and the larger narrative picture, where comprehensive scenarios provide a less prescriptive and more constructivist perspective on media environment to movement activity - or, coupling feedback to action, which is contrary to concepts discussed in (Burkholder, Grout, & Palisca 2010).

Within each sonic image, both reactive and predictive models of performance are conveyed through dual-layered compositions, with the lower voice aurally depicting an ideal movement strategy and the upper voice aurally depicting deviation from the ideal. The ultimate goal is to be rewarded with the voices accompanying each other in a pleasing, non-dissonant composition. If after a set, the patient has consistently performed ballistic movements, the upper voice would contain short, tin-sounding notes, dissonant intervals, and fast-moving density of notes when compared to the lower voice, which would contain the sound of a gentle, rolling ocean. The two voices would not match in speed or in timbre, and the patient would then be able to assess performance and plan future tasks accordingly.

In the current version of the system, there are 4 sonic images, each with sequences containing 8 "polychords" (in this paper, polychords describe groups of notes). Polychords were designed so that there is not necessarily a prominent progression, unlike the well-known harmonic progressions in the hospital-based system. This affords the opportunity for variance without the need for immediate familiarity, in order to gather proper meaning from the SME The polychords within each sequence are paired together, providing a higher-level composition over groups of movement activities, or sets. To toggle between pairs of polychords, a spatial threshold of at least halfway must be attained. This carries over from the concepts learned from and tested within the hospital-based system. (Duff *et al.* 2010) Rather than sonifying specific and prescribed parameters over an individual movement activity, depicted in **Figure 1**, a more generalized communication of movement performance can be attained. This gives us the opportunity to decrease the amount of real-time media information, to allow the patient to practice groups of efficient movement activities, and to create an interactive media narrative.

3.2.2.1 Sonic Image Tags

An "ideal" sequence signifies a proper movement for a group of activities. As shown in **Figure 2**, this composition has consistent pedal tones (in this case, C major without the third) throughout the sequence. The upper voice of each polychord progresses from D to B*b*, then A*m* to G, then F to E*m*, then E*b* to C. The reason behind this sound design is to create a directional, yet not forceful or abrasive, progression that is pleasing and is not an explicitly known or well-established sonic memory of chordal progressions. Similar to the chordal progressions embedded within the hospital-based system, this new compositional tactic provides the feeling of forward movement. This sequence becomes a recognizable signature for the patient to understand that recent groups of movement activities exhibited ideal movement performance.

Figure 2. Composition for "Ideal" Sequence

A “too slow” sequence signifies an overall too slow movement for a group of activities. As shown in **Figure 3**, this composition has decrementing pedal tones for each pair of polychords starting at C and ending at G. Its upper voice progression is D*bm* to B*m*, C*m* to B*bm*, B*m* to A*m*, then A*bm* to G. The reason behind this sound design is to enforce that there is movement, but also to convey a tone that is “dark” and "foreboding". This gives the forewarning: “Something is not correct”. The reason behind the pedal tones in these first two sequences is to provide consistency throughout the entire progression, in order to tie the polychords together. This sequence becomes a recognizable signature for the patient to understand that recent groups of movement activities exhibited movement performance of being too slow.

Figure 3. Composition for "Too Slow" Sequence

A “not smooth” sequence signifies an overall inconsistent movement for a group of activities. As shown in **Figure 4**, this composition does not have pedal tones, which, contrary to the first two sequences, enforces the concept that the movement activity is not consistent. It instead moves through “tetrachords” which are all based on C major. It begins: D to G and G to C, then C to F and F to B, then B to E and E to A, then A to D and D to G. The reason behind this sound design is to create a cyclical motion similar to how someone moving in a hesitant manner would travel. This sequence becomes a recognizable signature for the patient to understand that recent groups of movement activities exhibit an inconsistent movement performance.

Figure 4. Composition for "Not Smooth" Sequence

A “too fast” sequence signifies an overall too fast movement for a group of activities. As shown in **Figure 5**, this composition also does not have any pedal tones. It progresses through pentatonic scales. Each pair is one or two minor seconds away from each other, creating noticeable, jarring, fast-moving and non-resolute successions. This sequence becomes a recognizable signature for the patient to understand that recent groups of movement activities exhibited movement performance of being too fast.

Figure 5. Composition for "Too Fast" Sequence

3.2.2.2 Sonic Rhythmic Distributions

The sonic image tags listed above form a cohesive sound suite that can be generalized while implementing different sound samples without changing the semantic meaning of the SME The three primary semantics are consistency of movement activity duration, ideal acceleration and deceleration speeds during movement, and the peak velocity during movement. While sonic content is an important factor in relaying these semantics, it is rhythm that holds the key towards maintaining useful and consistent sonic media and movement information.

Table 3. Rhythmic density distribution per sonic image tag

	Density 4	**Density 3**	**Density 2**	**Density 1**
Normal	340ms	180ms	120ms	90ms
Too Slow	500ms	400ms	320ms	256ms
Hesitant	400ms	250ms	349ms	198ms
Too Fast	300ms	150ms	90ms	62ms

Another sonic attribute that can be manipulated is the speed filters. Mapping constraints help determine the most important attributes from which to focus and to pull the most valuable pieces of information. **Table 3** shows note durations for each of the densities within each of the speed filters for each of the sonic image tags. The densities are broken down as ratios of each other starting with the base (density #4). As acceleration continues and adaptable spatial thresholds are crossed, the system traverses through the densities below (from left to right). These densities fit within the spatial thresholds much like the harmonic progressions do within the hospital-based system, creating bell-like note densities.

Other sonic attributes that can be manipulated are the register range, change, and the probability of change. A low probability of change would keep notes within one or two registers, which allows for repeated notes.

Repeated notes convey to the patient that "beautiful, expressive melodies" are not being accomplished, therefore a low Brownian variable would be needed. Constraining to one or two primarily high (or primarily low) registers conveys the feeling of "being stuck", which is not an ideal situation. The patient must break out of the rut by correcting movement activities in order to create collections of more pleasing melodies.

Although much research has already gone in to the conceptual design of this composition model (see Section 2), defining the sonic space within the framework of a rehabilitation system that provides *post-set* sonic media information is relatively novel.

3.2.2.3 Assessment of the Sonic Media Environment

The efficacy of the affective summaries for the SME discussed in Section 3.2 was evaluated with 11 unimpaired subjects in an IRB-certified user study. The purpose of the study was to determine if there is a discontinuity between understanding sonic media environments given in real-time versus sonic media environments given after groups of movement activities. It helped inform and validate the sonic media designs used within the home-based system. The music conveyed a set of movements that were either all fluid, inconsistent, all too fast, or all too slow. During the user study, the subjects matched the sonic tags to the provided affective descriptors with over 90% accuracy.

