Alistair Riddell
The Australian National University
Centre for New Media Arts
121 Childers street
Acton, 0200
Australia
alistair.riddell@anu.edu.au

Sound Performance within an Augmented Reality Theatre Context

Abstract

This paper considers sound performance within a broader theatrical presentation using an augmented reality system. An examination of the modes of sound control and interaction reveal certain inherent difficulties and raises questions as to how sound might be more effectively performed in this context. Such dilemmas are increasingly common as Sound Art embraces more complex forms of sound and image synergies.
In addition, further questions arose in the process of compiling this paper, as to the appropriateness of placing a diverse range of performance demands on artists, and expecting that they will be able to maintain conscious control over their individual sound output in a multi-channel system.

Finally, the paper reflects on the outcome of the project and what the experience revealed with respect to future projects.

Introduction

The intention of the *Edible Audience* project[1] was to articulate a concept through the use of an augmented reality system (*AVIARy*[2]), which integrated images into live video projection and controlled sound diffusion, all through performer interaction. There was a noticeable absence of the technology during the performance with the emphasis being on the performers, the sound and the projection. This will be discussed in detail later in this paper.

The concept was actually quite simple and straightforward with a humorous, if rather dark perspective on the nature of consumption in contemporary society. The performance action centred on the 'consumption' of images of the audience and body parts, and was formally structured as an evolving narrative across the courses of a meal: Entrée, Main Course, Desert and a Toast. Although the technical implementation and the performance mostly ran smoothly[3], the complexity of the event gave rise to questions concerning how sound performance should be approached in this context.

[1] *Edible Audience* was performed as part of the Liquid Architecture 6 Sound Art festival at the National Gallery of Australia in July 2005.
http://www.nga.gov.au/LiquidArchitecture/bios.cfm
[2] *AVIARy* was written by Tim Barrass using the jARToolKit.
[3] This will be discussed later in the paper.

Performance Configuration

The technical details of the *AVIARy* configuration used in the *Edible Audience* performance has been extensively discussed elsewhere (Barrass. 2006. Riddell 2006) and will be reviewed here only in respect to the impact on the sound performance.

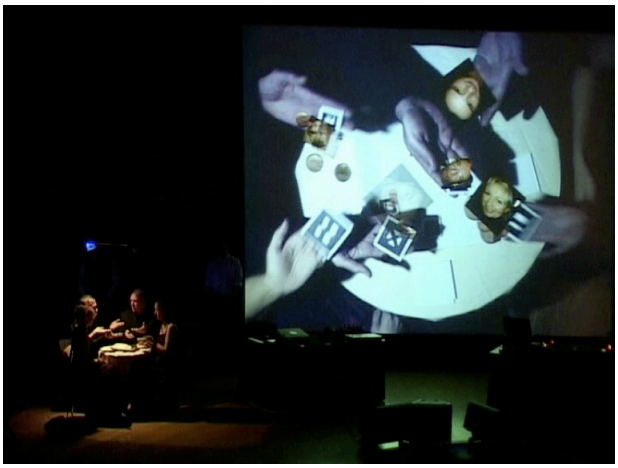

Figure 1. *Image from the Edible Audience performance at Liquid Architecture 6. Performers are on the left. The screen shows the fiducial markers with and without superimposed images.*

The performance involved six performers: four diners and two Waiters. The performance space was delineated by the tabletop[4], illuminated from above by a single light source (figure 1, lower left). The performance surface was then relatively small and intended to make both the performance intimate and *AVIARy* operation effective. The performers sat at the table and manipulated markers placed before them by the waiters. Each performer gestured with the markers in the spirit of an intimate, perhaps orgiastic, dining experience. The audience experienced this from a far or transfigured on the screen. The transfigured image on the screen included the addition of images of people's faces and body parts superimposed on the markers. Thus through extravagant gestures, the diners appeared to consume these images as part of the performance. The sound reinforced the visceral and rather surreal atmosphere of the theatre but was not itself directly imitative of the dining activity. If anything, the sound resonated in the space like weird restaurant

[4] This approach has precedents. See Berry et al. 2003.

