SONIFICATION AND VISUALISATION OF NARRATIVE: GETTING COMPUTING STUDENTS TO THINK AURALLY AND VISUALLY RATHER THAN AUDIO-VISUALLY

Paul Vickers

Northumbria University, School of Informatics
Pandon Building, Camden Street
Newcastle upon Tyne, NE2 1XE, United Kingdom
`paul.vickers@northumbria.ac.uk`

ABSTRACT

In this paper we describe a final-year undergraduate honours course that requires multimedia computing students to create narrative auditory and visual displays. The aim was to use discontinuity as a catalyst for creativity by getting students to tell a story in non-speech sound and in vision in which each medium conveyed a different mood.

1. INTRODUCTION

In his paper presented to ICAD 98 Somers [1] discussed his experience of teaching visual design based on the visualization of sound (and vice versa). Somers had design students produce visual representations of electro-acoustic sound clips, and got sound-composition students to create sonic representations of pieces of abstract art. The aim of both was for the student to perform structural analysis of the object in question and then to render the object's individual components and their structure in sound or vision. Somers found that the semiotic transformations from sound to vision (and vice versa) resulted in students producing work with better structural unity. Furthermore, the students improved their analytical skills through learning how to decompose visual and aural objects into their individual elements.

An interesting aspect of Somers' experiments is that by highly constraining the context of the design work, by asking the students to design representations with very specific boundaries, it was possible to improve the creativity of the students' work. Giving people a completely free reign (a blank canvas, if you like) seems to serve to limit creativity. The fact that the thing being represented was from an unfamiliar domain added to the contextual constraint.

Somers concluded that modern computing environments make it easy for designers to work in both the auditory and visual domains and that a bi-modal "*pedagogy may be an appropriate direction for design education in the 21st century*" [1].

Fishwick [2, 3] coined the term *aesthetic computing* to describe the application of art theory and practice in the design of computer systems and artefacts. The International Community for Auditory Display has been witness for more than a decade to the bringing together of artists and engineers to create useful interaction paradigms (e.g. see [4]). Such aesthetic computing is likely to continue to inform the design and construction of computing artefacts. So, spurred on by the recognition that computing students need to become cognizant of, if not literate in, design aesthetics, we were inspired by Somers' to apply some of his principles to multimedia computing students in an attempt to get them to think about their multimedia artefacts in a different light.

2. BACKGROUND TO THE WORK

The School of Informatics at Northumbria University runs an undergraduate BSc programme in Multimedia Computing. This is largely a computing science programme with an emphasis on multimedia technology (as opposed to a multimedia art and design course). Students on the programme have little, if any, background in music, art, or graphic design, and take courses in databases, computer programming, and data structures, as well as studying the production of multimedia material. Because the first year of the programme and much of its second year are common with other computing programmes in the School, by the time they reach their final year the Multimedia Computing students have a very strong traditional computing technology background. When the first cohort of students reached their final year in 2002/03 they had to take the core module CG087 Time-based Multimedia Assets.

The overall aim of this course is to extend students' theoretical understanding and practical skills in generating assets for multimedia, specifically in the time-based media of sound and video by:

1. Developing their abilities to originate, synchronise, and integrate time-based materials for use within a multimedia computer environment;
2. Providing them with an understanding of cross-media issues;
3. Providing them with a critical understanding of design techniques and technological issues relating to time-based assets;
4. And giving them an appreciation of the potential of time-based assets as communicative and accessibility tools.

Up until this point in their studies the students had tended to view multimedia objects as colourful (usually animated) things that make sounds (usually on a web page). Their cultural experience prior to coming to university was one of television and the movies. In their experience, the audio-only media (e.g. radio and CDs) were mostly used for music transmission. Thus sound tended to be either music or the soundtrack to a moving image.

We took the opportunity on this module to get the students to think about sound and vision as discrete and separate communication media. At the end of the course we wanted them to be able to communicate complex ideas, narratives, and concepts through non-speech sound alone and through vision alone. This would deepen their understanding of the semiotics of

aural and visual communication and should lead to a greater understanding and appreciation of how sound and vision interact to create experiences with emergent properties not evident in the two modalities separately.