4. CONCLUSION

The purpose of this paper is to document three years' worth of research on the sonic media environment for the home-based mixed reality rehabilitation system and to exhibit validated proof that meaning connects action to music. Although the home-based system has not yet been used or tested in real scenarios, there is a significant ground-truth basis that conveys the need for such systems. Proposed within this paper was to change the dialogue surrounding sonic media environments for rehabilitation purposes by researching viable solutions and designing a system to provide an immersive, multi-layered sonic media environment. Discussed, herein, is the motivation behind the project. Herein, the previous sound paradigms are defined alongside the proposed sound design, as well as the system design and the quantitative and qualitative solutions that target a new discussion for audio biofeedback.

5. REFERENCES

Burkholder, J. P., D. J. Grout, & C. V. Palisca. 2010. *A history of western music*. New York: Norton & Company.

Cirstea, M. C., & M. F. Levin. 2007. *Improvement of arm movement patterns and endpoint control depends on type of feedback during practice in stroke survivors*. Neurorehabilitation & Neural Repair, vol. 21, pp. 398.

Dourish, P. 2004. *Where the Action Is: The Foundations of Embodied Interaction.* Cambridge, MA: The MIT Press.

Duff, M., et al. 2010. *An Adaptive Mixed Reality Training System for Stroke Rehabilitation*. IEEE Transactions on Neural Systems and Rehabilitation Engineering, 18(5), pp. 531-541.

Essl, G. & P.R. Cook. 1999. *Banded waveguides: Towards physical modeling of bowed bar percussion instruments*. Proceedings of the International Computer Music Conference (ICMC). Beijing, China. October 22-27.

Gabrielsson, A., & E. Lindstrom. 2001. *The influence of musical structure on emotional expression*. Music and Emotion: Theory and Research, vol. 223, pp. 223-248.

Gardner, H. 1999. *Intelligence reframed: Multiple intelligence for the 21st century*. New York: Basic Books.

Ghez, C., R. Scheidt, & H. Heijink. 2007. *Different learned coordinate frames for planning trajectories and final positions in reaching*. Journal of Neurophysiology, 98(6), pp. 3614-3626.

Ghez, C., T. Rikakis, R.L. DuBois, & P.R. Cook. 2000. *An Auditory display system for aiding interjoint coordination*. Proceedings of the International Conference on Auditory Display. Atlanta, GA. April.

Headlee, K., T. Koziupa, & D. Siwiak. 2010. *Sonic Virtual Reality Game: How Does Your Body Sound*. Proceedings of the International Conference on New Interfaces for Musical Expression. Sydney, Australia. June 15-18.

Jaques-Dalcroze, E. 2007. *The Eurhythmics of Jaques-Dalcroze.* Rockville, MD: Wildside Press.

Papert, S. 1980. *Mindstorms: Children, Computers, and Powerful Ideas*. New York: Basic Books, Inc.

Prinz, W. 1990. *A common coding approach to perception and action.* Relationships between Perception and Action, pp. 167-201.

Prinz, W. 1997. *Perception and action planning*. European Journal of Cognitive Psychology, vol. 9, pp. 129-154.

Schmidt, R.A. 1991. *Motor learning principles for physical therapy*. In M.J. Lister, editor. Contemporary management of motor control problems: Proceedings of the II STEP Conference. Alexandria, VA: Foundation for Physical Therapy. pp. 49-63.

Siwiak, D., et al. 2011. *A home-based adaptive mixed reality rehabilitation system.* Proceedings of the ACM International Conference on Multimedia. New York, pp. 785-786.

Siwiak, D., J. Berger, & Y. Yang. 2009. *Catch Your Breath - musical biofeedback for breathing regulation*. Proceedings of the International Conference on New Interfaces for Musical Expression. Pittsburgh, PA.

Seitz, J. 2005. *Dalcroze, the Body, Movement, and Musicality.* Psychology of Music, vol. 33(4), pp. 419-435.

Sundaram, H., & T. Rikakis. 2006. *Experiential Media Systems.* In B. Furtht, editor. Encyclopedia of Multimedia (Vol. XXVIII, p. 989).

Thaut, M. 2007. *Rhythm, Music, and the Brain: Scientific Foundations and Clinical Applications*. New York: Routledge.

NOTATIONAL SEMANTICS IN MUSIC VISUALIZATION AND NOTATION

Lindsay Vickery
School of Music, Western Australian Academy
of Performing Arts

ABSTRACT

This paper examines a range of methods of exploiting the inherent semantic qualities of graphical symbols, colour and visual communication. Moody's Notations Theory is used as a starting point in the discussion of expanding the range of techniques for visualizing sound and instrumental notation. Recent findings in the understanding of semantic primes, visual language, perceptual metaphors and "weak synaesthesia" are examined and connections to existing sound-based fields such as spectromorphology, action-based scores, graphical and animated notation. The potentials for the use of colour to represent timbre both for descriptive analytical and prescriptive compositional tool in electroacoustic music is explored. Works by Cathy Berbarian, Luciano Berio, Aaron Cassidy, Vinko Globokar, Juraj Kojs, Helmut Lachenmann, Ryan Ross Smith and the author are discussed.

1. INTRODUCTION

The expansion of the possibilities for the musical score afforded by modern printers and computer screens has provided the opportunity to represent parameters of musical phenomena that were previously poorly captured by traditional Western music notation: most importantly continuously evolving parameters such as timbre and amplitude, and the depiction of complex sound events such as those found in electronic music.

The musical score is a time critical form of visualisation that, in the majority of cases, corresponds to sounds that unfold at a more-or-less defined temporal rate. For this reason there is a strong imperative for scores to employ symbols that signify sonic events with maximal efficiency. Moody's *Physics of Notations Theory* (Moody 2009) defines a set of principles to evaluate and improve the visual notation, that are pertinent to musical notation (See Figure 1.)

While Moody's principals are useful to consider in the creation of any form of notation, they are particularly pertinent to creators of music that challenges existing paradigms. As composers continue to explore increasingly idiosyncratic approaches to creating music, Cognitive Fit implies that notation should also evolve to flexibly reflect the materials that are represented. Approaches such as acousmatic music, microtonality, pulseless music, algorithmically generated music, guided improvisation, interactivity and mobile structure often poorly represented using traditional music notation.

