

INTEGRATING TEXT, SOUND, AND VISION IN AN INTERACTIVE AUDIO-VISUAL WORK

Roger Alsop

Victorian College of the Arts, Melbourne University,
School of Creative Media, Royal Melbourne Institute of Technology

ABSTRACT

This paper describes the integration of text, sound and vision in the development of an interactive audio-visual-textual work titled *'Yelling at Stars'*. It explores points of intersection between text, vision and sound and possible ways in which the performer and viewer can draw relationships between physical, visual and sonic gestures, developing links between each of those elements and seeing each as having equal prominence in the work.

The paper then outlines the algorithmic processes used in developing the work, in particular the considerations made when developing the web of interdependent links between text, vision and sound.

1. INTRODUCTION

This paper discusses *Yelling at Stars*, as an approach to developing Dannenberg's ideas to include spoken and written words as part of the aural and visual performance. Roger Dannenberg approaches "music and video as expressions of the same underlying musical "deep structure." Thus, images are not an interpretation or accompaniment to audio but rather an integral part of the music and the listener/viewer experience." (Dannenberg 2005)

Yelling at Stars is a collaborative work devised with amateur Willoh Weiland. The project uses sound and music, written and spoken text, visual images, and actors in a performance about communicating beyond the stars. It is designed so that each of those elements has equal weight, both in development, composition of the work, and consequently in performance. A main concern when making the work was creating a synergy between those elements, linking attributes of each, and then using those linkage points as the fulcrum around which each of those elements revolves.

It was decided that each element be a causal agent when performing the work, with a physical gesture beginning the performances. From there the performer and audience interact with the visual, textual and sonic elements, improvising within a prescribed structure. For example; the computer algorithms respond to physical gestures of the actor, and the actors then respond to that input. This process also allows for an interactive installation to be generated from the original work, in which the audience may take a generative/creative role in the work through similar interactions to those of the actor.

2. INTERSECTIONS IN TEXTUAL, VISUAL AND AURAL ART

The gesture, in all its definitions, may be seen as a fulcrum around which much art gains its affect on the audience. The following examples indicate possible links between these modalities without trying to make concrete comparisons or define methodologies. In visually oriented art, such as painting or dance, the gesture can be seen in the stroke of a brush in static visual art, the movement of an image in film art, or the movement of a body in dance; in conceptually oriented art, such as textual art, it can be seen as the motion between ideas; and in aural art, as movement between frequencies. W.J.T. Mitchell defines it well in Lia Markey's essay *gesture*: "Perhaps gesture is best understood as the moment when thought becomes visible, tangible, or palpable, staged and framed as form - something to be held and to hold us in mutual prehension."(Markey 2002)

Here I will be focusing on gesture in the temporally based art forms: film, dance, writing, and music. Each of these art forms requires time for its expression; the authors creating the sequence and flow of events, defining how the ideas contained within those events are revealed to the audience.

In each of these art forms there are well known processes and structures: music forms such as 'sonata', 'rondo', 'ternary' and so on; the textual forms such as the 'well made play', 'Mills and Boon' and the poetic forms such as 'sonnet', 'haiku', and 'mesostic'; and the filmic forms 'noir', 'action', 'slasher', or 'dogma'. These methods of presenting ideas to an audience have two vital functions, they provide a structure for the artist to work within, and they allow the audience to predict the type of ideas contained in the work and how those ideas will be revealed. If the form is known, either consciously or unconsciously, the audience can concentrate more on the ideas themselves, and less on trying to navigate the relationships between them.

Musical form has possibly the most overt and easily seen structures. In most cases these were designed at a time when the audience was unable to review a work, requiring that the developments and trajectories of the piece be clear to the audience on one listening. The sonata, A A' B A, form, where an idea is presented, repeated, varied and then presented again, is a well known method of presenting ideas, not exclusively musical, for the audience to easily learn and follow. Marvin Minsky refers to the form as a "teaching machine" saying that the sonata first explains "one idea, then another, and then recapitulate[s] it all". (Minsky

1981)

Novelty in any artwork is often seen in the way the author negotiates and navigates the form(s) being used. In order to effectively communicate, the audience must feel that their interest is piqued, and that they want to offer their attention, but not to the point of requiring too much effort. By following understandable forms these goals may be achieved, thus allowing the novel elements of the artwork to be appreciated.