3. A WALK IN THE PARK

The module is assessed through a single individual assignment that comprises a practical task and a reflective report. When designing the assessment we took the opportunity to get the students to think about the creation of audio and visual multimedia assets in a way that was new to them. We wanted to get the students to take a step sideways and reconsider their use and understanding of audio and visual media separately. Because the students were so used to thinking about multimedia as audio-visual presentations we chose to focus on the design of designed sonic environments (DSE) [5] and designed video environments (DVE). DSEs and DVEs are single-modality artefacts that aim to represent a story, a situation, or a scenario and to communicate information about the meaning of and the events in those scenarios.

The students were asked to design and construct a DSE and a DVE to act as narratives of a scene lasting between 75 and 90 seconds. Furthermore, the assets' durations had to be within 5 seconds of each other in order to allow a discussion of effects of discontinuity and complementarity (see section 3.2). The same basic scenario was to be used for both assets, but with one important difference: the mood. That is, the DSE and DVE had to explore the story in a different mood or emotion. For example, they might choose to tell a story about going for a walk in the park. The audio version could communicate a happy, carefree stroll on a bright spring day whilst the video suggests being tired and grumpy because it's cold, windy, and raining, and they're only going through the park as a shortcut.

Prior to embarking upon their designs the students were introduced to the fields of basic acoustics, auditory display, film language, and foley art. We explored program auralisation, auditory icons, earcons, auditory illusions and psycho-acoustic effects (e.g. Shepard-Risset tones and the cocktail party effect), and the use of sound as a communication medium both in the computing world and in films. As an example of the latter we looked at the famous shower scene in Hitchcock's *Psycho*, for which the screenplay has the following direction:

> "Naked, defenseless, and in shock, Marion screams while trying to ward off the assault with her arms. But the attack is relentless. The scene is intensified by *searing background music that seems to slash and scream*" [6].

The use of music as an auditory metaphor or symbol for the stabbing of the knife and the sense of danger and horror is very effective here.

To help create their DSE and DVE students had access to digital video cameras, Cubasis VST4.0, CoolEdit Pro, and Adobe Premiere.

3.1. Designed Sonic Environment

Using a mixture of MIDI and digital audio students had to create a designed sonic environment (DSE) to act as a narrative of a chosen scenario with a specific mood. Students were asked to think about what ideas and events they wanted to communicate and how this could be accomplished. They were allowed to use any combination of music (think of film music), auditory icons, sound effects, and earcons.

Somers [1] relates McCallion's experience [7] of encountering stereotypical thinking in design projects. Stereotypical thinking involves taking familiar ideas and restyling them slightly to create a new design that is not very different from what went before. To try to steer the students away from this we insisted that their DSEs must not rely on analogic sounds. Instead we wanted them to focus on metaphorical representations wherever possible. For instance, if they were representing a walk in the countryside then just playing back sounds of birds and cows would not be acceptable. Some analogic sounds to provide initial context to the piece were permissible but beyond that they needed to think very carefully about the metaphoric representation of mood, objects, and events.

3.2. Designed Video Environment

The DVE could be composed using a mixture of still and motion video. The students needed to think very carefully about how to capture the mood of the piece without using any sound. create a designed video environment (DVE) to act as a narrative of your chosen scenario and mood. The use of visual metaphors became especially important once the audio channel was removed.

3.3. Modality constraints, discontinuity, complementarity, and accessibility

Once the DVE and DSE had been created the students were asked to discuss how the constraint of working in a single modality affected their design decisions. They were asked to consider:

1. What limitations did this impose on their work?
2. What would they have been able to do better if they had been able to design an asset that used both sound and vision?
3. What aspects of the narrative were particularly easy or difficult to convey in each modality?

With recent changes in the law around the world regarding disability and access to information the students were required to discuss how this issue impacts upon the multimedia designer. In addition, because they were working with audio and video it was possible that they had used symbols that are culturally specific. Therefore, they also needed to discuss their work in relation to culture, internationalisation, and social constructs. By this the students were able to gain deeper insight into how audio and visual communication is interepreted/misinterpreted around the world.

Having considered the DVE and DSE in isolation we then asked the students to play both assets alongside each other thus merging them into a single audio-visual work. They then discussed what effects of discontinuity and complementarity emerged when the two assets were merged in this way. The following questions were given as discussion pointers:

1. Describe how the two assets give a sense of discontinuity. They were each designed to convey different moods. What effects does this discontinuity have on the whole experience? Does it lead to unexpected results? Does either of the assets take on different meanings at any point with the addition of the other modality?
2. Describe how the two assets complement each other. That is, how does the sound enhance the video (and vice versa) to create a more complete experience?