Muzak, yet entirely consistent with the nature of the event as a whole.

As can be seen in figure 1, the performers are to the left of the projection screen. The results of the *AVIARy* processing were clearly visible to the audience on the screen but the actions of the performers were less so. To operate successfully, *AVIARy* needed particular lighting for the USB camera to identify the markers. This proved to be a slight problem during the performance. The lighting and camera set up also determined the performance space as the camera needed to be relatively close to the computer running *AVIARy,* which was placed out of sight of the audience. In retrospect it was felt that the 'dining table' should ideally have been place under the screen to optimise the audience's view of the activities. Such a configuration would not have had any impact on the sound and was also not possible due to logistical issues with the other artists in the festival who were also setup across the stage area.

Software and Control

The controlling application, *AVIARy*, based on the jARToolKit[5], analyzed live video input and searched for pre-defined patterns, called "fiducial markers". When these markers were recognized images were superimposed on the projection of the markers, thus hiding the original marker images (see figure 1). These images could be moved around with the underlying marker and had 6 degrees of movement: X, Y, Z, yaw, pitch and roll.

These degrees of movement were originally to translate into control over the sound. X, Y and Z were used for position and amplitude respectively but the mapping of control to yaw, pitch and roll proved to be too complicated for the performers to manage without extensive rehearsal. These were not used in the performance.

On recognizing a marker, *AVIARy,* apart from initiating image processing, also sent OSC datagrams over a network connection to the audio application *SuperCollider 3 (SC3)* running on a Mac PowerBook. The network configuration sought to reduce the computational load that would arise on one machine also running *AVIARy*. Thus sound and image movement were synchronised sufficiently for live performance.

However, the sound was controlled by *AVIARy* and depended on whether *AVIARy* identified the markers. Recognition could be lost if the marker was moved too fast or tilted on too great an angle. This would terminate the sound with a slight fade out. If *AVIARy* recognized the marker again it would retrigger the audio. It was thus possible to rapidly trigger sounds depending on the recognition rate of *AVIARy*. This dependency was not ideal for sound but any alternative would have been more complex and not synchronised with the overall augmented reality presentation.

SuperCollider Operation

SC3 received OSC data from *AVAIRy* for each unique marker. OSCresponders therefore handled unique markers for 4 diners for each course (16) plus markers for the identification of each course including final credits.

During an early development stage, it was discovered that *SC3* would not parse the large number of variables needed for all the individual OSC Responders and so a compromise was adopted using a smaller number of variable, which were shared by the OSC Responders. Such a solution was potentially dangerous because these variables were subject to alteration by other Responders if events were not terminated correctly. No doubt there was probably a better way to do this but as time ran out and the script approached 1500 lines, the compromise was accepted. However, there was an unexpected anomaly during the performance due to the Augmented reality system's false recognition of certain marker images. This caused sounds to play intermittently. Irrespective of this problem it remained difficult to perform with subtle control over dynamics and spatialisation while maintaining a theatrical pose. Perhaps for this kind of performance lack of subtlety was not a bad thing.

Sound Diffusion

Audio output from *SC3* was through ADAT lightpipe from a *MOTU 828 mkII* to a Behringer *DDX3216* digital mixer, which then sent the audio to a Behringer *Ultragain PRO-8* connected to the speakers. This complicated configuration was necessary to accommodate the other performers and despite some concerns about stability of the system, it worked perfectly for the entire event.

The sound used for the *Edible* performance (processed audio files and some real-time synthesis) was output through an 8 channel sound system, which included 2 sub-woofers. The speakers were arranged around the audience with the sub-woofers in front on the stage.

Consequently, the performers were only partially able to perceive their control over their sound. This was probably a significant setback as it was an ongoing challenge to understand where the sound was and what was happening to it.

