Visualising Sound: Localisation, Feature Analysis and Visualisation

Jack Armitage, Kyle Molleson, Michael Battcock, Chris Earnshaw, David Moore, and Kia Ng
ICSRiM – University of Leeds
School of Electronic and Electrical Engineering
School of Computing and School of Music
Leeds LS2 9JT
United Kingdom
soundvis@icsrim.org.uk, www.icsrim.org.uk

Sound is an integral medium of communication in society. Despite its influence in everyday interaction, the fundamental features of sound and its impact are not commonly understood and often not considered. Advanced dissection and analysis of sound is often used to aid technology's understanding of its environment (e.g. in robotics and telecommunications). The aim of this project is to utilise sound information technology to enhance our understanding of sound and how we process it. Visualisation can provide a more accessible representation of complex sound analysis, and so it is investigated here in the context of musical performance and interactive installation. This cross-modal experience is informed by the phenomenon of synaesthesia and creative mapping to express its subjectivity by providing a user-unique experience. In addition to results and user evaluations, the paper concludes with plans for future development, focusing on the impact of the interactive installation on the user and ways in which the technology developed can be used in creative interaction with sound.

Sound. Visualisation. Cross-modality. Music. Performance. Installation. Arts. Synaesthesia. Mapping. Interactive. Multimedia.

1. INTRODUCTION

Sound visualisation is a process of constructing and manipulating connections between aural and visual perception. This mediation procedure, often referred to as mapping, must exhibit flexibility between direct and predictable relationships and abstract and unpredictable relationships in order to create a continually engaging and immersive experience. These requirements necessitate an intricate understanding of a given sound environment, which can be achieved through the development of sound information technology systems. Sound features obtained from digital signal analysis form the basis for modal transformation, along with a close reading of perceptual phenomena such as synaesthesia and previous subjective interpretations through art.

The technical aims of this project are:

- To develop a hardware system for real-time analysis of a sound environment
- To develop software tools for sound visualisation mapping

This project also aims to explore subjective perspectives of sound visualisation through new installation and performance work. Specifically, we want to see if it is possible to share our own perspectives and experience, with those of others, using interaction and mappings that change and evolve over time.

This paper begins with a brief history of sound visualisation and an overview of the predominant conceptual framework for informed visualisation mapping. A number of emerging technologies, which can extend the practice of sound visualisation, are examined in relation to our project aims. Sound visualisation is then considered in the context of synaesthesia and a selective history of the relationship between sound and colour imagery within the arts is presented. The design, development and validation of the proposed system are described, with reference to prototype test results. Finally, directions for future research and potential system applications are discussed.

2. BACKGROUND

The representation of sound using visual media has existed for many centuries, in many cultures and for many reasons. However, the toolset employed to visualise sound is constantly being updated and can be extended further by integrating advances in fields such as cognitive science, psychology and computer science. Concurrently, academic and artistic endeavours to create meaningful sound visualisation have increased and diversified in the last century. This section presents research into sound visualisation and examines work that is congruent with the project aims specified above.

2.1 Developments in Sound Visualisation

The traditional western musical score is perhaps the most well known method of visually representing music. Designed as a method of preservation and later used as a performance aid (Rastall 1998, p.5), it is only understood by trained musicians and does not provide a compelling visual accompaniment for the audience. It is especially unhelpful in realising non-western music that is less concerned with hierarchy and finite timespan such as Javanese Gamelan (Sumarsan 1995, p.107). It also fails to represent the abstract, textural sounds that permeate post-20th century music (Read 1987). In reaction to this problem, the score has been deconstructed, reconstituted and augmented by composers such as Cornelius Cardew, Iannis Xenakis and Karlheinz Stockhausen (Sauer 2009). In parallel to these developments, sound artists such as Oskar Fischinger and Daphne Oram took the more direct approach of optically reading graphical symbols and synthesising them, thereby establishing the practice of ‘drawing sound’.

