MAKING MUSIC WITH IMAGES : INTERACTIVE AUDIOVISUAL PERFORMANCE SYSTEMS FOR THE DEAF

M Grierson

Goldsmiths College Electronic Music Studios

Dept. of Music

Goldsmiths College

New Cross

London

SE14 6NW

UK

m.grierson@gold.ac.uk

<http://www.mickgrierson.co.uk>

Abstract - This paper describes the technical and aesthetic approach utilized for the development of an interactive audiovisual performance system designed specifically for use by children with multiple learning difficulties, including deafness and autism. Sound is transformed in real-time through the implementation of a Fast Fourier Transform (FFT) and translated into a moving image. This image is adapted so that relevant information can be understood and manipulated visually. Finally, the image is turned back into sound. The translation process is based on research in computer music, neuroscience, perception and abstract film studies, supported by the Arts and Humanities Research Council (AHRC).

INTRODUCTION

Through collaboration with the Sonic Arts Network, Whitefields Special Needs School, the South Bank Centre and the London Philharmonic Orchestra, a software system has been developed that visualizes sound in real-time in a way that allows hearing impaired individuals to interact with a specifically designed, experiential representation of sound.

Deaf children with learning disabilities can interact with the system through the use of a Nintendo Wiiremote, which provides additional vibration-based feedback when the audio signal peaks over a set amplitude. Users can load and play back sounds from the remote, whilst using the accelerometer to make real-time adjustments to the pitch, pan position and volume. Beyond this, the musical output of the software can be set to a number of transposable scale systems. This allows for a wide range of musical interaction.

Musical interaction is conducted through the alteration of the visualization itself. Through learning to control a number of specific gestures, the users can alter the appearance of the visualization instantaneously. The visualization is then converted back

into audible sound, perceptually occurring at the same time as the original sonic material. The sound and music is effectively 'translated' into useful visual information that can be sensibly interpreted and changed in real-time by users who are unable to hear and react to the sound.

THEORY: THE VISUAL REPRESENTATION OF SOUND

Several standardized methods for visual representation of sound are used everyday by musicians, engineers, and others that need them. These methods can be loosely divided into three types: Symbolic representation, average level indication, and spectral analysis. Symbolic representations, such as musical scores, graphic notations and markup languages display information using an agreed set of symbols. These forms of representation only convey a certain amount of information, given a specific set of limitations. Most importantly, their relationship to sound is not indexical (in the sense of Peirce [1])- i.e. the sounds and images are not in 1:1 relationships. Level indicators such as PPM, VU and dBu meters display changes in the average amplitude of sound signals. These are less like symbolic representations in that their relationship to sound is indexical. However, the amount of sonic information contained in these displays is limited. Information is only apparent in one dimension – amplitude. For example, it can be considered impossible to discern pitch relationships by viewing a level meter. Spectral analysis is a robust method for sound visualization. Spectrographs & sonograms function by decomposing sounds into a number of component sinusoids, and plotting their amplitudes on a graph. This is usually performed using a Fourier Transform, although other (related) methods exist.

In the case of providing visualizations for those with disabilities for the purpose of interaction, all of these methods could be improved, either lacking in sufficient information, or being difficult to understand. Symbolic representations often contain little textural or experiential information – that is to say, symbolic methods can be used to describe how a sound changes over time, and what pitches it might contain, but not what it is like to hear the sound. Level meters provide some experiential information, but this is heavily limited. Spectral methods provide a great deal of information with respect to the power spectrum of sounds – but this information is not presented in a way which makes it easily understandable, nor experientially relevant.

Audio Visualization in Experimental Film

Within the discipline of experimental film there have been many highly regarded attempts to visualize sound and music in ways that are experientially relevant. These include the works of Hans Richter [2], Viking Eggeling [3], Oskar Fischinger [4], Norman McLaren [5] and Len Lye [6]. To some extent, information about sound is translated into moment-to-moment changes in shape and texture in an attempt to echo the experience of sound. An excellent example of this practice can be found in the work of John and James Whitney [7]. These practices have been well documented by scholars such as Al Rees [8], P. Adams Sitney [9], and Malcolm Le Grice [10], among many others.

Importantly, these relationships can easily be shown to be arbitrary, containing little or no musical or detailed sonic information (similar to symbolic and level indication methods, whilst sometimes containing textural information). Despite this, what is key about this practice is that it emerges from the idea that elements of our sonic experience – specifically in terms of motion - can be experienced visually. As Sitney mentioned on the occasion of film-maker Stan Brakhage's death [11], Brakhage was committed to the concept of 'moving visual thinking', the idea that internal cognitive visual processes could

be made visible. Even though Brakhage had often discussed his practice as being related to music, there is a tension here between the desire to explore purely visual processes, and the concept that vision can be similar to musical experience. However, through this tension rises potential new solutions to the problem of how to effectively represent sound for the purposes of audiovisual interaction, particularly in situations where participants have a very different experience of sound.