Cognitive Fit:	use different visual dialects when required.
Semiotic Clarity:	there should be a one-to-one correspondence between semantic constructs and graphical symbols.
Perceptual Discriminability:	symbols should be clearly distinguishable.
Visual Expressiveness:	use the full range and capacities of visual variables.
Complexity Management:	include mechanisms for handling complexity.
Cognitive Integration:	include explicit mechanisms to support the integration of information from different diagrams.
Semantic Transparency:	use symbols whose appearance is evocative.
Graphic Economy:	keep the number of different graphical symbols cognitively manageable.
Dual Coding:	enrich diagrams with textual descriptions.

Figure 1. Moody's Physics of Notations theory (as summarized by Genon, et al. 2011:224)

Notation to capture the nuances of such explorations may require novel, semiotically clear, well defined visual languages that make full use of the range of forms of visual representation available to the composer, such as: colour, animation, hypertextual access, temporal coordination and multimedia integration (Vickery 2012a).

Colour provides a great potential for the formation of Perceptual Discriminability in a musical score. One obvious approach, for example, might be to employ a colour scheme that maximizes the distinctness of separate musical phenomena such as instruments, voices or sound sources. Similar requirements have been studied for the creation of data visualisation (Tufte 1990), transport maps (Green-Armytage 2010), and websites (Stanicek 2009). Recent research, however, has indicated strong perceptual correspondences between colour and a range of sonic phenomena (Prado-Leon, Schloss, and Palmer 2011), suggesting there may be more intrinsic semantic value to be gained from colouring the score.

This paper explores the implications of recent research in visual representation with particular reference to achieving semantic soundness though the use of perceptual metaphor and crossmodal correspondence.

2. SHAPE, COLOUR AND THE SCORE

Synaesthesia, a condition in which an individual experiences sensations in one modality when a second modality is stimulated (Ramachandran and Hubbard 2001:4), has been the subject of scientific enquiry for over two hundred years (Campen 1999:11). In the late 1960s Luria proposed that there are 'remnants' of synaesthesia in ordinary individuals "that many ordinary people have, which are of a very rudimentary sort (experiencing lower and higher tones as having different colorations)" (Luria 1968:22). In 1996 Marks noted that "there are natural correspondences between experiences in different sense modalities, and that these seem to be nothing less than "hard wired." (Marks 1996:61). Ramachandran and Hubbard have also proposed that "there may be natural constraints on the ways in which sounds are mapped on to objects" (Ramachandran and Hubbard 2001:19), citing Köhler's bouba/kiki experiment (Köhler 1929:224) as an example. In this experiment, "because of the sharp inflection of the visual shape, subjects tend to map the name kiki onto the (pointed, star-like) figure (…), while the rounded contours of the (other) figure make it more like the rounded auditory inflection of bouba" (Ramachandran and Hubbard 2001:19).

This phenonmenon has come to be known as Weak Synaesthesia (Martino and Marks 2001) or simply Crossmodal Correspondence (Deroy and Spence 2013). Martino and Marks differentiate between strong and weak forms of synaesthesia as follows:

Strong Synaesthesia:

One stimulus is perceived, the other is experienced as an image;
the correspondences are both idiosyncratic and systematic;
the definition of the correspondences is absolute;
associations are literal and semantic.

Weak Synaesthesia:

Both stimuli are perceived;
the correspondences are systematic;
definition of the correspondences is contextual;
associations are metaphorical and semantic.

(Martino and Marks 2001:63)

One of the key concepts underlying Traditional Western Notation (as well as many visual representations of sound such as the spectrogram) is the vertical spatial depiction of frequency in which higher frequencies are also vertically higher on the page. This interpretation has been shown to be supported by apparently basic latent mapping inherent cross-modal understandings in infants as young as 1 year old (Wagner, Winner, Cicchetti, and Gardner 1981) and pan-culturally (Eitan and Timmers 2010:419). Eitan and Timmers suggest that "pitch metaphors, while culturally diverse, may be based upon basic underlying mappings, stemming from bodily-based inter-modal interactions with the physical environment" (Eitan and Timmers 2010:407).

Walker has proposed that cross-modal correspondences are ordered in clusters as shown in Figure 2. Walker claims that "the same core correspondences should emerge whichever sensory feature is used to probe them, confirming that the en bloc alignment of the dimensions is context invariant" (Walker 2012:1806). Likewise Eitan and Timmers have suggested that "such implicit and automatic correspondence of positions and directions on the vertical spatial plane with non-verbal behavior may engender a ''second-order" mapping of ''high" and ''low" auditory pitch into features such as valence, mood or social hierarchy, as well as physical features like size and mass" (Eitan and Timmers 2010:407). They propose that:

> *The percept of pitch involves two contrasting magnitude representations. On the one hand, as pitch ''rises" its metaphorical height, intensity, and visual lightness increase; on the other hand, however, its metaphorical mass, size, and quantity decrease.* (Ibid:420)

These findings suggest guidelines that might inform the creation of "semantically transparent" notation for non-traditional musical sound sources. Exploring these ideas, the work *unhörbares wird*

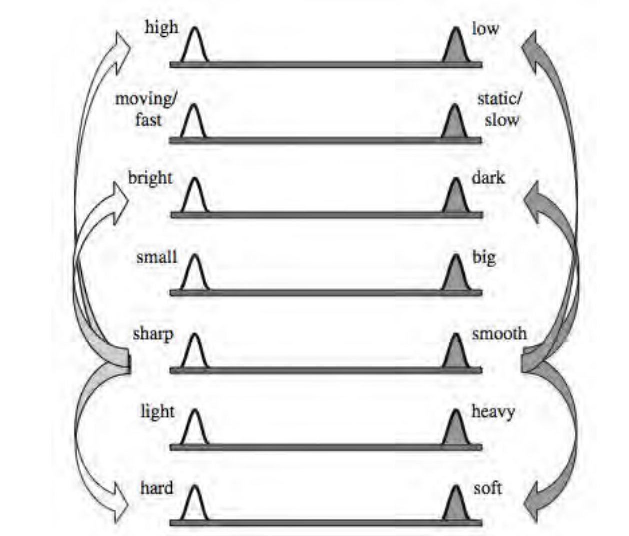

Figure 2. Proposed *En Bloc* alignment of cross-sensory features contributing to cross-modal perception (Walker 2012:1793)

hörbar [2013] (the inaudible becomes audible) used a spectrogram as the basis for the score for flute, clarinet, viola, cello, percussion and electronics. Using a spectrogram as the basis for a score poses a number of challenges, as Grill and Flexer have indicated, spectrogram "visualizations are highly abstract, lacking a direct relationship to perceptual attributes of sound" (Grill and Flexer 2012:591). In particular the "spatial" representation of the sonogram lacks the relational quantifiers of a traditional score that presents the representation of sonic events in the context of a tempo and frequency grid. This raised issues concerning the identification of parameters such as pitch, timbre, dynamics and orchestration, the issue of synchronization of multiple performers and importantly the resolution of the spectrogram itself (See Figure 3.)