4. DISCUSSION

On the whole the students accepted the challenge of the task very enthusiastically. The requirement to think in terms of audio and video alone spurred some imaginative thinking, though, as discussed later on, it became clear that some students were not able (or, perhaps were unwilling) to design true stand-alone DSEs and DVEs.

4.1. Communication of narrative

With the recent emphasis on distance and on-line learning technologies the study narrative as a communication form within computing environments is becoming increasingly popular. Furthermore, as Back [8] points out, narrative takes auditory constructions on a step from metaphor by incorporating a temporal element. The inclusion of time means that behaviour can be modelled and so narrative, or story-telling becomes possible.

Whilst story-telling may not seem relevant to computer science students, we can view many interactions (both human-computer interaction and computer-mediated human-human interaction) as narratives. According to Back [8] "…the desktop 'metaphor' does operate as a narrative in some instances: the animation sequences on the opening or closing of a file, folder, or application is one instance of this." Thus by taking the student back to a familiar idea of narrative (telling a 'real' story) and working with them to communicate that narrative using a single modality, the stage is set for them to be able to transfer these ideas into multimedia and interaction designs. As Back says

"[by] conceiving of the sound design task as telling a story, the sound designer can make a more detailed analysis of needed sonic detail and decide on appropriate technical approaches. The finished design will thus present a richer and ore consistent feel to the user " [8].

That said, the idea of narrative structure was problematic for some students. Probably because they were computing rather than design students, their main concern was often the technical creation of well-produced sound, or video with lots of effects and fades. This led to some nice sounding DSEs that had no discernible story line and DVEs which had more effects than story line[1]. Typically, those who had trouble with creating narrative flow found the sonic narrative to be more problematic to evoke than the visual. This is probably because they are used to seeing pictures tell a story whereas audio narrative is generally done through speech (e.g. sound tracks and radio plays); the concept of non-speech narrative was unnatural. These students either focused wholly on trying to create a sense of mood or atmosphere without reference to narrative flow, or else they fell into the trap of producing very coarse direct and analogic representations of events; these event sequences came across less as a story line and more as like a series of unrelated or random pictures in a slide show. Even where there was an identifiable event sequence students often failed to manage the transitions from one sonic or visual event to the next. Sometimes, the change from one sound or picture to another was so jarring that it resulted in a complete loss of context.

Those who did engage with the task of creating a sonic narrative also fell into two broad camps: those who tried to create a strong sense of story line and through metaphoric

representation of events (typically using sound effects and auditory icons) and those who took a more traditional movie soundtrack-type approach. The 'soundtrackers' would stitch together samples of found music (often orchestral) in an attempt to generate a sense of movement and mood. Whilst the exact story line was much less obvious (or should we say, open to more interpretations), there was a clear sense of drama, mood, and atmosphere. Furthermore, when listened to alongside their DVE the overall effect could be quite powerful. This suggests than many students tackled the creation of the DSE and the DVE as a single project where both assets were really designed to be experienced together rather than in isolation. Not many really managed to create DSEs that were genuinely interesting in their own right.

4.2. Single modality constraints and the metaphor

As part of their submission students were required to comment upon how they reacted to the constraints placed upon them by having to work in a single modality. For people who have grown up in an audio-visual world the challenge of telling a story using only pictures for the DVE and only non-speech audio for the DSE was stimulating. Students reacted well to the challenge and often commented on how they had to change their way of thinking about how to communicate ideas. In one example, a student put a desk and chair on an empty beach to portray a sense of isolation and abandonment following an intensive computer-game-playing session. Ordinarily he would have used a combination of dialogue and mood music to support a more realistic video sequence. Without audio he was forced to consider metaphors and the use of film language to get his ideas across.

The metaphor is a central feature in many auditory displays with certain timbres and motifs standing for interface objects and data. We hope that by thinking about metaphoric representation in such an overt way in this course the students have gained an appreciation for the use of metaphor in interaction design more generally.