2.1 Gesture as an element in textual, visual, and aural art

Examples of the gesture in text can fit each of the areas being considered: the visual (through the shape of the text on the page or screen), the sound of the word and the sequence of words used to express and develop an idea.

John Cage's *62 mesostics re Merce Cunningham* demonstrate links between spoken and written text, sound and visual images. Here various fonts and font sizes are used, and the words have some meaningful relationship to Cunningham. Cage says that the visual representation of the text on the page "may be used to suggest an improvised vocal line having any changes of intensity, quality, style, etc., not following any conventional rule." (Cage 1971) See Figure 1.

QuickTime™ and a
TIFF (LZW) decompressor
are needed to see this picture.

Figure 1: Mesostic 59 from *62 Mesostics re Merce Cunningham*

The words "taketa" and "naluma" demonstrate the way in which people draw a reference between sound and vision. Here most people take "taketa" to refer to the more jagged shape and "naluma" to refer to the more rounded shape.

This could be for a number of reasons, one being the way in which the two words are made in the vocal tract, "naluma" being glides, nasal, and labial, and "taketa" being stops, both glottal and dental, and plosives, influences the way in which the words are heard at

an emotional and instinctual level. The forward motion of the lips and tongue in "naluma" engendering confidence and openness, and the backwards motions in "taketa" engendering hesitance and offensiveness or defensiveness. See Figure 2.



Figure 2: Images of "naluma" and "taketa" (Lexicon 2007)

Bettina Knapp discusses the emotional qualities of word sounds in her discussion of James Joyce's short story *Eveline*. Knapp sees certain word sounds as having more than their lexical meaning. Knapp sees that Eveline's experience is "accentuated by [Joyce's] complex [musical] system of figures of speech and by his use of stressed and unstressed phonemes". An example is seen in the sentence: "Eveline sat at the window watching the evening invade the avenue." here "Eveline", "evening", "invade" and "avenue" encase "window" and "watching"; here Knapp contrasts the /w/, to which she ascribes "an airy, breezy quality, indicative of the need to move about", with the /v/, to which she ascribes opposite qualities. (Knapp 1988; Alsop 2003)

"Naluma" and "taketa" can also be seen as larger movement gestures than those of the vocal tract. "Naluma" could be seen as representing flowing movements and rotational forces in a choreography, and "taketa" as representing jagged movements and linear forces.

Introducing his book chapter Meaning in Musical Gesture Fernando Iazzetta quotes Boethius: "How does it come that when someone voluntarily listens to a song with ears and mind, he is also involuntarily turned toward it in such a way that his body responds with motions somehow similar to the song heard?" (Iazzetta 2000). The way in which music can inspire informal and unpremeditated movement indicates a subconscious link between aural input and the physical gesture. This relationship can easily be seen in the startle reflex in infants, a simple movement of bringing arms and legs towards the chest when loud, unexpected noises are heard that can also be seen in adults in the same circumstances. When more sonorous and expected sounds are heard, such as when listening to music, Boethius's body movements are often seen and can be quite predictable, for example the raising of the chin when logically ascending melodies are heard. The gestures of a conductor and the rock guitarist are another example of a direct mapping between sonic and visual gesture.

The above discussion outlines some of the ideas that informed some of the processes used in developing the algorithms used in the *Yelling at Stars* project outlined below. The intersections between the sound of words, the meanings that are instinctively

ascribed to words and word sounds, the shapes of images, the physical motion that shapes can infer, and the way that sound and music can generate physical gestures in performers and in listeners, all informed the development of the algorithms.

3. THE ALGORITHMS

The algorithms used in generating the relationships between the sonic, textual and visual elements are developed in the Max/MSP/Jitter graphical, object based programming environment. It allows rapid development of programs using provided objects and objects made by third party developers. This broad and integrated palette from which to build gives the programmer the opportunity to create systems through which to experiment and explore ideas that can integrate and effect visual and sonic data.

3.1 Developing the algorithms

When developing the algorithms five elements were considered: visual input, used to map physical/visual gesture; textual input, both spoken and written word which is used to set the semantic/conceptual context; and the textual, visual, and sonic output.

The visual input creates the expression of the piece by mapping the gestures of the performer or participant to spoken and written elements drawn from a database, allowing the performer/participant to improvise within a set framework. The mapping of the actions of the participants takes two forms, their motion through a set of quadrants, and the speed within each of those quadrants.