Today, computer science has opened up a broad range of technical solutions for audiovisual artists, but the problem of mapping is deceivingly complex. For example, audiovisual mapping requires useful, flexible and meaningful taxonomies for both senses. A comparison of two approaches to the taxonomy of sound illustrates the ‘semantic gap’ described by Lew *et al.* (2000) that exists in sound analysis. Schafer (1977) set in motion an acoustic ecology framework (Truax 1978) for curators to describe sound based on cultural, temporal and geographical resonances. On the other hand, IRCAM’s music information retrieval (MIR) project CUIDADO (Peeters 2004) represents an example of computer music approaches to sound categorisation (see also Bullock 2008, p. 46). Here sound is separated into six categories: temporal shape, temporal feature, energy features, spectral shape features, harmonic features and perceptual features. Ultimately, humans rely on context and experience to understand music (Meyer 1956), which is absent from the determinism of signal processing.

Real-time and generative sound visualisation can potentially benefit from any techniques, which broaden the context of sound features extracted from a digital signal. Machine learning and artificial intelligence can be applied to digital signal analyses to register high-level features like musical gestures, which can be segmented and matched for similarity (Gillian 2011, Bloit 2011). This area of research is continually improving automated musical understanding. Spatial awareness can also add contextual understanding by linking the sound space and the visual space. Sound source localisation is commonly used in teleconferencing systems and in robot audition (Hu 2008). The availability of this technology means it is now investigated as a live performance tool (Trondheim Voices 2011). In this example, spatial recognition of the performers informs sound processing rather than a visualisation. Part of this project involves exploring the potential of sound source localisation as a context enhancement for sound visualisation.

2.2 Colour-Sound Perspectives and Synaesthesia

Synaesthesia is a neurological condition in which stimulation of one sense leads to the involuntary stimulation of another sense. The condition is predominantly known for manifesting abstract colour relationships with graphical symbols, tastes and especially music. Despite synaesthesia being an inherently subjective experience, colour-sound relationships have been proven to be consistent across larger sample sets. For instance, the relationship between the perceived pitch of a sound to the overall lightness of any visualisation has been established. In several tests (Hubbard 1996 and Ward *et al.* 2005), almost all participants in both synaesthete and non-synaesthete groups preferred uniformly lighter colours for higher pitches. Another common feature amongst colour-sound synaesthetes is the development of specific colour palettes for musical instruments. This is shown in Ward *et al.* (2005), where each instrument has its own pitch-colour mapping; palettes for particular groups of sounds tend to show more variation between synaesthetes than generic pitch-colour relationships.

The body of synaesthesia research has grown significantly in the last fifty years, but our understanding is very much incomplete. However, composers have been aware of this phenomenon for much longer, and in a few cases this ability directly informed their work. Alexander Scriabin, a composer and pianist in the early 20th Century, was one of the composers well known for the influence that sound-colour synaesthesia has had

on his work, although it has been questioned as to whether he experienced the condition himself or not (Galeev & Vanechkina 2001). In the score for his piece Prometheus: The Poem of Fire (1910), Scriabin includes a part for the "Luce" (a colour organ), as a means for the piece to be performed as both an auditory and visual experience.

Scriabin's associations between specific notes and colours are well documented, as seen in Table 1, and played an important part in the composition of his pieces. Due to the subjective nature of Synaesthesia, this mapping style and colour scheme is by no means uniform amongst composers with the condition. For example, Oliver Messiaen experienced a completely different set of colours for each piece he composed, varying from mode to mode (Bernard 1986). In this paper, we seek to model the mappings based on these composer as case studies to aid the development of our mapping strategies.