Through working in a single modality students began to appreciate just how powerful and flexible sound and vision are as communication media. For many, this course was the first occasion where they had to stop and actively listen to sound rather than 'simply' hearing it as one facet of some larger experience.

Many students reflected on the difficulty they faced in choosing non-analogic sounds to portray key objects and events in their DSE. The temptation to rely on bird sounds to carry the whole ambience of a park was great. Some managed to compose or find appropriate mood music to enhance the experience. A frequent theme was the use of a metronomic pulse to signify a heartbeat, itself a metaphor for the emotional state of the main character. Others used cultural references to situate the DSE. In one example, a student used a treated sample of the theme music from a UK TV ski programme to place the story in an alpine setting. To those not exposed to British television in after the late 1970s this reference would likely go unnoticed.

The constraint of silent video led to a several submissions that lacked a clear mood. The film soundtrack is such a direct communicator of mood that the use of visual cues and metaphors for this purpose had gone unnoticed by some students. On the other hand, some managed to use film metaphors such as lighting very effectively. One student reported spending three days trying to get the correct ambient lighting for his outdoor piece. It was noticeable that those who availed themselves of the tutorial sessions devoted to setting up shots,

---

[1] In one case, a video wipe was used every ten seconds, or so, which rather than create a sense of movement between scenes just served to irritate the viewer.

lighting, and composition fared better than those who relied on Adobe's After Effects software to do the work for them.

4.3. Discontinuity and complementarity

One of the most interesting aspects of the work was when the students commented on what happened when they played their DSE and DVE alongside each other. Both artefacts were telling a narrative in the same scenario but with different moods. We asked the students to discuss two issues:

1. How the discontinuity of the two environments affected their overall perception when played simultaneously, and
2. How the DSE and DVE complemented each other.

Students discovered that the sound and vision often interacted in unexpected ways that changed the entire meaning of what they were watching. In one example, a student created a comical animation of a spaceship taking off from earth, flying far out into space, landing upon a new planet and claiming it for the USA at which point a funny-looking alien appears and tells the astronaut to leave. The astronaut gets back into his ship and flies home. When the DVE is viewed in isolation it is very redolent of 1960s cartoons and has a great sense of fun. The metaphor of repeatedly showing the space ship moving from the bottom of the screen to the top gave a clear sense of movement and pace. The overall mood came across as jolly, adventurous, and lighthearted. The complementary DSE, on the other hand, portrayed a great sense of awe. This student was one of the 'soundtrackers' who took orchestral arrangements to create a tapestry of the unknown with some majesty (much like the composers of the various modern Star Trek TV franchise theme tunes have attempted to do). Both the DSE and the DVE were enjoyable in their own right, but when played simultaneously created a whole new experience brought about by the discontinuity of the different moods of the two artefacts. Now, when accompanied by the majestic and awe-inspiring soundtrack the mood of the once comic video changed entirely. Now, instead of a lighthearted trip through space with the denouement of the comic alien at the end, the mood became sinister and threatening. A sense of unease settled on the piece as the DSE transformed the meaning of the DVE.

Students would sometimes find that sonic and visual events that they thought would clash actually complemented each other. In fact, it was usually the case that when this happened the sound complemented, or enhanced, the video rather than the other way around. The most pleasing examples of complementarity were those that were unintentional. A few students story-boarded their DSE and DVE so carefully that the events in each were synchronized exactly. In these cases, although the sound and vision did work together it was not as pleasing as the serendipitous effects when non-planned interactions took place.

5. CONCLUSIONS

Some students observed that seemingly incompatible audio and video sequences would work together in a complementary manner, or would work together to remove the meanings of the individual artefacts by creating a new emergent overall mood or feeling. Filmmakers and those working in the media have long been aware of some of the interaction effects between sound and vision. However, software developers not normally educated in these matters could easily and unwittingly communicate unintended messages to users through the addition of auditory displays to interfaces. Such interaction effects need to be studied more formally with the aid of artists, designers, composers, and sound designers to come up with some guidelines appropriate to the further development of auditory interfaces and displays. Even well-motivated and well-intentioned auditory displays that aim to complement a graphical visualisation may, in fact, change the meaning of the visualisation.