Neuroscientific Evidence for the Effectiveness of Audiovisual Congruence

Evidence supporting the view of experimental film-makers that the experience of sounds and music can be fundamentally linked with the experience of moving images continues to grow. Key examples demonstrating the stimulation of multisensory cells in the Superior Colliculus by audiovisual interaction include the McGurk effect [12], and the double-flash illusion [13]. In addition, evidence that seemingly arbitrary links between sound and image can improve event perception is emerging [14]. The idea that a visualization based on perceptual congruence may aid in translating the experience of sound for non-hearing populations has not been effectively tested, and there may be several problems with this view – especially considering that the precise nature of audiovisual relationships often appears arbitrary and is not always well understood. Our research does not necessarily claim to have scientifically demonstrated a definitive link between a particular type of visualization and a particular type of sound. Instead, we offer a technically useful and experientially relevant method for translating sound to an image and back again that may prove to be effective for enhancing musical interaction, particularly (but not exclusively) for the deaf and hard of hearing. Most importantly, we have chosen a visualization technique directly inspired by experimental film practice, and neuroscientific enquiry that directly relates to the structure of the visual cortex [15].

METHOD

This project combines experiential approaches that share similarities with symbolic and indexical forms of audio visualization (such as those found in experimental film making practice), with filtered spectral analysis and signal processing methods (FFT and related processes). This introduces a level of information not commonly found in similar approaches. Importantly, not all the available spectral information is used in the visualization. Attempts have been made to reduce the amount of information so that only the most important elements are retained. Once the sound has been translated into image,

it can be manipulated in real-time by a user. Importantly, the manipulated image can then be turned back into sound. This process occurs quickly enough that the users remain unaware of any significant time delay. Crucially, it is the real-time editing of this image which results in the perceptually relevant changes in the sound.

Figure 1. Max/MSP Patch showing an FFT process being sent to a video buffer

Figure 2: The FFT data expressed as 1 dimensional video information

The project was realized using Cycling 74's Max/MSP/Jitter, based on Miller S Puckette's Patching environment [18]. The sound is converted into a real-time spectrogram (see Figure 1). This spectrogram is loaded into a one-dimensional 32 bit floating point array, 2048 bins in size (see Fig 2.). It is then modified to appear as a set of two-dimensional concentric rings. In order to preserve processor load, the calculation from one to two dimensions is done using the computer's video hardware. This is achieved using a texture mapping process, with the one dimensional array being stretched across an OpenGL Surface. The OpenGL surface can be manipulated using a third dimension, to alter the appearance of the concentric ring formation with very limited CPU overhead. This results in an additional speed gain, and is useful for filtering the visual information without altering the material in the video buffer itself. This means that the image can be made more perceptually relevant without losing valuable data that is needed in order to

resynthesise the signal. Most importantly, all data manipulation continues to be performed on the one dimensional array, whilst being displayed in three dimensions.

The rings are spaced non-linearly so as to more accurately mirror human perception of audible frequencies. The exact spacing of the rings is achieved through the application of a logarithmic function on the visualisation of the 2048 point FFT spectrum. This function is applied by mapping the information onto a 3D OpenGL sphere (with non-uniform dimensions). This results in some spectral information loss with respect to the display. This is acceptable for two reasons. First, the spectral information is still maintained in the one dimensional array. The shifting in pitch appears relative, whereas if the display were linear, the shifting would appear to spread out as the fundamental frequency rose. Secondly, it is not required that the display be scientifically accurate, just that it remains perceptually relevant, whilst still resulting in an accurate re-translation back into sound. There is a precedent for the representation of sound and music in a concentric ring formation that is clearly demonstrated in the field of early 20th century film making and animation, such as the visual music tradition. Oskar Fischinger's silent 1924 film *Spiralen* [16] uses a number of concentric and spiral formations in what can be described as an early attempt at producing a visual metaphor for the experience of sound. In addition, this type of representation is echoed in the work of computer graphics and visual music pioneer, John Whitney. Although neither of these artists exclusively used the concentric formation, the iterative tunnel effect which it produces fits with the forms that continually appear in this canon.

Further reasoning behind the choice of visualisation comes from studies in neuroscience and perception. In Bresslof, P., et al [17], a number of visual patterns are identified as persisting in various forms of visual art and experience generally as a result of their direct mathematical relationship with the structure of the visual cortex. These visual configurations are roughly divided into four categories - tunnels and funnels, spirals, lattices (which include honeycombs and triangles), and cobwebs. The formation described here fits more than one of these categories at various times depending on the nature of the sound being analysed / re-synthesised. However, due to the perspective illusion that it evokes, it most closely fits within Bresslof et al's 'Tunnels and Funnels' category.

Figure 4. Still image from *Spiralen* (1924) by Oskar Fischinger (left), and a screengrab from the audiovisualisation software (right).

