

Composition for Ubiquitous Responsive Sound Environments

Dan Livingstone and Eduardo Miranda

Computer Music Research

School of Computing, Communications and Electronics, University of Plymouth, Drakes Circus Plymouth
PL148AA United Kingdom
dlivingstone@plymouth.ac.uk eduardo.miranda@plymouth.ac.uk

Abstract

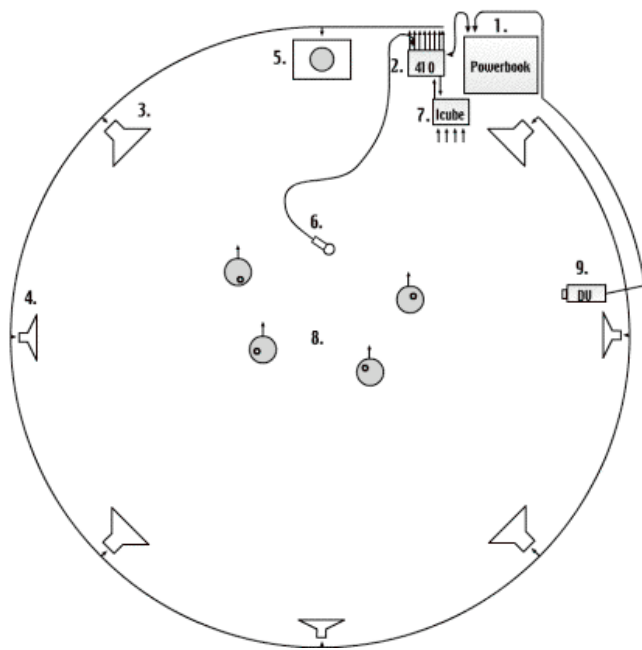
As new forms of social interaction with sound are developed through hardware, software and ubiquitous technologies it follows that emergent behavior, gesture capture and motion tracking will increasingly play a compositional role within generative and reactive sound environments. This case study defines an adaptive system, which enables participants within these spaces to have a tangible influence on the compositional process. Both individual and collaborative interaction modes are considered in the context of generative and real time systems, which are dynamically affected by user presence.

1 Introduction

Interactive music systems are often designed to provide engaging gestural control, enable new forms of musical expression, and are generally accepted to include three classes of compositional algorithms; sequencing, generation and transformation [Rowe 1993]. This case study establishes an integrative model for process driven collaboration [Livingstone 2003] within responsive compositional environments, by detailing the flow of interaction between composer/participants, a responsive sound environment and an adaptive compositional process. The system regenerates a soundscape dynamically by mapping 'known' gestures to influence diffusion and spatialization of sound objects created from evolving data, degrees of control are determined by clarity and scale of gesture, and the system is designed to adapt to these interactions by initial 'call and response' feedback within the structure of the composition. This is seen as a beneficial extension of the performer/performer relationship [Lippe 2002]. The structure of each 'response' is stored in memory

and compared to previous instances via mapping of the properties of each sound event ie; gestural trigger, related to concurrent sound object and its diffusion properties. This sequence of properties that encapsulate or encode each response instance can be visualized using additive synthesis as a model, each element (timing, diffusion, gesture mapping, synthesis and concurrent sound objects can be stored in a series of envelopes or partials that codify a discrete sound event. The interplay of sound events evolves as a dialogue is established between users and the system. The sound objects themselves are designed to initiate this dialogue, as 'psychoacoustic triggers' to interaction and are transformed through interaction. The system is given a collection of 'gestures' or patterns and building blocks for a range of sounds. Initial 'response' patterns are based on relating intentional gesture/movement to sound diffusion, and in(attention)al [Mack & Rock 2000] gesture/movement to re-synthesis of sound objects in memory. Compositional parameters are designed through continued observation of interaction through the vision system combined with live environmental sensor data. The system includes listener objects which can 'live sample' ambient sound material or intentional sound input with environmental data, enabling an adaptive approach for capturing sound events stored in short term memory as data parameters only; the system transforms this combined data and live audio which includes the acoustic properties of the physical environment to new compositional material. The integrative model discussed has been prototyped on a small scale using an apple laptop, icube, m-audio firewire410, 7.1 sound output, environmental sensors input through I-cube midi interface and video tracking for gesture capture. Software has been written in cycling74's version of Max/MSP/Jitter for OSX. This system (fig1 responsive system) provides an intuitive interface where simple gestures are used to interact with a 'real world composition' [Costa, Manzolli, Verschure 2003]

that is; an adaptive sound environment that is designed to be responsive to new interactions, as opposed to a pre – defined reactive systems approach.



