Artificial Intelligence systems as a solution to subjective video sensing in Contemporary Performing Arts.

Dr Garth Paine

Senior Lecturer, Music Technology, School of Contemporary Arts

Research Associate, MARCS Auditory Laboratories

University of Western Sydney

Email: ga.paine@uws.edu.au

Abstract

This paper discusses approaches to realtime motion tracking in contemporary dance. It outlines some problems with current techniques and proposes, through previous research, some alternative approaches that could provide much richer data sets for realtime sonification and visualization of choreographic patterns.

1. Background.

While undertaking my PhD (Paine, 2002b), I developed a number of approaches to the mapping of gesture (tracked using video based systems, VNS1 and Cyclops), which identify issues involved in the sonification of gesture (Mulder *et al.*, 1997) in a manner that creates a visceral engagement with the quality of outcome. The following are excerpts from that research as a way of contextualizing the need for the AI research discussed below.

During 2000, while I was the Australia Council for the Arts, New Media Arts fellow at RMIT University, I began some research that attempted to expand the interactive environment research (Moser & MacLeod, 1996) I had been doing over the years prior to include dynamic levels of intelligence in video tracking of movement and behaviour patterns.

Through my previous research involving interactive immersive sound installations and composing using interactive music systems for the Australian Dance company, Company in Space, I formed the opinion that for an interactive system to be substantially more complex and sophisticated than the current first or second order responses, a level of artificial intelligence had to be introduced between the sensing stage and the mapping of the sensed data to synthesis parameters (Bongers., 2000; Paine, 2002a). A relatively linear mapping of input data to a limited and fixed number of synthesis parameters does not support the evolution of system response over time. In order to pursue a model of interaction that goes beyond 'response' to a dynamic and intelligent relationship between the interactive agent(s), (human(s), engaged with the system/environment/installation, the system must be conditioned by it's accumulated experience (histogram), being able to evolve responses accordingly (Ascott, 1997).

Such a system would require a level of cognition in the form of a software infrastructure that could establish the patterns of interaction based on historical knowledge, and act accordingly. Neural networks are one possible approach, as are Hidden Markov models, both developed for pattern recognition tasks, and capable of being trained with an initial set of sensitivities, and able to evolve those sensitivities in response to varied input over time. Such systems built into interactive environments or interactive dance works may allow the system to be trained to:

- **Recognize** individuals from their gesture patterns and movement characteristics (a useful feature for training an interactive environment to respond independently to different people, and also useful for interactive dance performances where it could be compositionally valuable to attribute different response patterns to different dancers).

- **Make** subjective, qualitative judgments about the observed movement or gesture patterns, so that the system could determine the intent of the movement or gesture. Qualitative data of this kind would greatly extend the scope of current systems that respond to changes in light intensity per frame only, providing data that allow the calculation of speed of movement, position of movement, acceleration, relationship between two bodies etc. The accumulation of subjective, qualitative data would make additional layers of intention based aesthetic responses

1 See http://homepage.mac.com/davidrokeby/vns.html (08/08/05)

available through the structured control of more sophisticated synthesis algorithms. These additional layers would in turn make for much more refined, unique and individualized interaction, creating realtime sonifications or visualisation that tightly mirrored the nuance of the sensed gesture.

- **Control** vast numbers of synthesis variables in a structured manner, directly related to the subjective, qualitative data output of the neural network, and in turn control much richer aesthetic outcomes. For synthesis output to reflect the minutiae of individual gestures, the synthesis algorithms must become much more sophisticated, which in turn requires more variables, more than would be easily controllable in a non-structured manner. Artificial Intelligence systems could provide a mechanism to control vast numbers of synthesis algorithms in a controlled and subjective, qualitative manner, and as such would be a valuable addition to an interactive systems development tool kit. Artificial Intelligence systems could, through object-oriented programming, directly support the ideas of dynamic orchestration (Paine, 2004).

- **Analyse** the aesthetic output of the interactive system, and generate new sonifications or visualisation algorithms that would extend, or fine-tune the aesthetic scope of the output of the system. This may see the output algorithms of an interactive, responsive environment evolving over time, so that the response patterns of the installation adapt to an accumulated knowledge of how people interact with it, and in so doing may totally discard the algorithms the artist/composer(s) established for the piece. This kind of development is currently occurring in artificial life animated worlds both in exhibition environments and online (Sommerer & Mignonneau, 1998).

An examination of cybernetics(Wiener, 1948) moves the design approach away from the foundations of existing musical or computer sciences practice into an area of contemporary exploration of the phenomena of the natural world and the human condition, a common platform for creative exploration, and areas that in my view, are vital considerations in the development of a new artistic paradigm.

In order to explore these possibilities, I undertook some research with Dr Dinesh Kant, Senior Lecturer in Biomedical Engineering Research at the School of Electrical and Computer Systems Engineering, RMIT University, Melbourne, Australia.