These precedents provide theoretical weight to the idea that a very specific type of concentric visualisation may be effective with respect to the goals of the project. The Concentric ring display format was chosen in an attempt to encourage the users, most of whom were autistic in this case, to focus on the image as a focal point of their experience, and to become sensitive to the complex and potentially delicate variations in the sound that result from the user altering the image through their own interaction. It is also being employed to hold their attention, and to give them an indication of the morphology of the sound and music that other, able bodied participants are experiencing. Sonic transformations are performed by moving the FFT bins in both directions. These bin shifts are logarithmic – frequency relationships remain coherent throughout musical octaves. In addition, phase information is rotated, enabling the system to operate as a phase vocoder. As such it provides smooth real-time pitch correction, and can be used as a high quality real-time time-stretching device. In addition, the system allows for spectral freezing and Fourier synthesis through manipulation of frozen spectral fragments.

RESULTS

Working alongside the Sonic Arts Network, Whitefields school, and in collaboration with experienced workshop leader, Duncan Chapman, various versions of the software were tested with a number of users. Feedback was very positive with respect to the effectiveness of the visualisation. Immediately it became apparent that the software gave children with multiple learning disabilities including deafness the opportunity to appreciate the way their own sound making was experienced by others. Through comparing the experience of 'video-listening' to pre-recorded and live sound material with the experience of their own interaction and sound making activity, users were able to more effectively appreciate the effects of sound in the world around them. As such, the tool was considered immediately useful in (re)habilitation, giving non-hearing users a simple way of having meaningful relationships with sound and music.

Two significant problems were identified and overcome throughout the testing period. First, the system allowed for non-hearing individuals to create sounds through Fourier synthesis. It was found that although this could be entertaining, it could produce a wide variety of unpleasant sounds. By request, these functions were disabled. In addition, it was noted that the graphical interface could be developed to enable teachers and students to use the system unaided by the workshop leader. As a result of this process, a GUI was designed through collaboration with an information designer with knowledge of teaching environments.

Figure. 5: The Graphical User Interface, designed in consultation with information designer Myah Chun.

CONCLUSION

This system has been used in number of collaborative workshops and performances at the South Bank Centre, London. These featured members of the London Philharmonic Orchestra working with disabled students from Whitefields school. These events were aurally and visually documented by the Sonic Arts Network, and several compositions were directed by Duncan Chapman. The system has been improved to allow ease of use for both hearing and non-hearing players of varying ability so that they may collaborate in performance in ways that are immediate and inherently musical.

The software was made publicly available free of charge in May 2008 by the Sonic Arts Network [19]. In addition, some source code will be released as part of the AHRC funded project, Cognitive and Structural Approaches to Contemporary Computer Aided Audiovisual Composition. This project will be demonstrated as part of an upcoming BBC technology article, due for publication in Summer 2008.

References

- [1] HARTSHORNE, C, Weiss, P & Burks A.W. (Eds): Peirce, Charles Sanders (1931-58): Collected Writings (8 Vols.), Cambridge, MA: Harvard University Press.
- [2] RICHTER, H: *Rhythmus* 21, 1921.
- [3] EGGELING, V: *Diagonal Symphony*, 1922-24.
- [4] FISCHINGER, O: *Studie No. 6*, 1930.
- [5] MCLAREN, N: *Allegro*, 1939.
- [6] LYE, L: *A Colour Box*, 1935.
- [7] WHITNEY, J, Whitney, J: *5 Abstract Film Exercises*, 1944-45.
- [8] REES, A L: A History of Experimental Film and Video, BFI, 1999.
- [9] ADAMS SITNEY, P: Visionary Film: The American Avant-Garde, 1943-78 New York: Oxford University Press, 1974.
- [10] LE GRICE, M: Abstract Film and Beyond, MIT, 1977.
- [11] ADAMS SITNEY, P: "Jusqu'a son dernier souffle", Cahiers du cinema 578, pp. 50-1, Editions de l'Etoile, April 2003.
- [12] MCGURK, H and MacDonald, J: "Hearing lips and seeing voices," Nature Vol 264(5588), pp. 746-748, 1976
- [13] SHAMS, L., Kamitani, Y., Shimojo, S: What you see is what you hear, Nature Vol(408), pp. 788, 2000.
- [14] EFFENBERG, A.O: "Multimodal Convergent Information Enhances Perception Accuracy of Human Movement Patterns," Proc. 6th Ann. Congress of the European College of Sports Science (ECSS), Sport und Buch Strauss, p.122, 2001.
- [15] BRESSLOFF, P.C., Cowan, J.D., Golubitsky, M., Thomas, P.J., Weiner, M.C: What Geometric Visual Hallucinations Tell Us about the Visual Cortex, *Philosophical Transactions of the Royal Society B*, 356, 2001.
- [16] FISCHINGER, O: *Spiralen*, 1924.
- [17] BRESSLOFF, P.C., Cowan, J.D., Golubitsky, M., Thomas, P.J., Weiner, M.C: What Geometric Visual Hallucinations Tell Us about the Visual Cortex, *Philosophical Transactions of the Royal Society B*, 356, 2001.
- [18] PUCKETTE, M: "The Patcher," Proceedings, ICMC, International Computer Music Association, San Francisco, pp. 420-429, 1988.
- [19] Available from <http://www.sonicartsnetwork.org>