To create an audible sense of homogeneity in the sound for each 'course', the same sample was used for each performer but the starting point in the sample was randomly chosen. In this respect, the collective sound would be slightly different in a way similar to canonical form. In practice the difference was not always enough to help the performer readily identify their sound. If a marker was intermittently recognized by *AVIARy*, the sound would restart accordingly which might have conformed to the chaos of the diner's gestures but made the logic of the sound hard to comprehend.

There was, possibly, too much sound happening as a consequence of continuous participation by the performers. This was driven by the imperative of making the markers visible all the time and so the sound was also present. The ability to have sound events with autonomous internal structures (generated sequences of events) triggered by the markers would have been, perhaps, more ideal. Exactly what the real-time control would be in this situation is another question.

[5] Software source and documentation - www.hitl.washington.edu/artoolkit/ (Referenced 17-03-06)

Performance Reflections

While the tabletop provided an easily recognizable gesture space for the 4 performers, it was also a complicated space to navigate with a view to manipulating sound. The X and Y coordinates were translated into points on a circle and then into positions around the 8 speakers but the result was not always coherent and audible to a performer while three other, somewhat similar, sounds were also being presented. As the performers could not reach to the camera for maximum amplitude, this level was adjusted to occur near head height and meant that the sound would not always be at maximum amplitude but vary was the performer moved the markers up and down. If the performer's marker happened to be lower than the others, then their sound would be quieter and less recognisable. Of course, this varied moment by moment as the performers jostled for positions in the performance space.

From an audience perspective, the vertical representation of the table top on the screen added to a confusion or failure to read the movement of the sound as defined by the table surface. To the audience, images of hands moving across the table did not readily equate to the sound moving around them. While the audience would have eventually, perhaps, understood the relationship between the movement of the markers and the movement of the sound around them, it would have taken considerable concentration to follow for some time the activities of one of the performers.

When considering the nature of the sound interaction and performance space I was acutely aware of the prospect of idiosyncratic interpretations by the performers who had little experience with this kind of context. The question is whether this was a problem and whether it had a negative impact on the event as a whole. In this respect Rebelo observes:

> A performance space, in the context of non-linear digital media structures, implies sophisticated analysis in the areas of gesture, one-to-many communication schemes, individual presence, idiosyncratic action, and instrumentality. The performing body operates in a space of expectation, in a space that tends 'towards more or less coherent systems of non-verbal symbols and signs'. (Rebelo 2003)

In the case of the *Edible Audience,* the performers where required to behave in a theatrical manner, evident in exaggerated dining gestures and a stylized ritual of consumption. I imagine that this theatre, while not entirely clear to the audience in the early part of the performance, became more understood as they accepted the performance translated into sound and image. This performance behaviour required on the part of the performers, a degree of concentration and forced a division of attention between the acting and sound control. In addition, the performers had to be aware of how they were manipulating the markers with respect to the image and thus the overall function of the augmented reality system. It is clear from figure 1 that *AVIARy* was not always able to recognize the markers given the way the performers were handling them. However, in figure 1, it can be seen that there are six markers visible but *AVIARy* is only recognizing four. Reflecting on this from a sound perspective, there would have been four discrete sounds present, although there were six markers in view. In fact, this is how it was set up in the *SC3* code and would have functioned according to that design. Yet this visual presentation is confusing. Clearly a problem existed here given that the demands on the marker control at a visual level superseded those for sound control. It was difficult to view the markers as part of a performance sound instrument after only a few rehearsals. So much emphasis had been placed on their visual significance to the *AVIARy* system.

Underlying Process

> I am not a great believer in explanatory/didactic art in which the artist makes sure the audience appreciates the process, I do try and make the process, or in some cases the constraints evident with view to how the work is presented. (Rebelo. Interview with Kyle Dickau)

Any collective intention of making the "process" visible in the case of the *Edible Audience* project can be answered in the negative. However, the project did assert a technological imperative as inherently part of its nature and design. Although, however present this technological aspect may have been, in reality consideration was prioritised towards problem solving and aesthetic matters. In this respect, the project concurs with Rebelo's principle and any didactic ambitions were thus not explicitly emphasised. Yet it cannot be overlooked that, by the very inclusion of a complex configuration of technologies, the audience were to perceive a pre-eminence of process driven by technology.