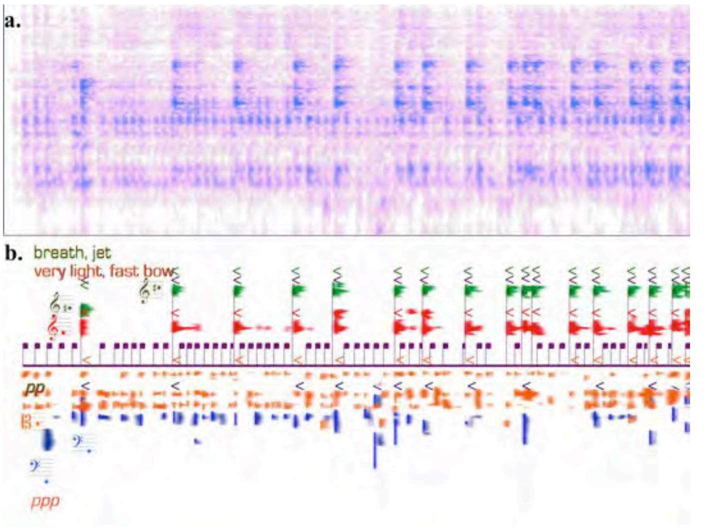

Figure 3. The opening 20 seconds of *unhörbares wird hörbar* [2013]: a.) shows the spectrogram upon which the work is based and b.) shows the score using graphical features of the sonogram as "noteheads".

In order to maintain a level of "graphic economy", a resolution of roughly 60ms/px was used for the spectrogram-score. This resolution allows the performer to view elements of the sonogram that represent what Curtis Roads refers to as "basic units of music structure (…) complex and mutating sound events on a time scale ranging from a fraction of a second to several seconds" (Roads 2002:3-4) while at the same time reading at an acceptable scroll rate of 2.35 cm/s (Vickery 2014a).

It was necessary to represent "Perceptual attributes" of the sonogram in a manner that that was maximally efficient and semantically sound, and therefore prominent features of the spectrogram are indicated using:

- "floating" traditional staff/clef/pitch symbols to specify pitch;
- the thickness of each player's line to represent dynamics; and
- transparency of the line (along with textual indication) to denote specific forms of timbral variation, from regular instrumental sound to diffused tones, "coloured noise" (Eimert 1955:4). in Stockhausen's terminology

The prominent shapes depicted in the original spectrogram are mostly retained, allowing the performer to calculate glissandi, minor fluctuations in pitch and timbral variation based on their interpretation the of colour, shape and size of the "noteheads".

The performers are synchronized by presentation of the scrolling score on networked iPads. This allows the acoustic instruments to remain coordinated with a spatialised re-sonification of the spectrogram that is played simultaneously.

The orchestration of individual instrument parts was colour-coded: flute - green, clarinet - red, viola - orange, cello - blue and percussion – purple with the aim of maximizing the distinctness of each part. Research at The Visual Perception and Aesthetics Lab at the University of California Berkeley, however, suggests that there is a high degree of correlation between mappings of colour-to-sound in non-synaesthetes. Griscom and Palmer have proposed that there are systematic relationships between colour and a range of musical phenomena including timbre, pitch, tempo, intervals, triads and musical genres (Griscom and Palmer 2012, 2013).

Human vision utilizes only three types of colour-sensitive cone cells - red, green and blue – and hues with frequencies between these (such as yellow, cyan and magenta) are perceived through stimulus to multiple cone cells. For example the frequency of yellow light falls between red and green and is detected by stimulus to both the red and green cone cells and as a result appears "lighter" than purple even though it frequency is not as "high". In this way, visual perception differs greatly from auditory perception. Figure 4. shows the a notional colour spectrum based on human visual perception.

The notional colour spectrum provides a palette from which colours representing sonic features or instruments might be chosen in a musical score. For most people this chart appears segmented into families of similar hue (yellows, oranges, tan, green-blue etc) and distinct but related hues may

Figure 4. A notional colour spectrum based on human visual perception from white to black (based on CIELAB colour space (Hoffman 2003) and Bruce MacEvoy's Artist's Value Wheel (MacEvoy 2005).

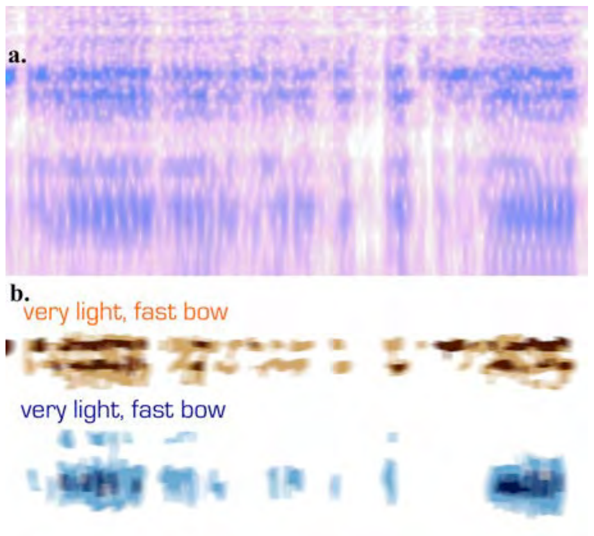

Figure 5. Colour as an indicator of timbral variation in the viola and cello parts of *unhörbares wird hörbar:* a. corresponding spectrogram and b. viola and cello parts.

lend themselves to the representation of timbral variation within a sonic feature or instrument.

As the number of represented features increases, however, so does the difficulty of discriminating between hues required. Green-Armytage suggests a palette of 27 tones based on white, yellow, orange, lime, green, turquoise, blue and purple and their lighter or more saturated counterparts, as a template for colour representation (Green-Armytage 2001).

Grisolm and Palmer have explored the idea of using cross-modal associations to define semantically sound mapping a range of instrumental timbres against colours in a two dimensional red/green: yellow/blue field. Grisolm and Palmer have also demonstrated that "color choices for multiple timbres are well predicted by an average combination of the component timbres" (2013). Interestingly they have observed, for example, that the yellow-blue value is correlated with attack time, whereas average red-green value is correlated with spectral brightness (2013).