This approach was taken after seeing participants interacting with a 'dancing machine' installation I created. Here the movements of participants caused bass, drum, guitar, percussion, and so on, riffs to be triggered, forming the sounds of an entire band when movement was detected in each of eight quadrants. This allowed participants to learn the results of their actions, so that after about three or so minutes they were able to control the sounds that were produced, forming their own mix as they danced. The quadrants were taken from a visual input from a video camera placed above the 'dancing area', as shown in Figure 3 below.

Drums 1	Guitar 1	Bass 1	Keyboard 2
Bass 2	Keyboard 1	Drums 2	Guitar 2

Figure 3: The positions of various sound groups in the 'dancing machine' installation.

This positioning of sounds allows the participant to create a four-piece band by moving in any of the three central crosses, and the riffs played by each quadrant change if no motion was detected after a certain time. Similar processes are used in the 'Yelling at Stars' algorithms, where motion through different quadrants triggers images, texts and sounds, and the amount of motion in any

one quadrant influences the speed at which images are presented and their placement on the screen.

As different images, sounds and texts are presented the actor/participant improvises, in generating the output. As the piece develops the role of the algorithms becomes more dominant and the role of the actor as a causal agent wanes. Through this process the actor becomes an equal agent to the sonic, visual and textual elements in the performance.

Willoh Weiland wrote the text, the narrative following the trajectory of someone wanting to generate a relationship with the unknown, using the extra-terrestrial as the metaphor. The text as presented in performance loosely follows the text as written, but allows for interaction from musical, physical and visual gestures to influence the ordering and position on screen. For example, when the amount of motion increases the motion of the text presented on the screen may also increase, or the separation between semantic areas of the seen text and the spoken words increases.

The music, sound and visual images for *Yelling at Stars* are also affected dynamically in performance; here the gestures (both the physical ones of the actor, visual ones of the images on screen, and the semantic ones of text, and text font and size) influence the sounds being heard, just as those sounds influence the visuals.

The main goal in developing the algorithms was to ensure that an interactive and interdependent web between all elements was achieved. To this end a web of data pools was created that contained the text, visual images, and sonic elements. Each of these pools was filtered through appropriate effects procedures, such as visual or sonic granulators or amplifiers, or font size and style manipulators. An example in performance might be that as the text image increases in size the pitch of a sound may decrease and the spoken text may become more semantically fragmented.

4. CONCLUSION

The relationship between what an audience may view as three distinct elements, text, sound and visuals, can be blended as the audience develops an understanding of the interaction and interdependence between those elements.

At time of writing one performance to an invited audience had occurred. In this performance much of the interactiveness built into the system had been removed in an attempt to gauge the degree to which the audience could apprehend and parse the wide varieties of input.

In general, responses to the work indicated that the audience had an instinctual understanding of the relationships between the various elements. It was also indicated that the opportunity to develop the work to a higher degree of interaction from the audience should be taken.

As *Yelling at Stars* is further refined, a greater sense of the relationships and potential mapping between the various elements will develop, allowing for more predictable and concise communication between author, performer, and audience to be achieved.

1. Alsop, R. (2003). The Ineluctable Modality of the Audible: Exploring the sound worlds of James Joyce's Ulysses. Acoustic Ecology: An International Symposium. Melbourne, Australian Forum for Acoustic Ecology. <http://www.acousticecologyaustralia.org/symposium2003/index.html>
2. Cage, J. (1971). SIXTY-TWO MESOSTICS RE MERCE CUNNINGHAM. New York, Henman
3. Dannenberg, R. B. (2005). "Interactive Visual Music: A Personal Perspective." Computer Music Journal 29(4): 25-35
4. Iazzetta, F. (2000). Meaning in Musical Gesture. Trends in Gestural Control of Music. M. M. Wanderley and M. Battier. Paris, Ircam - Centre Pompidou.
5. Knapp, B. L. (1988). Music, Archetype and the Writer: A Jungian View. Pennsylvania, The Pennsylvania State University Press.
6. Lexicon (2007). Sound Symbolism, Lexicon. <http://www.lexicon-branding.com/process2aSound.html>
7. Markey, L. (2002). gesture, chicagoschoolmediatheory. http://www.chicagoschoolmediatheory.net/glossary2004/gesture.htm#_ednref6
8. Minsky, M. (1981). "Music, Mind, and Meaning." Computer Music Journal 5, No 3(Fall) <http://web.media.mit.edu/~minsky/papers/MusicMindMeaning.html>.