> The more we talk about collaborative work and inter-disciplinarity, the more I think interesting practice revolves around the differences and disruptions between disciplines, rather than in their merging. (Rebelo. ibid.)

The nature of the *Edible Audience* collaboration was multi-dimensional, beginning with the project concept from a chance discussion, through to the formation of the system and finally the performance. It is probably thus fair to say that Rebelo's comments resonate with the reality of the project where whatever "merging" took place, it was a fluctuating phenomenon over months, and in the final performance, sometimes evident and sometimes not. However, the collective aspiration of the artists was to convey a homogeneous experience rather than simultaneous and disparate sensory experiences. That the performance was constructed from activities that were not always conducive to a homogeneous ideal reflects the vicissitudes of this experimental project.

The Body in Collaborative Performance

The "performing body" while not a new concept to digital artists, is often a new experience that cannot be anticipated, necessarily planned for or the basic skills quickly acquired. Typically, the demands are different for each performance.

Personal experience (Riddell 2005) suggests that gesture control of sound in free space, while apparently not confronting in concept, is particularly challenging in practice. A comparative examination of the *Edible Audience* performance and earlier *HyperSense Complex* performance practice (figure 2) reveals some important differences in sound performance configurations.

Figure 2. *HyperSense Complex performing at "Glitch", Candy's Apartment, Sydney. November 2003.*

In common with both performance practices is sound control through arm/hand movement in free space. That is, a space essentially defined as variable or unconstrained in respect to physical movement. In the case of *HyperSense*, the articulation of the fingers controlled the sound. Position sensation and limitations of finger movement informed the performer about control of the sound. But frequent movement of both the arms and hands arbitrarily in space contributed to a sense that there was an optimal place to locate the hands such that movement of the fingers could be given primacy of attention. In this respect, the hand actually was the dominant unit if only from a cognitive perspective. If the hands were not comfortably located in space then finger movement was also compromised.

One needs only consider the piano keyboard as an example, to gain an idea of an optimal place for the hands. The keyboard provides access to a frequency series laid out from left to right. It is immediately appreciable that a comfortable range for performing lies in the middle region, directly in front of the performer. This correlates to the range of frequencies that are most often used in traditional musical composition. Extension of the arms and bending of the wrists to play at the extreme ends of the keyboard is physically uncomfortable.

Exactly the same experience applied in the case of *HyperSense*. A more comfortable placement of the hands is in a forward direction with the upper arm parallel to the body and the forearms raised slightly. The hands could, at times, be rotated in an expressive manner to intimate an emotional relationship with the nature of the sound being produced.

The performers in *HyperSense,* however, were also aware that the placement and movement of the arms facilitated a connection between performance action and the audience. It was therefore important to exaggerated movement of the arms even though it was not conducive to the best control over the sound. Moving the arms also relieved strain on the fingers. Interestingly, even though the fingers moved without restriction, a long period of conscious and precise control and articulation of the fingers seemed to bring on a noticeable sense of strain.[6]

In contrast, the gesture movements required in the Edible Audience performance were specifically arm and hand. The difficulty began with picking up the markers and holding them in a manner that could be presumed to be visible to AVIARy (we were not to look at the screen). This required that either the marker lay flat on the fingers or that it was held on the edges so as not to obscure the pattern. Exaggerated movement under those conditions, particularly if the arm needed to be elevated, was awkward and often self consciously indecisive due to the negotiation of the other performers' hands and arms.

Under these conditions, optimal sound control was not possible. During the development of the sound control system it was becoming increasingly apparent that subtlety and nuanced performance gestures where not going to rapidly evolve in such a short time. And they were perhaps not desirable from a theatrical point of view. As a consequence, sound control development began to focus on fewer control parameters and those with the most visual impact. In this respect, the theatre of the performance began to impose demands that paralleled those for the visual dimension but were also not entirely complementary.