Table 1: *Scriabin's Note-Colour Associations (Galeev & Vanechkina 2001)*

Note	Colour according to Scriabin
C	Red
G	Orange-Pink
D	Yellow
A	Green
E	Whitish-blue
B	Similar to E
F#	Blue, Bright
Db	Violet
Ab	Purplish-Violet
Eb	Steel colour with a metallic sheen
Bb	Similar to Eb
F	Red, Dark

2.3 The Influence of Music on Visual Art

The mediums of visual art and music often collide in the development process of a creative work (Champa 2000). Historically these labels have been separated; due mainly to the media each relies upon. Music is an inherently time based art form, whereas static visual art transcends time and is able to be experienced non-linearly. Many painters have produced works that have been described as 'musical', notably Delacroix and Kandinksy (Verdi 1968). Equally, composers Debussy and Ravel became synonymous with the impressionist movement alongside painters such as Monet and Cezanne (Palmer 1973). During the early 20th century artists such as the surrealist painter Paul Klee began experimenting with the concept of visual sound and is credited with providing "a more profound and unique solution to the problem of the relationship between music and the visual arts" (Verdi 1968).

Klee used music as a viable medium to draw influence from by transposing musical gestures into visual form. His works often focused on creating dynamic movement within the fixed medium of painting in the spirit of Walter Pater's maxim, "All art aspires towards the condition of music" (Pater 1877). It is noted that Klee sought to "discover what universally applicable aesthetic properties could be isolated from the accomplishments of the titans of music" (Kagan 1983) and how these aspects could be translated into an effective visual stimulus. The depth and simultaneity characteristics of objects in Klee's later work are given the musical description 'polyphonic painting'. The artist describes polyphonic painting as "superior to music in that, here, the time element becomes a spatial element" (Klee 1964). This translation of a musical attribute into a static visual effect allows the viewer to experience multiple pictorial objects simultaneously, negating the passing of time as a barrier for musical expression in visual form.

Our approach to the visualisations will integrate the concepts of synaesthesia outlined in Section 2.2 and Klee's style of musical gesture by populating the augmented virtual space with sound objects that vary in size, colour and animated agility. With the added localisation parameter allowing the objects to appear, flourish and diminish according to the users' position, we will attempt to reference the effect of constant motion that permeates Klee's work.

3. DESIGN AND DEVELOPMENT

In this section we describe the embedded system that analyses its sound environment and transmits the results for use as an input for visualisation and interactive response. The system can be conceived using the conceptual diagram in Figure 2, which also illustrates a perception-action cycle. Figure 3 depicts the block diagram and data flow of the system, which is described in four areas: input (data capture), embedded digital signal processing (DSP), visualisation and prototype testing. Developments in each area are outlined and important design considerations are discussed.

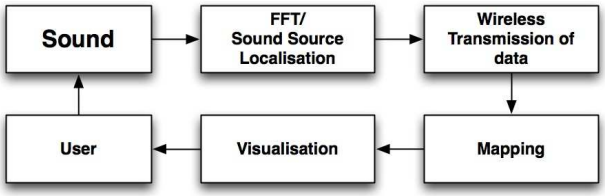

Figure 1: *Conceptual diagram of the proposed sound visualisation system*

3.1 Input

The data capture process involves a number of hardware and software configurations. The microphone array forms the structure of the input section. A printed circuit board (PCB) contains the amplification and filtering components used to prepare the sound signal for digital signal processing within the embedded processor.

3.1.1 Microphone Array
Initial experiments showed a distinct advantage for arrayed microphones over distributed microphones for source localisation. *N* microphones are required to perform accurate source localisation on *N-1* sound sources. The number of microphones permissible in this application is limited by the processing capacity of the embedded system and the minimum sample rate required to analyse a signal bandwidth similar to the human auditory system.

MEMS (MicroElectrical-Mechanical System) microphones are used for data capture as they are the current industry standard for microphone applications within small devices such as smartphones, video cameras and teleconferencing systems. MEMS microphones were chosen for their small size and ease of integration with the prototype components. Operational amplifiers bring each microphone's signal to an appropriate level for signal processing. As digital amplification would have lowered the overall signal resolution due to quantisation, analogue amplifiers are used. The amplifiers also act as low pass filters through the use of an extra capacitor, preventing audio aliasing errors.