My objective was to explore the possibility that Neural Network computing techniques could be used to recognise gesture patterns(Camurri & Volpe, 2004) so that;

- Gestures could be categorised using subjective human criteria (i.e. violent, loving, inclusive etc),

- Gestures could be used to identify the individual (i.e. distinguish between individuals movement characteristics),

- It would be possible to track multiple individuals with a single camera view. This would mean the tracking of gestures unique to individuals so that different response patterns could be associated with individuals within a group, ie. Individual dancers in a dance troupe.

- It would be possible to develop evolving mapping strategies that would change, based on subjective analysis by the neural network system (i.e. the characteristics of the current and historical movement gestures).

A number of experiments were carried out which illustrated that it was likely that a Neural Network could be trained to differentiate between two simple activities (drinking a glass of water, and writing a page of notes). There was also some positive indication that with further development, it would be possible to differentiate between the individual subjects. This research was not realtime, it used Matlab to build a back propagation Neural Network to analyse the data, a process that took several days for each subject.

A great deal of further research and development work is required before such a system produces reliable results, or most importantly before the system can run in realtime.

But the inclusion of AI in such interactive systems could produce much more advanced analysis-synthesis relationships; a system that not only tracks subjects, but is able to analyse input data in a subjective manner (by which I mean a human experience based ability to categorise sensed gesture), the sophistication of which can evolve in direct relationship to the history of input.

Furthermore, it is not currently possible to reliably track and identify a number of individuals within a video frame in an interactive system. I propose that an AI based system would be able to identify nuances in individual movement, and subsequently be able to reliably identify the individual who is the source of the gesture.

Mapping for Immersion

Immersion involves creating a three-dimensional interactive, responsive environment that envelops the

exhibition visitor in such a way that they feel engaged and captivated.

This experience is defined as distinctively and qualitatively different from the experience of listening to sound and music from loudspeakers as a detached observer.

Simon Emmerson (Emmerson, 1994) alludes to the importance of a perceivable relationship between performance and the outcome of a performer's gestures. For instance, a mouse connected to a laptop computer can be used to create massive and very fast changes in an electronic music performance. The size of the movement belies the outcome. Emmerson (Emmerson, 1996) discusses these issues in relation to the perceived location, or rather 'dislocation' of the sounding source in acousmatic music and much electroacoustic music using multi-channel diffusion/spatialisation techniques.

These issues have a direct relevance to the mapping of human movement to sound and vision creation in an interactive system. The weight of the gesture must be translated into a sound quality that communicates back to the mover something about the 'perception' of the movement. If the movement is large then the sound must change in scale with the movement. If the gesture is small then the timbre or texture of the sound must respond in concert with the gesture. The sounds may also come closer to the interacting body when intimate gestures are sensed or 'run' way when subject to aggressive gestures. In this way, it becomes apparent that the response of the installation is a direct result of not only the movement and gesture, but also the quality of the movement, and therefore a reflection of the intent of the person initiating the gesture. Such an interactive experience communicates a visceral, individualistic perception of engagement that is an imperative outcome of a successful interaction.

The spatialisation of the sound within the installation must be considered from two perspectives:

1. The creation of a sound field that creates a sense of immersion. Such a sense of immersion is generated when the sound field seems continuous, and when one is not aware of particular loudspeakers as the point source of the sound.

2. Individual elements of the sound field that are directly responding to the current movement should be diffused in such a way that they appear to have a spatial relationship with the source of the gesture in the installation. The sound source should be positioned as if the user has created a disturbance in the sound field. This requires a dynamic spatialisation system, placing each layer of the sound score in relation to the position of sensed movement.

Emmerson outlines the *"Three great 'acousmatic dislocations' established in the half century to 1910. These are: (1) Time (recording), (2) Space (telecommunications, telephone, radio, recording) and (3) Mechanical causality (electronic synthesis, telecommunications, recording)".* (Emmerson, 1994):98

These three categories can be adopted for the generation of sound within an interactive, responsive environment. They can be described as follows:

- **Time**: The speed of response of the environment to the gestural input must be such that there be sufficient immediacy for the user to perceive a direct relationship with their movement. In this usage, there is an attempt to prevent the 'dislocation' of time being argued by Emmerson.

- **Space**: Space, as discussed above, must be considered both in terms of the diffusion of sound, and the way in which the architectural space contributes to the construction of the interactive, responsive environment experience.

- **Mechanical causality**: The relationship between the causality and the response within an interactive, responsive installation is of utmost importance. The mapping of response patterns must generate a relationship that is immediately perceivable. The mapping must communicate something about a qualitative relationship, drawing a parallel between the quality of gesture and the nature of the change in response.

Emmerson goes on to say *"The aim is to be clear that in abandoning any reference to these 'links of causality' the composer of electroacoustic music – especially that involving live resources – creates a confusion (even a contradiction) and loses an essential tool for the perspective and engagement between the forces at work." (Emmerson, 1996):1*

This same clarity in intent is vital within an interactive, responsive environment. The mappings between movement and outcome may be many and varied; they may change in relation to the number of people in the space (GITM2), or the current dynamic of movement within the exhibition (MAP1, MAP2,

2 see http://www.activatedspace.com/Installations/ (08/08/05)

Gestation)3, and so they may offer an ever deepening relationship as the user understands the finer characteristics of the interface, but they must always be clear and immediately perceivable.