Both the planned and the serendipitous audio-visual interactions observed in many of the combined DSEs and DVEs show how important it is to get the design of sonifications correct when they are going to be used alongside visualisation. The unplanned complementarities and discontinuities are the most instructive in highlighting the need for designers of auditory displays (especially those that use musical forms and structures) to seriously consider and explore the interaction effects with any visual displays.

If nothing else, it was great fun to get computing students not only to indulge their love of building things, but to do so in such a blatant artistic and creative context. The design work was constrained by having to work in a single modality. However, some students failed to address the issue of narrative flow, and their stories were very dull with few events. On reflection it might have been better to further constrain the students by providing a given story as the context for the work (much as McCallion suggested for architecture students [7]). Having a given narrative structure would mean they could focus their efforts more on the representational aspects rather than spending effort writing a good story. We are considering getting next year's students to design a DSE and a DVE for a given poem (or perhaps some haiku). A poem provides metre and rhythm as well as a sense of narrative (though one which may be open to many interpretations), and so offers a highly contextualised and constrained environment within which to work. That said, some of the best submissions came from students who not only managed to create effective DSEs and DVEs but who also came up with engaging and interesting story lines.

The poor transitions in some submissions between events in the narratives which resulted in a jarring effect (and even a loss of context) exemplified what can happen when a non-linear narrative (as happens in dialogues with a computer interface) is sonified. Back commented that it is hard "…to create interactive narratives because…artistic control of … [the shape and ordering of events] … is weakened or absent, with control being given instead to the user" [8]. This exercise serves to show how important good transition between the representation of events is and how difficult it can be, especially when the precise order of those events cannot be predicted (as would be the case in many auditory display environments, e.g. program control flow in a program auralisation). Vickers and Alty [9, 10] showed how sonifications based upon musical grammars can be especially effective at maintaining context and smooth flow between events in a non-linear narrative.

Exercises of this type showed students how they can use sound and vision as separate communication streams, but that design decisions for one impact upon the perception of the other through interaction effects. Research needs to be undertaken to see how aesthetics can be properly addressed by the designers of auditory displays to ensure that the sonifications and visualisations not only work well on their own but that they work together, and with the intended effect.

A greater use of people trained in the graphic and sound arts will enhance the work of sonification designers and this interdisciplinary work needs to be strongly encouraged.

6. ADDITIONAL FILES

The presentation of this paper will be accompanied by examples of the DSEs and DVEs produced by the class. For copyright clearance and file-size reasons these examples may not necessarily appear on the proceedings CD.

7. ACKNOWLEDGEMENTS

My thanks go to Dr Alf Watson of Northumbria University who taught the video aspects of this course. Internal moderation of the assessment task was undertaken by Fin McMorran, a fine-artist who teaches multimedia to computing students at Northumbria University; her comments and support were invaluable. Finally, thanks go to all the students of CG087 who made the experience so enjoyable and the assessment so much fun to grade.

8. REFERENCES

[1] E. Somers, "A Pedagogy of Creative Thinking based on Sonification of Visual Structures and Visualization of Aural Structures," presented at ICAD '98 Fifth International Conference on Auditory Display, Glasgow, 1998.

[2] P. Fishwick, "Aesthetic Programming: Crafting Personalized Software," *Leonardo*, vol. 35, pp. 383-390, 2002.

[3] P. Fishwick, "Aesthetic Computing," MIT Press, 2004.

[4] G. Kramer, "Auditory Display." Reading, MA: Addison-Wesley, 1994.

[5] E. Somers, "Designed Randomness," vol. 2004.

[6] New Arts Library. (1998). *Alfred Hitchcock's Psycho Screenplay.* Retrieved February, 2004, from http://www.paradiselost.org/psycho.html

[7] S. McCallion. (1997). *They Have a Better Idea...Do You?* Retrieved February, 2004, from http://www.fastcompany.com/magazine/10/one.html

[8] M. Back, "Micro-Narratives in Sound Design: Context, Character, and Caricature in Waveform Manipulation," presented at ICAD 96 - Third International Conference on Auditory Display, Palo Alty, CA, 1996.

[9] P. Vickers and J. L. Alty, "When Bugs Sing," *Interacting with Computers*, vol. 14, pp. 793-819, 2002.

[10] P. Vickers and J. L. Alty, "Using Music to Communicate Computing Information," *Interacting with Computers*, vol. 14, pp. 435-456, 2002.