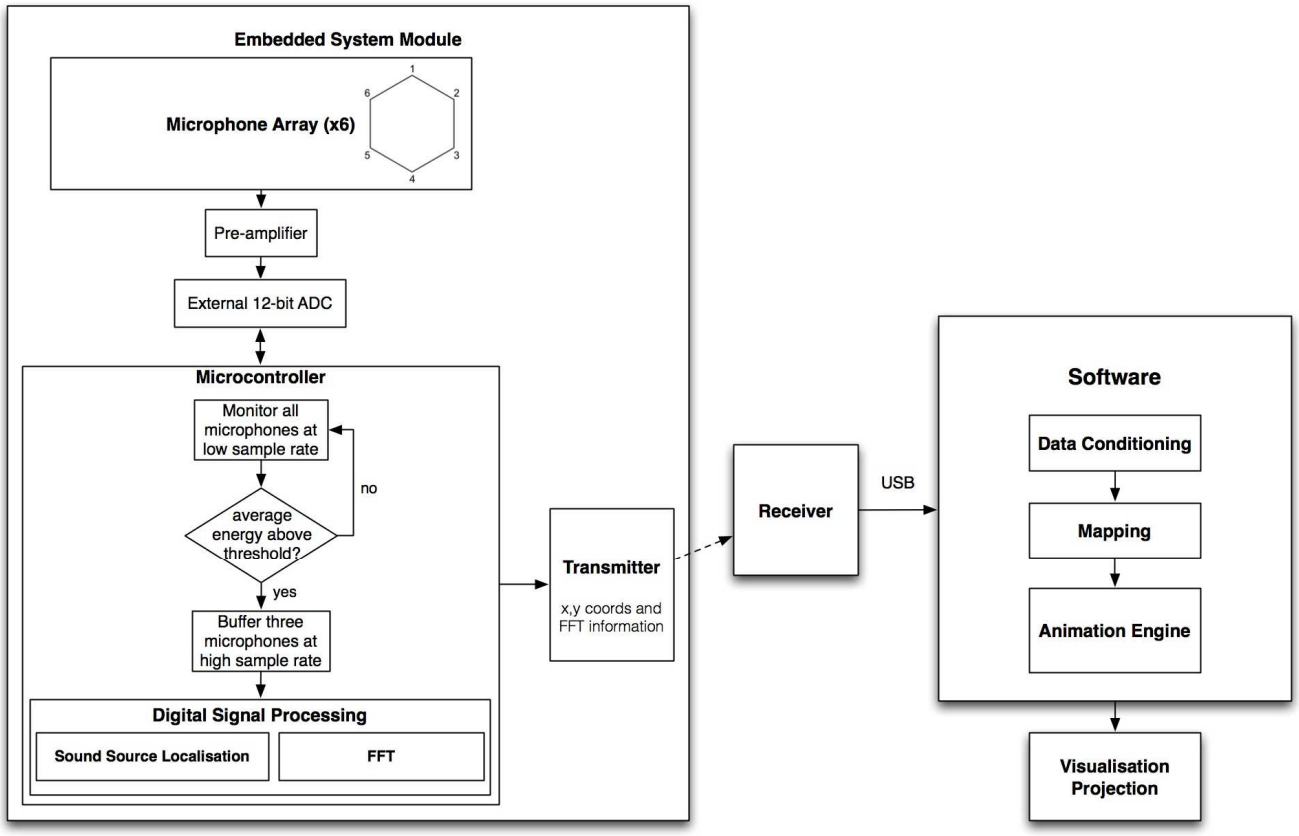

Figure 2: *Block diagram and data flow*

3.2 Embedded DSP

The embedded software performs two key tasks; sound localisation and feature extraction using frequency components extracted via FFT (Fast Fourier Transform). In order to achieve near real-time operation from the processing speed available on the embedded system platform, the algorithms used for both tasks underwent thorough optimisation.

3.2.1 Localisation
Sound localisation is used in this project to allow the visualisations to be calibrated to the sound source. In order to calculate the location of a sound source the time difference of arrival (TDOA) is measured by comparing the microphone signals. This method works across multiple microphones by arranging them equidistantly in a hexagonal array. The TDOA is the basis of the sound localisation algorithm and is explained visually in Figure 3.