2. Dynamic Orchestration

As discussed above, the orchestration of a truly interactive environment must be dynamic, to vary in accordance with variations in the dynamic of movement, relationships and choreographic interplay. In my own installation and interactive dance works, I have changed the weight of the sound texture in such a way as to reflect the weight of the sensed gesture.

Creating a perceivable link between the weight of gesture and the density of texture, provides a visceral, tactile quality to the interactive experience. This relationship draws on traditional instrument design, where a more intense engagement with an instrument generates a change in timbre that reflects a more complex overtone structure. In general, acoustic instruments also illustrate a relationship between energy input and amplitude of output.

This consideration led to the mapping of movement to intensity used in MQM and GITM.

Amplitude is supported in these two works by changes in:

- **The intensity of sound**. For instance, the sounds in MQM and GITM range from meditative drones and slow rhythmic patterns through to loud distorted tones and complex polyphonic sonic objects. This change in the character of the sound represents intensity.

- **The period of sound events**. The sounds in MQM and GITM reduced in length in direct relationship to their intensity. Gentle sounds are longer (08 - 20 seconds); intense sounds are shorter (3 - 8 seconds). The variation in sound file duration brings about a slowing or speeding up of the rate of sound events. This change in the speed of sound events creates a variation in density (meditative/ energetic) and the subsequent increased complexity of orchestration.

MAP1 freed the sound/gesture relationship allowing exhibition visitors to create changes in the sound environment by explicitly varying individual parameters.

The floor space of MAP1 was divided into four large sliders, each controlling a different synthesis parameter. Each synthesis variable would move to the position of last activity within its zone. The rate at which the parameter moved was determined by the sensed activity in the target field (dynamic activity = fast, meditative activity = slow, and all gradations in between). In this way, MAP1 was played like an instrument.

Other variables were determined by the sensing field (64 fields in four rows of 16) with the highest activity. These variables include a range of variation (jitter) in some variables, i.e. grain length and pitch, which created an increasing rate of change in the focus of the primary setting in direct relationship with the increased intensity of sensed dynamic of movement. The centre of the variable range was, however, set as described above.

Beyond the rate of change of synthesis parameters, MAP1 separated the synthesis parameters from a direct relationship with the dynamic of movement and gesture in favour of precise control of parameters based on spatial plotting within active zones. This approach was continued in the latter works (MAP2).

Gestation explored a much more layered approach, seeking to employ aspects of qualitative second order analysis. Acceleration, size of movement, and proximity of bodies was used to form behaviour patterns for each sonic layer. The dynamic orchestration therefore took on an additional layer of association with the gesture and choreography, having both timbral and spatialisation characteristics subject to sets of rules of interaction. This formed a much more responsive environment.

3. Conclusion

The development of a realtime sensing system embodying AI modules for the performing arts would move approaches to sonification and visualisation of performance characteristics much further in the direction of force-feedback research, where the affordances between gestural nuance and the characteristics of sonification and visualisation outcomes would provide a direct and visceral engagement with the environment. Such a system would open up possibilities for interactive performance that would change the very nature of the relationship between making and performing work.

References

Ascott, R. (1997). *The technocratic aesthetic: Art and the matter of conciousness*. Paper presented at the CAiiA Research Conference, Conciousness Reframed, art and conciousness in the post-biological era, University of Wales, Newport.

3 MQM, GITM, MAP1, MAP2 and Gestation are interactive environments developed by Garth Paine

Bongers., B. (2000). Physical interfaces in the electronic arts. Interaction theory and interfacing techniques for real-time performance. In M. Wanderly & M. Battier (Eds.), *Trends in gestural control of music*. Paris: IRCAM - Centre Pompidou.

Camurri, A., & Volpe, G. (2004). *Gesture-based communication in human-computer interaction: 5th international gesture workshop, gw 2003: Genova, italy, april 15-17, 2003: Selected revised papers*. New York: Springer.

Emmerson, S. (1994). *'live' versus 'real-time' cmr 10:2 pp. 95-101*. Amsterdam: Harwood Academic Publishers.

Emmerson, S. (1996). Local/field: Towards a typology of live electronic music. *Journal of Electroacoustic Music, 9*, 10-12.

Moser, M. A., & MacLeod, D. (Eds.). (1996). *Immersed in technology: Art and virtual environments*. Massachusetts: The MIT Press.

Mulder, A., Fels, S., & Mase, K. (1997). Mapping virtual object manipulation to sound variation. *IPSJ SIG Notes, 97*(122), 63-68.

Paine, G. (2002a). Interactivity, where to from here? *Organised Sound, 7:3*(3), 295 - 304.

Paine, G. (2002b). *The study of interaction between human movement and unencumbered immersive environments*. Unpublished PhD, RMIT University, Melbourne.

Paine, G. (2004). *Gesture and musical interaction: Interactive engagement through dynamic morphology*. Paper presented at the NIME, Hamamatsu, Japan.

Sommerer, C., & Mignonneau, L. (Eds.). (1998). *Art @ science*. Wien: Springer-Verlag.

Wiener, N. (1948). *Cybernetics*. Cambridge, Massachusetts: The MIT Press.