(Fig1 responsive system)

1. G4 Apple laptop
2. M-audio Firewire 410 audio/midi interface
3. Set of eight active speakers positioned in either 7.1 surround format or two tiers simulating multiple height speaker system in atria, Portland Square.
4. Indicates offset speaker location of 2nd higher tier of four speakers
5. Additional active sub bass speaker
6. Microphone (AKG CB300) for live sampling
7. Infusion systems Icube interface.
8. Wireless 'composer & listener objects – clusters of sensors provide either environmental data or direct interaction) – can also use Bluetooth enabled mobile phone for diffusion control.

The system illustrated provides an adaptive framework that can be used for both specific compositional installations or for controlled experiments in sound perception/interaction/reaction to develop new compositional methods.

2 Interaction

Primary interaction with the system is via gesture capture using a real time feed through two fixed cameras, one overhead for general movement and orientation relative to physical space and one in front of the user localized to capture left or right hand movement. The vision system data is captured with Cycling74's Jitter software using matrix objects to track, map and compare a limited palette of symbols, *circle*, *line*, *triangle*, *square* and *cross* to user gestures drawn in the air with a finger/hand. Resolution of this action over time is increased using CV.jit externals by

Jean-Marc Pelletier (<http://www.iamas.ac.jp/~jovan02/cv/>) to track depth, speed and direction of gesture, the base information or algorithm for each symbol is stored in a matrix so the sequence of numbers that correlates to a 'known' gesture can have levels of accuracy and relative scale, allowing subtle variations in captured gestures to be identified. Max **mtr** (*multi track sequencer*) **capture** (*stores number streams*) and **env** (*Script-configurable envelope editor*) objects are used to store and compare these sequences with real time input from the vision system. This approach creates short term memory, allowing the system to match 'known patterns' (reactive system) but also to identify repeated unknown patterns which can then be added to long term memory as new symbols (responsive system), for example drawing an 's' several times will add this as a new symbol, previously unknown. This gestural composition process of interaction enables small scale gesture (individual – see fig 2) but can also be mapped to larger scale (group –see fig 4) behavior, for example social groups viewed from above can intentionally recreate symbols collaboratively by forming patterns or 'known' symbols tracked with similar methods as an individual hand gesture, this is achieved by designing an adaptive composition system and applying a methodology that allows for adaptive resolution, this approach can be considered as diachronic emergentism as consideration of the acoustic perception of sound objects is a key factor in sustaining the collaborative real time compositional process through effective sound design and spatialization to influence participants behavior and establish a musical 'dialogue'. Both the system and users are sensitive to the environmental properties of the composition environment; the localized portable version (fig 1) includes sensors for ambient temperature, light, and air movement, 7.1 sound diffusion, gesture capture and 'composer objects'. (A large-scale implementation of this system using additional data from a building management system with extended social interaction is being developed for field-testing in the Portland Square building, University of Plymouth. UK. <http://www.arch-os.com/>)

2.1 Conscious interaction

During individual interaction a parallel is drawn between recognized/non-recognized data combinations and intended/unintended actions recognized data is mapped to specific processes, drawing a *circle* will trigger a rotary pan with speed and direction, straight *lines* give panning settings, a *triangle* will create a new envelope for current sound object, a *square* defines a measure of time and a *cross* fades current sound object/s. Of course superficially this symbol based approach leads to a limited palette of compositional possibilities, one must consider that this limited palette provides a clear framework for interaction in line with the 'call-response' model which also enables the software design to factor user reaction to current real time

outputs. An agent based sound object can pan itself relative to user position, so any recognized intended action is subject to current system and environment parameters this approach can be described as multiple low resolution events combined to provide a more sophisticated higher resolution 'world view'