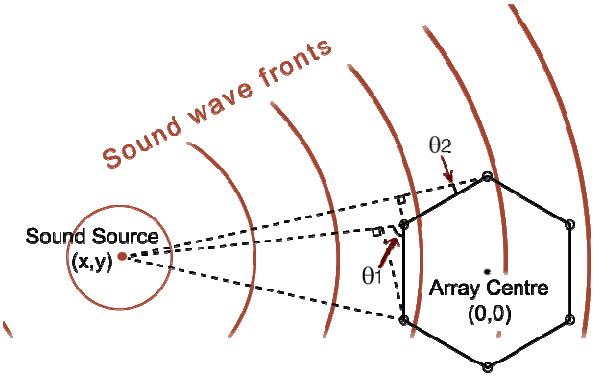

Figure 3: *How the sound waves create a TDOA at each microphone in the array due to their propagation*

The computational cost of correlation is compensated for by selectively deploying it. The six microphones in the array are initially monitored at a low sample rate. When a minimum amplitude threshold is crossed, sound energy averaging identifies the three most active microphones, buffers them at a higher sample rate and performs cross correlation.

$$(f * g)[n] = \sum_{M=\infty}^{\infty} f^{*}[m] \cdot g[n+m]$$

Equation 1

The discrete form of the cross correlation shown in Equation 1 is applied to two signals to compute the time delay between them, known as the time lag. This equation takes two time domain signals and moves them in time to determine phase difference. The lag value received is the value used to evaluate where a sound source is with relation to the microphones.

$$GCC-PHAT(d) = IFFT\left\{\frac{X_1 \cdot X^{*}_2}{|X_1||X_2|}\right\}$$

Equation 2

Generalised cross-correlation with phase transform (*GCC-PHAT*, shown in Equation 2) is another method to find correlation between signals and their lags at different samples in the signal. This equation takes the frequency domain representation of one signal and correlates it with the complex conjugate of the second signal. *GCC-PHAT* is used in the final project as it is much more efficient and robust especially under reverberant conditions. Although a lot of Fourier transform processing is required, this equation is far superior with regards to CPU time as shown with comparisons MATLAB.

With the lag values from two pairs of microphones in the array, a source coordinate can be calculated. The final location coordinate is transmitted to the visualisation software.

3.2.2 Fast Fourier Transform (FFT)
FFT is used to transform a time domain signal into the frequency domain. FFT is a mathematically more efficient way to compute the Discrete Fourier Transform (DFT) and therefore is preferred in programming. FFT can be computed in *NlogN* computations whereas the DFT requires *N2* this is done through exploiting repetitions found when computing each term.

The FFT is, as demonstrated in this paper, being used for both analyses of frequencies of great interest within the audio signal and as a subroutine to acquire a full frequency domain representation of the signal for use in the *GCC-PHAT* algorithm. In order to send only the results of most interest over the wireless link to the visualisation software, the embedded system can select certain points from the FFT. The results of interest might be the fundamental frequencies, harmonics based on that fundamental frequency or the top most prominent frequencies and their relative magnitude.

3.2.3 Data Output
The results of sound source localisation and FFT analysis produced on the embedded system form the basis of visualisation. This data is sent wirelessly to the computer hosting the visualisation software. Transferring this data wirelessly allows for discrete placement of the module and distance from any fan noise from a computer or projector. It also increases the module's portability and ease of use, extending opportunities for potential applications of the project such as a live performance tool.

3.3 Visualisation

The data received from the hardware module is broadcast on an internal server on the computer hosting the visualisation software. This allows for any necessary data conditioning to be scripted in before the data is visualised. At this stage, higher-level audio features such as those discussed in Section 2.1 are realised and then distributed as mappable parameters.

A node-based visual development platform is used to produce the visualisation. Within this platform it is possible to incorporate 3D models, animations, particle models and physics engines. Mappings can also be read and written which is the method used to investigate known mappings such as those of Scriabin and Messiaen.