Such observations may provide indications of how best to represent timbral information in coloured scores. Figure 5. shows detail from the score of *unhörbares wird hörbar* at 3m 35s, demonstrating the use of colour to indicate timbral variation in the viola and cello parts.

In another work The *Lyrebird: Environment Player* [2014b] a Max patch was built along these lines to visualise sonic features of field recordings. The score represents the frequency and amplitude of the single strongest detected sinusoidal peaks as rectangles drawn on a scrolling LCD object.

Brightness, noisiness and bark scale data derived using Tristan Jehan's `analyzer~` object are used to determine the luminance, hue and saturation of each rectangle. In contrast to a spectrogram, only principal sonic features are depicted, however timbral features are reflected in the changing colour of the rectangles. Figure 6. shows a simple example in which one of the long-crescendo F#s from the clarinet part of Messiaen's *Abîme des Oiseaux* is shown represented as a spectrogram (using Chris Cannam's *Sonic Visualiser* software) and the *Lyrebird Environment Player.* This example illustrates the representation of continuous timbral and amplitude changes over the duration of the note.

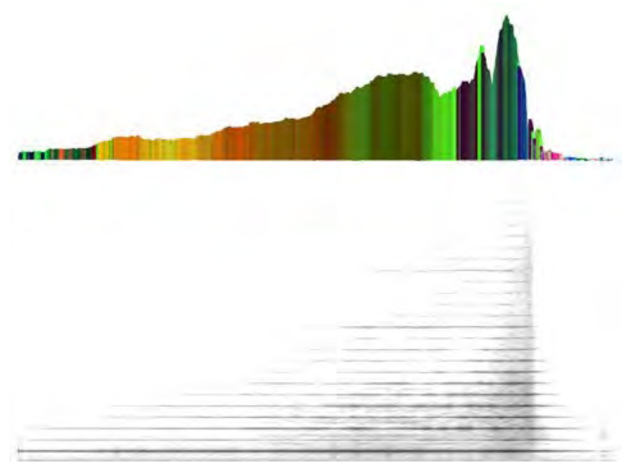

Figure 6. One of the crescendo F#s from the clarinet part of Messiaen's *Abîme des Oiseaux* represented as a spectrogram and the *Lyrebird Environment Player.*

This approach also has application for the analysis of electroacoustic music, somewhat alleviating the problem of "demonstrating coindexation and segmentation due to the difficulty in illustrating differences in timbre" (Adkins 2008) in a spectrogram and provides an (almost) realtime feature analysis of the recording in which contours and timbral shifts are readily recognizable.

Figure 7. shows a representation of Pierre Schaeffer's *Étude aux Chemins de Fer*, clearly delineating segments of the work created with varied source materials by consistently colouring sound objects of the same materials. The insert shows the whistle that occurs at approximately 112 seconds into the work and illustrates the "Doppler" effect that is heard through a change of both vertical height (pitch) and colour (timbre).

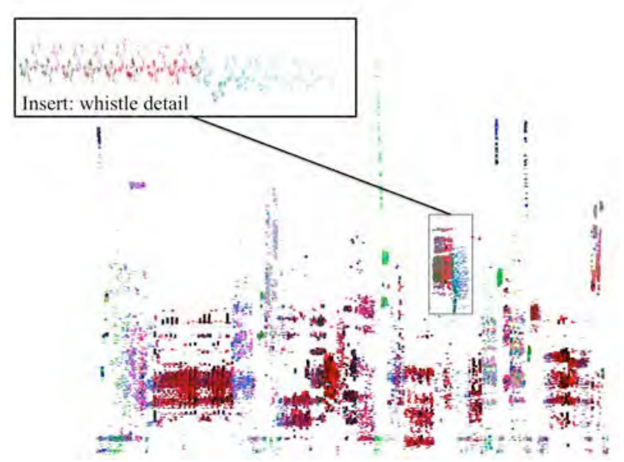

Figure 7. Pierre Schaeffer's *Étude aux Chemin de Fer* represented in the *Lyrebird Environment Player*, illustrating segmentation by colour (timbre). The insert shows the Doppler effect on a whistle that occurs at approximately 112 seconds in greater detail.

3. VISUAL METAPHOR

Moody refers to the evocative appearance of Semantically Transparent notation, implying that such notation should ideally be inherently sensible to the reader. In addition to current research into "weak synaesthesia", other fields of research potentially contribute to the creation of semantically transparent scores. Wierzbicka has investigated the concept of "Semantic Primes", innately understood concepts that cannot be expressed in simpler terms (Wierzbicka 1996), visual language, the creation of semantic graphical symbols for non-verbal communication (Horn 1998) and the concept of perceptual metaphors, which proposes that physical experience and embodiment bring about heuristic understandings that can be expressed in metaphors (Marks 1996). Marks claims "metaphors reflect processes of thinking and, consequently, appear not just in language but in perception as well" (Marks 1996: 39). Arnheim expressed a similar concept that he termed isomorphism, "according to which processes which take place in different media may be nevertheless similar in their structural organization" (Arnheim 1949:157).

A useful starting point in discussing semantically transparent notation from the standpoint of semantic or perceptual metaphor is Patel, Schooley and Wilner's collection of visual principals that convey meaning in graphic symbols (See Figure 9.). It was developed to evaluate Picture Communication Symbols, "a popular augmentative and alternative communication symbol set" (2007:65).

Gestalt:	Proximity, Similarity, Common region, Connectedness
Semantic Attributes:	Increment, Anthropomorphis, Possible outcomes, etc.
Cartoon Conventions:	Emotion, expression, Motion, Physical phenomena, Speech balloons, Embodied experience, Cartoon metaphors, Arrows
Compositional Distinctions:	Symmetry, Asymmetry, Repetition, Singularity, Juxtaposition, Exaggeration,
Line Interpretation:	Horizontal lines, Vertical lines, Active lines, Converging lines, Diverging lines

Figure 8. Visual Principals that Convey Meaning in Graphic Symbols (Patel, Schooley and Wilner 2007).

These visual principals are pertinent to music of a more textural nature. They suggest a way of conceiving of sound not unlike Denis Smalley's "spectromorphology", which provided "a list of terms, some of them technical, some more metaphorical, which can be used to interpret the function-significance of an event or context" (Smalley 1997:115).

The terminology employed by Smalley is often quite abstract (for example: "emergence" or "transition") and yet comprehensible by composers and listeners, at least in part through heuristic understandings based on "physical experience and embodiment".

Similarly Kramer argues for the presence of a level of semantic understanding of abstract structures in tonal music, based "gestures that sound characteristically like transitions, climaxes, contrasts, and other such conventions" (Kramer 1986 p. 140).