(fig2 small scale interaction - Gesture is tracked in relation to live sampled sound (sonograph) in this instance a new symbol 's' is added to short term memory creating a new envelope from data associated with this pattern or number sequence for diffusing a sound object in real time)

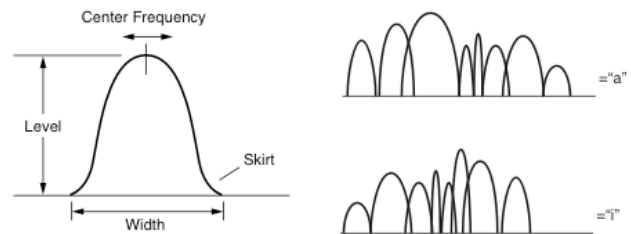
enabling the system to interject compositionally significant events [Camurri 2000] Composer objects; wireless interfaces capturing localized environmental data can also be manipulated by users to gain a higher level of influence over sound object creation/design. Both custom built sensor clusters and blue tooth mobile phone interaction are being prototyped, initial experiments show two distinctly different modes of intended interaction, users manipulating composer objects seek *dynamic control* over sound composition whereas mobile users have *positional control* of one element of the soundscape in relation to other participants for example. These modes of *intended* interaction all have the potential to establish a compositional dialogue with the system.

2.2 Subconscious interaction

When the system is not tracking known symbols or direct influence from composer objects, listening agents are used to mediate subconscious interaction. An example of subconscious interaction can be seen in figure three, an overhead motion tracking camera detects one slow moving and two fast moving social groups, connecting these groups relative co-ordinates creates a *triangle* the compositional system may recognize this or can be taught this variation based on the equilateral triangle pattern stored in long term memory, in this compositional system the *triangle* shape

influences the envelope driving the Formant synthesis element of a current sound object A *square* symbol influences timing and is identified in relation to the overall tracked area this can be applied to the current envelope or if a specified time has passed since the last identified symbol it is applied to current sound diffusion patterns. For example an 'ah' shaped formant is initiated on recognizing a *triangle*, environmental parameters excite resonators or anti resonators allowing the 'ah' to shift from vocal to nasal, if a *square* is formed timing values are adapted, a *circle* symbol morphs the formant to an 'ai' shape or other vowel shape by shifting the center frequency or manipulating the envelope generators through time based timbral shifts. A Yamaha FS1R rack synthesizer is used for real time control of Formant and FM synthesis, the system use a software interface created in MAX/MSP to control it. Of course users of the system may become attentive of this 'subconscious' process and choose to *intentionally* change the current sonic structure, collectively forming a uniform *square* for a longer duration to extend the temporal properties of a current sound object. Alternatively if no symbols are recognized live data from the composer/listener objects or selected parallel data from the building management system is used to reform base sound objects through FM synthesis.

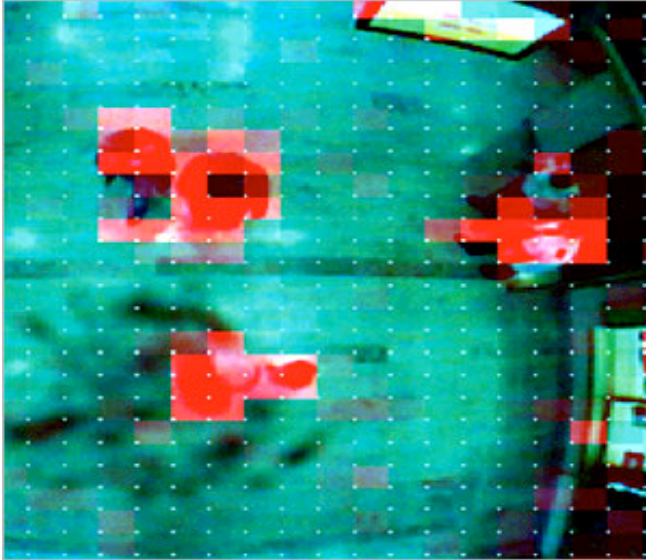
(Fig 3 basic formant shaping)



environmental data changes are usually slow in interior environments so these elements are mapped to the timbre and color of sounds created, providing an overall structure for the real time composition that is responsive to either the ambient light, temperature and air movement (test system) or to selected data real world data from the building management system.