3.3.1 Feature Mapping
The overall mapping strategy can be described in three categories; audio features generate visual matter, localisation governs the physical manipulation of the visual matter and collected data is used to construct a narrative from the beginning of usage to the present.

In order to map sound features to visual matter, a custom sound feature taxonomy has been developed which uses a hierarchical structure. At the lowest level are basic mathematical descriptors similar to those mentioned in Section 2.1. These descriptors are then used as building blocks to construct higher-level features such as gestures and rhythms. High-level features are stored during usage, providing a reference for new material, which can be matched for similarity. The sound taxonomy is mapped to a visual taxonomy, which is also constructed in a hierarchical manner. The visual taxonomy is founded on abstract geometric elements such as lines and shapes, which can be arranged into patterns and more discernible objects to reflect the detection of high-level audio features.

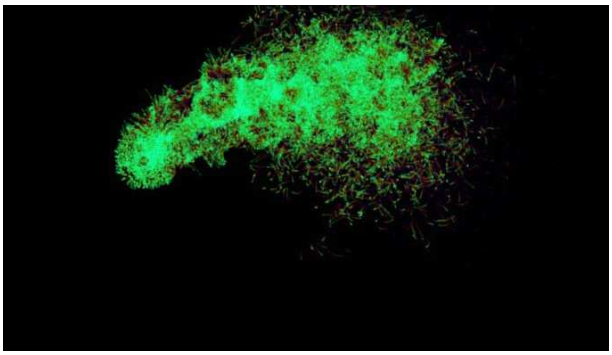

Figure 4: *Localisation animation with path trajectory shown as a particle trail*

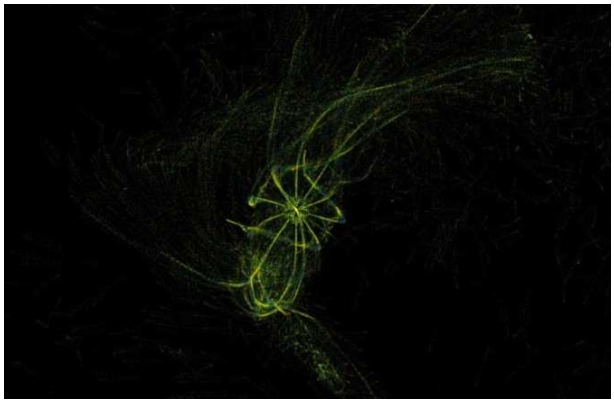

Figure 5: *Particle based visual sound object, coloured in response to pitch*

The localisation data is used to drive the physics elements of the visualisation. Applying physics to the data allows for parameters such as velocity and trajectory (Figure 5) to be visually represented. The predicted future position can also be established and incorporated.

In order to provide a continually engaging experience during usage, mappings change over time. The intention is to keep the audience or user in a constant state of attention by blending familiar mappings with new ones, or even abruptly changing the mapping altogether. Changes can be automated or controlled by the performer. Specific locations in space can trigger mappings changes, as can specific musical material. The mapping changes can also be triggered by any other device which can be interfaced with the visual development platform, such as MIDI instruments, smart-phones or any other sensor data.

3.4 Prototype testing

Development software MATLAB has been utilised to compare the embedded system processing speed when performing cross correlation with time domain signals and frequency domain signals. Results showed that for 200,000 samples, time domain processing took approximately 40 seconds whereas cross correlation using the frequency domain data was almost instantaneous. These tests enabled the software design strategy to include conversion of each audio signal into their respective frequency components resulting in further optimisation of the cross correlation algorithm.

A number of experiments to test the effectiveness of the prototype have been carried out in a controlled environment. An acoustic laboratory equipped with an acoustic measurement system has allowed for intensive testing on the prototype including, distance, angle, volume and sound source (sinusoidal, white noise, speech and music). Figure 6 depicts an experiment testing the

localisation algorithm and accuracy of the source angle detection.