What is required for sound control is an understanding of the defined boundaries of action and an unambiguous awareness of the resulting sound. In traditional instrumental practice the musician has, at least, a tactile relationship with the instrument, which should concur with the sound produced. In electronic music, this is possible in some cases but not all and there can be cases (as in this instance) where such conditions cannot apply.

Conclusion

As many sound artists appreciate, a visual component within a performance project can make significant demands on a performer's attention unless configured in such a way as to balance the functional relationship between sound and image. This suggest taking a simpler approach to the outcome but not necessarily to the technology involved, as long as it is not the primary point of the event.

It is possible that the *Edible Audience* project contained, along with a strong visual/theatrical element, too many metaphors, that the audience had to interpret from their sensory experience. In other words, there was too much going on and a sustained synergy of the experience was too challenging and ultimately elusive.

[6] It was a relief to return to a more natural (less self conscious) movement of the fingers after a performance.

This complexity was perhaps due to an over enthusiasm for the project on our part and led to an overly ambitious realization. It is, however, interesting to consider that the project embraced and was effectively defined by the concept. To have made it simpler or less complex, even reduce some of the expectations in the presentation format, would possibly have diminished the motivational energy required to work on the project. In addition, many works of this nature are highly experimental and there is a desire on the part of artists to pursue the logic of experimentalism to the boundaries of the work and in the time frame allotted to its realisation.

Where the author's reservations primarily lie is in the collective attention to the sound. Certainly the project required the skills of a diversity of artists, in particular, those with visual and technical skills. However, the performance might have been better, from a sound point of view, if all the performers had formidable prior experience with sound in live performance.

Given that this was not a necessary condition of the project, an alternative might have been to simplify the sound world and present a minimal palette of sound objects to be performed. Nonetheless, this approach would still have required conscious control and performance skills on the part of the performers.

The other tangible shortcoming was the lack of adequate rehearsals. While each rehearsal shaped certain operational matters, they did not mature or clarify the performance practice substantially. Certainly not where the sound was concerned. The rehearsals/testing phases did not really prepare us for the final performance. Granted in practice, each performance space has its own peculiarities when it comes to experimental performance. Adequate rehearsal time in a performance space just prior to the performance is rare and therefore the possibility of unforeseen problems is higher.

However, probably more important than extensive rehearsals for one performance would have been earlier performances in different venues under different aesthetic contexts. If there had been several performances prior to LA6 then problems with the concept would probably have been ironed out well in advance.

Finally, it would be wrong to conclude that the *Edible Audience* project, as discussed here, should not have taken place. In the context of the festival, the project had a distinct and highly engaging character. Negotiating its complexities was a significant learning experience in itself, creating for the performers, confidence to consider and undertake future projects of an equally challenging and rewarding nature.

Acknowledgements

Thanks go to fellow *Consumers*: Stephen Barrass, Tim Barrass, Anita Fitton, Peter Morse and Onoclov. Special thanks also to Lea Collins for sound engineering the festival.

References

Barrass, S. 2006. "Music in Collaborative Virtual Environments" *Virtual Reality*. eds, Léonie Schäfer and Steffi Beckhaus. London: Springer (forthcoming).

Berry, R. Makino, Mao. Hikawa, Naoto and Suzuki, Masami. 2003. "The Augmented Composer Project: The Music Table" Proceeding of the *Second IEEE and International Symposium on Mixed and Augmented Reality*.

Rebelo, P. 2003. "Performing Space" *Organized Sound*. 8, 2. 181-86.

Rebelo, P. *Interview with Kyle Dickau*. ed, Sylvia Borda. www.ontherundesign.com/Artists/Pedro_Rebelo3.htm (25th March 2006).

Riddell, A. 2006. "Edible Audience: What about this Gastronomic Performance Translated as Data Art?" Proceedings of the *Sounds in Translation* conference. E-Press. The Australian National University. (forthcoming).

Riddell, A. 2005. "HyperSense Complex: An Interactive Ensemble" Proceedings of the *ACMC'05* conference. QUT.