2.3 Resolution

The portable system illustrated (fig1) is currently located in a small office the Portland Square Building, so while local gesture capture and synthesis methods are being tested and refined (small scale interaction) the full building management data and camera streams are monitored, allowing comparative analysis and continued prototyping for the real world system. The concept of 'resolution' has been a valuable tool in developing methods for tracking and synthesis that can be 'transposed' to a complex social environment with embedded ubiquitous technologies.



(Fig 4 large-scale interaction, Arch OS vision system, Portland Square – example of how subconscious interaction can be given meaningful compositional attributes)

3 Composition

The compositional approach is one of continued observation and refinement of the interaction process; the system has been given a subtle ‘voice’ through basic formants [Styger, Keller 1994], which are combined with more spatially specific fm synthesized sound objects. Output from the vision system tracking either small-scale gestures or large-scale movement is monitored to identify possible musical relationships that can be used to influence behavior. This is an approach of considered sound design which produces effective base sound material that is not overly complex, the system, site and participants influence these base materials either directly or indirectly to form new sonic structures that reflect the movement and physical properties of the compositional environment. This approach can be considered as an adaptive variation of spatial music. It is also a learning process, a number of tasks have been developed to field test these compositional processes.

4 Conclusion

A prototype responsive system has been developed which integrates a range of computer music techniques to provide a compositional approach to generative or interactive music. In this paper we described one prototype example/system ... key findings are that through a person centered design approach to interaction on an intimate level (hand gestures) that in principle an effective compositional dialogue can be established; novel interactions can be recognized and ‘learnt’ both by the system and those interacting with it. A strategy for deployment of this system on a larger scale, using the atria connecting offices and

teaching spaces, has been outlined for Portland Square building, University of Plymouth, UK.

Interaction design for larger social groups is being refined based on initial observation of motion of people through these spaces. (Fig 4) Ongoing research seeks to identify methods of recognizing social ‘intended’ interaction with the system by field-testing and refining the symbol based compositional process discussed (2.1). This research establishes process driven collaboration as a compositional methodology. Future applications could include the capture of ‘perceptual constructs’¹ [Livingstone 1998] as sound signatures of those participating with these systems. The potential for large scale adaptive games systems integrating personal mobile technologies with large scale social ‘learning’ environments offers significant potential for interdisciplinary research.

References

- Camurri, A. (2000) “Artificial Intelligence Architectures for composition and performance environment”. In *Miranda, E. R. Readings in Music and Artificial Intelligence (Contemporary Music Studies)* Routledge pp.163-188
- Costa, Manzolli, Vershure (2003) “Constructing Sonic Structures with Visual Roboser”. In *Proceedings of the IX Brazilian Symposium on Computer Music, SBC* pp131-136
- Lippe, C (2002) “Real time Interaction Among Composers, Performers, and Computer Systems” *Information Processing Society of Japan SIG Notes, Volume 2002, Number 123*, pp. 1-6.
- Livingstone D. (2003) “Emergent Behavior in the context of Reactive Compositional Environments”. In *Proceedings of the IX Brazilian Symposium on Computer Music, SBC* pp235-240
- Livingstone D. (1998) “The Space between the Assumed Real and The Digital Virtual”. In *Ascott, R. Reframing Consciousness – art, mind & technology Intellect Books* pp138-143
- Mack, A. Rock, R. (2000) *Inattentional Blindness* (Cognitive Psychology) MIT Press pp.215-255
- Rowe, R (1993) *Interactive Music Systems: Machine Listening and Composing*, Cambridge, The MIT Press.
- Styger, T. & Keller, E. (1994). “Formant Synthesis”. In *E. Keller (ed.), Fundamentals of Speech Synthesis and Speech Recognition: Basic Concepts, State of the Art, and Future Challenges* Chichester: John Wiley. pp. 109-128
- ¹ A perceptual construct in this context is a series of spatial relationships created by an individual manipulating sound objects as reference points in relation to physical space using the mental model of a wire frame cube to enhance spatial perception of sound objects, positional coordinates for each diffused sound are captured and compared with the users description or individual mental model of ‘their’ cube in relation to the physical environment and cubes of other people.