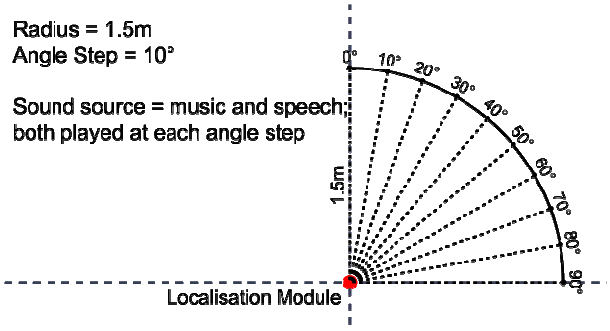

Figure 6: *Localisation testing on prototype using music and speech as sound sources*

The efficiency of the embedded DSP cross-correlation algorithm was compared against MATLAB's internal algorithm using the same acoustic data. This comparison showed our system to be accurate to within 2-3° for perpendicular sources and 4-5° for peripheral sources. Through experimentation and optimisation iterations the software could be upgraded and made more efficient.

4. CONCLUSION AND FURTHER WORK

This paper has presented the research, design, implementation and evaluation stages of a system which can provide robust sound source localisation, feature analysis and visualisation. The hardware module can transmit audio analysis results wirelessly which can then be transformed into visualisation for use in an installation or live performance. Synaesthesia research is used as one of the possible mapping strategies for visualisation. It is suggested by initial tests that this strategy will result in a more visceral and meaningful translation from sound to vision.

In this paper, we focussed on sound visualisation in the context of live musical performance and interactive installation. Although this paper explores these perspectives separately, the system could involve both the performers and the audience simultaneously. In such a scenario, each audience member could view their personal mapping using an augmented reality smart-phone application. Individual mappings could then be scaled to larger group mappings; an emergent visualisation could arise from many users influencing a global mapping. This would enable further research into synaesthesia for different groups of musical instruments and complex textures.

Other future work involves exploiting sound source localisation and high-level feature analysis in human-robot musical interaction (Solis & Ng 2011). Music psychology research has shown the importance of non-verbal communication in ensemble performance (Cambell 1995). Behaviours such as eye contact and bodily gesticulation allow performers to coordinate and aid the audiences' understanding of the performance (Shutter-Dyson & Gabriel 1981). Further development would enable the capture and analysis of musical interactivity between performers, which could be calibrated with other gestural data analysis such as motion capture. Designing human-robot interaction around the analysis of human-human musical interaction is therefore desirable and possible. For example, a robot could perceive a musical change, localise it to the appropriate performer and acknowledge them through a familiar non-verbal gesture. Musical robots are a key long-term research area for music education and home entertainment.

Our mobile audition system also has potential usage as a sensory aid for people with hearing impairments or autists with sensory processing issues. Establishing a sound visualisation mapping with a user with a hearing impairment would allow them to navigate sound visually. Similarly, an autist user with sensory processing difficulties could use sound visualisation as part of a sensory integration therapy program (Schaaf 2011). Research into sensory replacement and augmentation has far reaching potential and consequence and is seen as a long-term interest.

5. REFERENCES

Bloit, J., Rasamimanana, N., and Bevilacqua, F. (2010) Modeling and segmentation of descriptor profiles with segmental models. *Pattern Recognition Letters*.

Bullock, J. (2008) *Implementing audio feature extraction in live electronic music*. PhD thesis, University of Birmingham.

Campbell, P. S. (1995) Of Garage Bands and Song-getting: The Musical Development of Young Rock Musicians. *Research Studies in Music Education*, 4: 12.

Champa. K. S. (2000) Painted Responses to Music: The Landscapes of Coret and Monet, Morton. M. L., and Schmunk. P., *The Arts Entwined: music and painting in the nineteenth century*. Garland Publishing, New York.

Galeyev, B. and Vanechkina, I. (2001) Was Scriabin a Synaesthete? *Leonardo*, 34(4), pp. 357–362.