Blackburn and other have extended Spectromorphology into the visual field hypothesis as both a descriptive analytical and prescriptive compositional tool in electroacoustic music by Giannakis (2006), Thoresen (2007), Blackburn (2009, 2011), Pasoulas (2011), and Tanzi (2011).

Blackburn cites the cross-modal quality of acousmatic sound "that "it is frequently reported that, in concert, acousmatic music has the powerful effect of conjuring imagery, shapes, trajectories and spaces, which we as listeners proceed to describe verbally" (Blackburn 2011:5). She proceeds to outline "a new graphical vocabulary based on spectromorphology" (2011:5) that conforms to many of the principals outlined by Patel, Schooley and Wilner (see Figure 9.). Blackburn's graphical vocabulary not only visualizes individual "sound units" but shows how they can be "strung together to form longer phrase lengths" or "morphological strings" (Blackburn 2009).

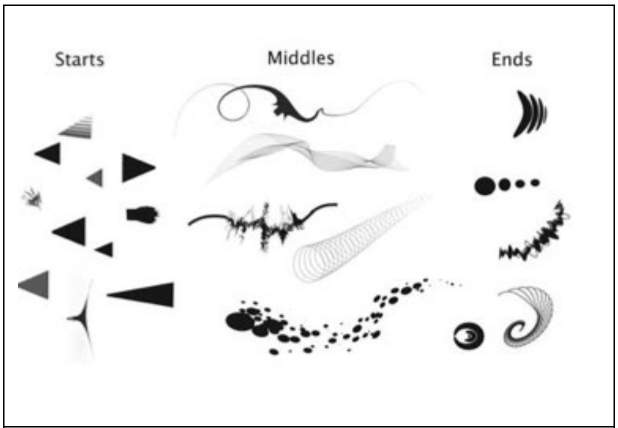

Figure 9. Spectromorphological shapes explored by Manuela Blackburn (2011).

Blackburn also emphasizes the use of perceptual metaphors, stating that words that are "more readily visualized ie. *spiral, flock, stream* and those with a clear associated physicality ie. *fly*, *drift*, *attack*, appear better suited for informing sound material creation" (Blackburn 2009).

This approach is equally applicable to instrumental music. In *Agilus Mimoid, Symmetriad* (2012) the instrumental score establishes a set of visual conventions that reflect a similar line of enquiry. The notation is proportional both horizontally and vertically: the duration of the note is equivalent to its spatial length (and scroll-time) and intervals and relative pitch of the instruments are always represented by the same vertical height.

This leads to the interesting condition that instruments performing the same pitch are notated "on top" of one another. In a traditional score this situation is avoided, however the networked scrolling score allows for the each individual's parts to be "brought to the front" while maintaining synchronization. This state of affairs allows the performer to view the full score and their own part simultaneously, providing a proportional visual representation of the other parts (see Figure 10a.).

Figure 10b. shows a passage from the work exploring semantic attributes – the broadening of an ensemble unison into a "cloud" of sounds of varied pitch and then "falling". The performers are only given the starting pitch and highest or lowest pitch of the "cloud" and are left to determine the exact note used to represent the visual figure themselves.

In Figure 10c. "coloured noise" is represented semantically by a textured line, suggesting a continuously changing timbre, for each instrument accompanied by textual performance instructions.

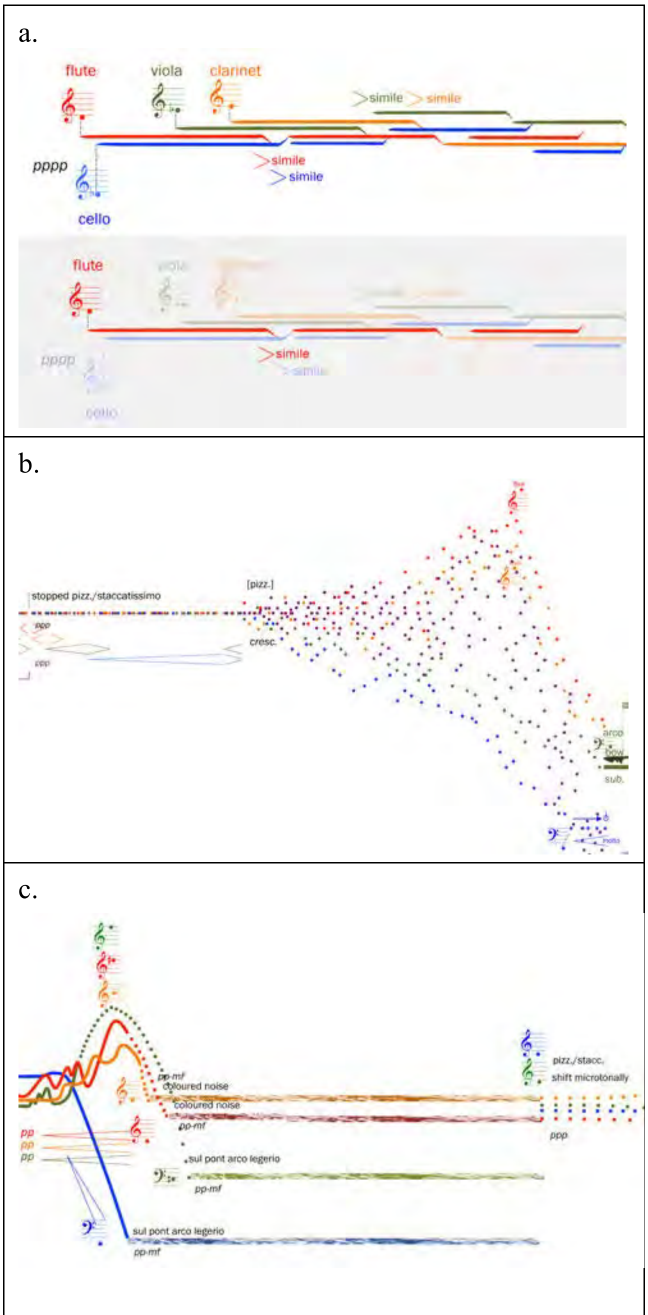

Figure 10. Excerpts from the score of *Agilus, Mimoid, Symmetriad* (2012b): a. vertically and horizontally proportional notated and score and part view; b. notation with semantic "cloud"-like and "falling" attributes; c. textured lines used to represent continuously changing "coloured noise" texture.