Gillian, N., Knapp, B., and O'Modhrain, S. (2011) An adaptive classification algorithm for semiotic musical gestures. In: SMC (Sound and Music

Computing), *8th Sound and Music Computing Conference*. Padova, Italy 6–9 June 2011.

Goina, P. and Polotti, P. (2008) Elementary Gestalts for Gesture Sonification, *Proceedings of the 2008 International Conference on New Interfaces for Musical Expression (NIME–08)*. 150–3. Genova, Italy.

Howells, T. (1944) The Experimental Development of Color-Tone Synaesthesia. *Journal of Experimental Psychology*, 34, pp. 87–103.

Hu, J-S., Huang, C-Y., Yang, C-H., Wang, C-K., and Lee, M-T. (2008) Sound source tracking and speech enhancement by microphone array on an embedded dual-core processor platform. In: *IEEE Workshop on Advanced Robotics and its Social Impacts (ARSO)*. Taipei, China, DOI 10.1109/ARSO.2008.4653624, 23–25 August, 2008.

Hubbard, T. L. (1996) Synaesthesia-like Mappings of Lightness, Pitch, and Melodic Interval. *American Journal of Psychology*, 109(2), pp. 219–238.

Kagan, A. (1983) *Paul Klee/Art & Music*, Cornell University Press, New York

Kandinsky, W. (1977) The Psychological Working of Colour in Concerning the Spiritual in Art. pp. 23-26, Dover Publications, New York.

Klee. P. (1964) The Diaries of Paul Klee, trans. Pierre B. Schneider, R. Y. Zachery and Knight. M., University of California Press, Berkeley/Los Angeles.

Lew, M. S., Sebe N., Djeraba C., and Jain R., (2006). Content based multimedia information retrieval: State of the art and challenges. In: ACM Trans. *Multimedia Computing, Communications, and Applications*, 2(1): 1–19. 2006.

Meyer, L. B. (1956). *Emotion and Meaning in Music*. Chicago: University of Chicago Press.

Solis, J., and Ng, K. (2011) *Musical Robots and Interactive Multimodal Systems*. Springer Tracts in Advanced Robotics, Springer.

Palmer, C. (1973) *Impressionism in Music.* Hutchinson, London

Pater, W. (1877) The School of Giorgione, *Fortnightly Review*.

Peeters, G. (2004) *A large set of audio features for sound description (similarity and classification) in the CUIDADO project*. Paris: IRCAM Recherché.

Rastall, R. (1998) *The Notation of Western Music.* Leeds: Leeds University Press.

Read, G. (1987) Source Book of Proposed Music Notation Reforms. Connectitut: Greenwood Press, Inc.

Sauer, T. (2009) Notations 21. New York: Mark Batty Publisher.

Schaaf, R. C. (2011). Interventions That Address Sensory Dysfunction for Individuals with Autism Spectrum Disorders: Preliminary Evidence for the Superiority of Sensory Integration Compared to Other Sensory Approaches. In: B. Reichow, P. Doehring, D. V. Cicchetti, and F. R. Volkmar. *Evidence-Based Practices and Treatments for Children with Autism*. Springer US, pp. 245–273.

Schafer, R. M. (1977) *Our Sonic Environment: the Tuning of the World*. New York: Knopf.

Shuter-Dyson, R. and Gabriel, C. (1981) *The Psychology of Musical Ability*, 2nd edition. Methuen.

Sumarsam (1995) *Gamelan: Cultural interaction and musical development in central Java*. Chicago: University of Chicago Press.

Trondheim Voices (2011) [Live performance]. In: NIME (New Instruments for Musical Expression), *11th International Conference on New Instruments for Musical Expression*. Oslo, Norway 30 May – 1 June 2011.

Truax, B. (1978) *The World Soundscape Project's Handbook for Acoustic Ecology*. Vancouver: ARC Publications.

Ward, J., Brett, H., and Elia, T. (2005) Sound-Colour Synaesthesia: To what extent does it use cross-modal mechanisms common to us all? Cortex.