Patel, Schooley and Wilner also propose the use of "Cartoon Conventions" to communicate visually. In a broad sense, this includes any manner of visual representation of the physical world, and as such encompasses tablature-based forms of notation. Examples of tablature notation include systems commonly used for guitar and gamelan notation, but can also be found more experimental scores such as Berbarian *Stripsody* (1966), Berio *Sequenza V* (1966), Globokar *?Corporel* (1985) or Lachenmann *Pression* (1969-70).

Kojs defines musical scores that employ tablature to direct physical movements "Action-based music" which "*emphasizes the artistic exploration of mechanical actions which are used to control all aspects of composition, including its conception, form, instrumentation and instrumental design, performance and score* (Kojs 2009: 286). (See Figure 11).

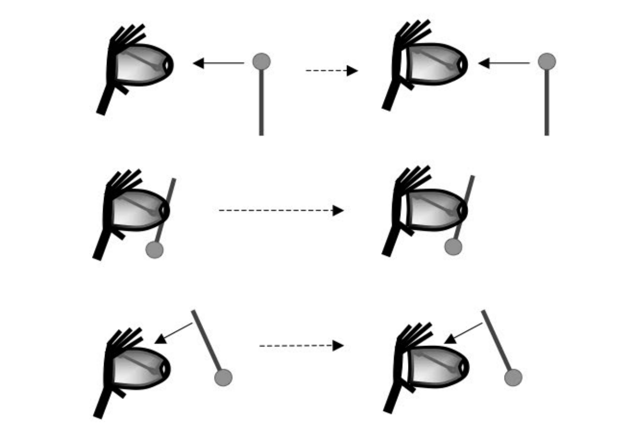

Figure 11. "Cartoon Conventions" used Juraj Kojs' *At and Across* for Slovak sheep bells and cyberbells (2007)

Berbarian's *Stripsody* embodies the principals of cartoon conventions as a mode of convention. The score is peppered with actual cartoon characters, speech bubbles and "action words" accompanying knock out punches and so forth. Lachenmann's *Pression* on the other hand actually depicts a cello fingerboard and defines the actions of the work against the tablature style physical depiction of their execution.

Such works might equally be represented with an fixed fingerboard image with animated notation superimposed upon it. Ryan Ross Smith's *Study no. 8 for 15 percussionists* (2013) (Figure 12.) is a animated tablature score depicting the movement of the mallets of 15 individual performers each represented by a figure. The smooth pendulum-like movement of the mallet symbols in this work allows the performers to anticipate the point at which they will strike the small grey circles on each side of the figure representing the instruments. This approach is relies on kinaesthetic understandings of motion rather than visual synchronisation.

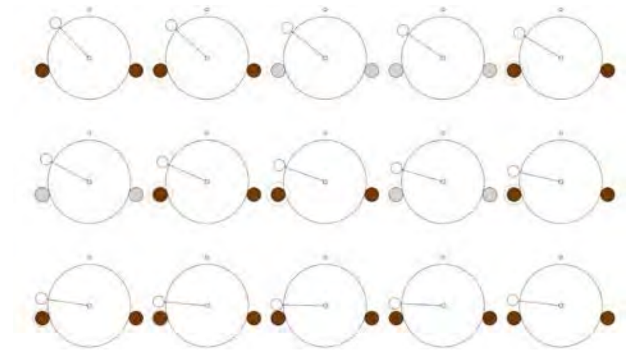

Figure 12. Excerpt from Ryan Ross Smith's *Study no. 8 for 15 percussionists* (2013)

When tablature and notation are combined it is often because additional non-traditional physical actions are required by the composer. One example is the polyphonicization of different components of performative technique that are normally unified into the single goal of "note production". An early example of this technique is Berio *Sequenza V*, in which the trombonist is directed to move the slide according to one contour while blowing (regardless of the outcome) at times defined by a separate stave.

The works of Aaron Cassidy expand this approach, often notating different components of instrumental technique on up to ten independent, simultaneous staves. This radical approach is the product of "experimentation with the polyphonicization of the various components of performative, physical action involved in producing sound in/on an instrument (…) the final resulting sounds of the piece are not in fact denoted in the score as such but instead arise as "aural byproducts" of the interaction of the (…) decoupled layers" (Cassidy 2000).

In Cassidy's *What then renders these forces visible is a strange smile (or, First Study for Figures at the Base of a Crucifixion)* (2007-08), the solo trumpeter is expected to simultaneously read 10 systems prescribing rhythmic values for embouchure tightness and articulation, breath pressure, tuning slide position and rhythm and positioning values for each valve. (See Figure 13.)

Works exploring polyphonicization of actions such as Cassidy's perhaps deliberately confront the limitations of traditional music notation, not to mention music reading itself. However, if these aims are set aside, the representation of multiple parameters using the techniques outlined by Patel, Schooley and Wilner and exploiting the semantic potentials of colour and shape, affords the potential of substantial simplification of complex multi-parametric scores.

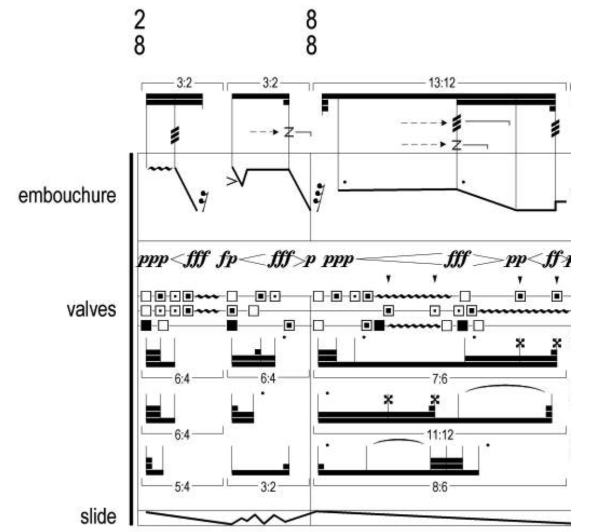

Figure 13. Aaron Cassidy What then renders these forces visible is a strange smile (or, First Study for Figures at the Base of a Crucifixion) (2007-08).

4. CONCLUSION

Many of the evolving techniques and pre-occupations of composers demand more continuous control of multiple musical parameters. The methods of representation and evaluation issues discussed here, especially when coupled with the affordances of the screen score, provide opportunities for composers, in instrumental and electroacoustic domains to capture the nuances of such works.

The issue of efficient and semantically sound notation is crucial for the development of effective notation for the screenscore. It is hoped that the current expansion of interest in “weak synaesthesia” will continue to contribute to the understanding of semantic “short-cuts” to communication in music notation.

The real-world applications of the emerging variety of methodologies for presenting notation on screen remain unexamined. One possible strategy for the evaluation of these techniques is the use eye-tracking technology to observe the connection between what is seen and what is performed and heard. Further work by this author will attempt to establish an understanding of the interaction between readers and the screenscore.

5. REFERENCES

[1] Arnheim, R. (1949). The Gestalt theory of expression. *Psychological Review* 56, 156-171

[2] Blackburn, M. (2009). Composing from Spectromorphological vocabulary: Proposed application, pedagogy and metadata. http://www.ems-network.org/ems09/papers/blackburn.pdf

[3] Blackburn, M. (2011). The Visual Sound-Shapes of Spectromorphology: an illustrative guide to composition. *Organised Sound* 16:5-13

[4] Campen, C. (1999). Artistic and psychological experiments with synesthesia. *Leonardo* 32 (1): 9–14.

[5] Cannam, C., Landone, C., and Sandler, M. (2010). Sonic Visualiser: An Open Source Application for Viewing, Analysing, and Annotating Music Audio Files. *Proceedings of the ACM Multimedia 2010 International Conference*. Firenze, Italy.

[6] Cassidy, A. (2000). Asphyxia. http://www.aaroncassidy.com/music/asphyxia.htm"

[7] Deroy, O., and Spence C. (2013) Weakening the case for ‘weak synaesthesia’: Why cross-modal correspondences are not synaesthetic. *Psychonomic Bulletin & Review* 20: 643–664.

[8] ,Eitan, Z., and Timmers, R. (2010). Beethoven’s last piano sonata and those who follow crocodiles: Cross-domain mappings of auditory pitch in a musical context. *Cognition* 114:405–422 (419)

[9] Genon, N., Amyot, D. and Heymans, P. (2011). Analysing the Cognitive Effectiveness of the UCM Visual Notation System Analysis and Modeling: About Models. *Lecture Notes in Computer Science Volume* 6598 221-240: 224

[10] Giannakis, K. (2006). A comparative evaluation of auditory-visual mappings for sound visualization. *Organised Sound* 113: 297–307.

[11] Green-Armytage P. (2001). Colour zones, connecting colour order and every- day language. 9th Congress AIC, Rochester, *Proceedings of SPIE*, 4421: 976–979.

[12] Green-Armytage, P. (2010). A Colour Alphabet and the Limits of Colour Coding. Colour: *Design & Creativity* 510: 1–23.

[13] Grill, T., Flexer, A. (2012). Visualization of Perceptual Qualities in Textural Sounds. *International Computer Music Conference 2012*, Ljubljana, Slovenia

[14] Griscom, W. S., and Palmer, S. E. (2012). The Color of Musical Sounds in Non-Synesthetes. Paper presented at the *12th Annual Meeting of the Vision Science Society*, Naples, Florida.

[15] Griscom, W. S., and Palmer, S. E. (2013). Cross-modal Sound-to-Sight Associations with Musical Timbre in Non-Synesthetes. Paper presented at the *13th Annual Meeting of the Vision Science Society*,, May 10 -15, 2013 Waldorf Astoria, Naples, Florida.

[16] Hoffman, G. (2003). *CIELab Color Space*. http://docs-hoffmann.de/cielab03022003.pdf

[17] Horn, R. E. (1998). Visual language: Global communication for the 21st century. Bainbridge Island: MacroVU.

[18] Jehan, T. (2001). *Analyzer~*. http://web.media.mit.edu/~tristan/

[19] Köhler, W. (1929). *Gestalt Psychology*. New York: Liveright.

[20] Kojs, J. (2007). At and Across for Slovak sheep bells and cyberbells. MS

[21] Kojs, J. (2009) The Language of Action and Cyberaction-based Music: Theory and Practice, *Journal of New Music Research*, 38(3): 285-294

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

[23] Lachenmann, H. (1969). *Pression*. Wiesbaden: Breitkopf & Härtel.

[24] Luria, A.R. (1968). *The Mind of a Mnemonist*. New York: Basic Books.

[25] MacEvoy, B. (2005) *Artist's Value Wheel*. http://www.handprint.com/HP/WCL/vwheel.pdf

[26] Marks, L. E. (1996). On perceptual metaphors. *Metaphor & Symbolic Activity*. 11: 39-66.

[27] Martino, G., and Marks, L. E. (2001). Synesthesia: Strong and weak. *Current Directions in Psychological Science*, 10, 61–65.

[28] Moody, D.L. (2009). The "Physics" of Notations: Towards a Scientific Basis for Constructing Visual Notations in Software Engineering. *IEEE Transactions on Software Engineering* 35, 756–779

[29] Patel, R., Schooley, K. and Wilner, J. (2007).Visual Features That Convey Meaning in Graphic Symbols: A Comparison of PCS and Artists' Depictions. *Assistive Technology Outcomes and Benefits* 41: 62-80.

[30] Prado-Leon, L. R., Schloss, K. B., and Palmer, S. E. (2011). Color, music, and emotion in Mexican and US populations. *New Directions in Colour Studies*. Amsterdam: John Benjamins.

[31] Ramachandran, V.S., and Hubbard, E.M. (2001).Synaesthesia - A Window Into Perception, Thought and Language. *Journal of Consciousness Studies*, 8(12): 3–34.

[32] Roads, C. (2002). *Microsound*. Cambridge: MIT Press: 3-4.

[33] Smalley, D. (1997). Spectromorphology: Explaining Sound-Shapes. *Organised Sound* 2(2) (1997): 107-126.

[34] Stanicek, P. (2009) *Color Scheme Designer*. http:// http://colorschemedesigner.com

[35] Tufte, E. (1990). *Envisioning Information*. Graphics Press.

[36] Vickery, L. (2012a). The Evolution of Notational Innovations from the Mobile Score to the Screen Score. *Organised Sound* 17(4): 128-136.

[37] Vickery, L. (2012b). *Agilus, mimoid, symmetriad*. MS

[38] Vickery, L. (2013). *unhörbares wird hörbar*. MS

[39] Vickery, L. (2014a). Exploring a visual/sonic representational continuum. *International Computer Music Conference 2014*. (Forthcoming)

[40] Vickery, L. (2014b). Lyrebird: Environment Player. MS

[41] Wagner, S., Winner, E., Cicchetti, D., and Gardner, H. (1981). "Metaphorical" mapping in human infants. *Child Development*, 52, 728-731.

[42] Walker, P. (2012). Cross-sensory correspondences and cross talk between dimensions of connotative meaning: Visual angularity is hard, high-pitched, and bright. *Attention Perception and Psychophysics* 74:1792–1809.

[43] Wierzbicka, A. (1996). *Semantics: Primes and universals*. New York: Oxford University Press.