MusiCam – an Instrument to Demonstrate Chromaphonic Synesthesia

Derek Yau and Rachel McCrindle

The University of Reading, School of Systems Engineering, UK

siu02dy@rdg.ac.uk, r.j.mccrindle@rdg.ac.uk

Abstract

Inspired by a type of synesthesia where colour typically induces musical notes the MusiCam project investigates this unusual condition, particularly the transition from colour to sound. MusiCam explores the potential benefits of this idiosyncrasy as a mode of Human Computer Interaction (HCI), providing a host of meaningful applications spanning control, communication and composition. Colour data is interpreted by means of an off-the-shelf webcam, and music is generated in real-time through regular speakers. By making colour based gestures users can actively control the parameters of sounds, compose melodies and motifs or mix multiple tracks on the fly. The system shows great potential as an interactive medium and as a musical controller. The trials conducted to date have produced encouraging results, and only hint at the new possibilities achievable by such a device.

Keywords: colour organ, audio synthesis, musical controller, synesthesia, HCI

1. Introduction

The term synesthesia (from the Greek, syn = together + aisthesis = perception) means 'joined sensation' and as such refers to an involuntary physical experience in which the stimulation of one sense causes an additional perception in a different sense or senses (Cytowic 2003). For example, a synesthete might feel, see or taste a person's voice as well as hearing it; might detect a scent on seeing a particular colour; or when looking at printed black numbers might see them in colour, each with a different hue. Synesthetes represent a group of otherwise 'normal' people who experience the ordinary world in extraordinary ways due to their senses of touch, taste, hearing, vision and smell becoming mixed up rather than remaining separate (Ramachandran 2003).

2. Aims and Rationale of the MusiCam Project

The MusiCam project aims to explore a particular type of synesthesia where colour will typically induce musical sound. Whilst it is impossible to replicate exactly different synesthetes' personal perceptual abilities, through the use of digital multimedia techniques we can create virtual synesthesia (MIT 2006) demonstrations that give a close approximation to what a person with synethesia might experience. In addition to studying the perceptions of synesthesia we explore the issues of colour sound combinations more generally with regard to using them as a computer based communication medium, for relaxation or rehabilitation purposes or as a static or interactive ambient display.

With these goals in mind, there are three intrinsic areas of interest which have influenced and inspired the development of the MusiCam system:

- The colour/sound phenomenon and audio/visual art.
- The synesthesia condition and chromaphonia in particular.
- Human Computer Interaction (HCI), musical controllers and Interconnected Musical Networks (IMN).

Colour/pitch scale systems were devised as early as the 16th century and 'colour organs' [note 1] developed by the 18th century (Moritz 1997). In these early systems the relationship between colour and pitch was normally determined by the inventor of the system using mathematical techniques. Even today's counterparts still use scales that are fundamentally drafted from assumption; but with computers frequently involved in producing arbitrary mappings between colour and sound (Abbado 1988). In chromaphonia (coloured hearing), a type of synesthesia (Cytowic 2003), those with the condition will naturally associate a particular colour with a corresponding musical note and in effect create their own colour scale; often this is unique for each synesthete.

Proceedings ArtAbilitation 2006: ©ArtAbilitation/SoundScapes ISBN 87-7606-015-2

Aside from investigating the characteristics of the condition, we believe that discovering novel ways to interact with the personal computer will improve human creativity and in consequence improve communication and cooperation between humans. The benefits of sound and music in computer interaction are often forgotten even though auditory stimuli may leave longer lasting impressions than visual stimuli (Darwin 1972), better recall can be attained if information is received aurally as opposed to being read, and humans can react faster to auditory stimuli than to visual stimuli (Goose 1998). Music is one of the most highly structured auditory mediums and communicates through parallel streams, which can be effortlessly extracted and differentiated by our human cognition processes. Additionally, the inspirational qualities of music naturally spawn a multitude of research areas such as music in education, music therapy and more recently Interconnected Musical Networks (IMNs) which allow players to independently share and adapt each others music in real time (Weinberg 2001, Machover 1996).

3. The MusiCam system

The MusiCam system uses image processing techniques to extract colour data from a webcam. The colours that result from the extraction determine what note, track or instrument is played. Adding more colours can result in richer textures and increasing the level of a particular colour manipulates the dynamics of the system. Different banks of sound clips can be loaded into the system generating new sound/colour mappings, changing the character of the system and forcing different user interactions. For example, loading an ambient sound bank encourages minimal or moderate exchanges and suits therapeutic purposes whilst loading a DJ sequencer promotes more aggressive and complex interactions as user mix multiple tracks in real-time.

The test platform currently used is displayed in Figure 1 below. This particular arrangement is composed of a slowly rotating plinth on which coloured objects are placed and a mast which houses the webcam, positioned so as to allow a section of the disc to be examined.

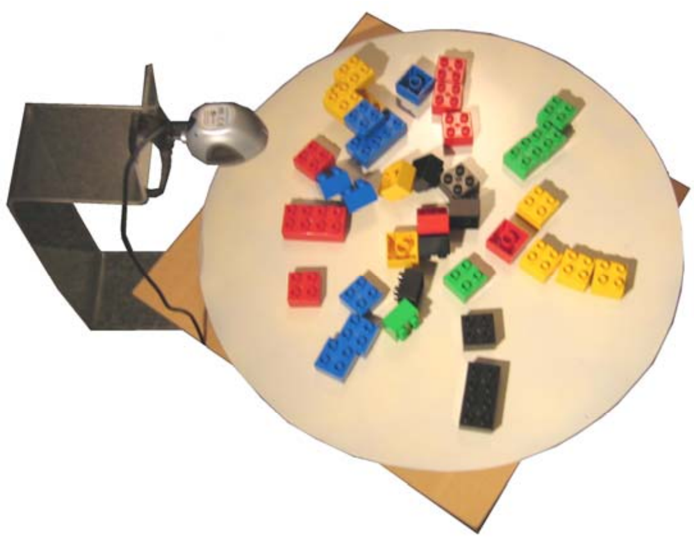

Figure 1 Test platform

In this configuration, users can carefully organise blocks in an orderly manner away from the webcams view. This can lead to structured and composed connections as clips are played in specific timing with one another. However, since the plate is continually spinning, even few or little interactions with the blocks can still result in continuous, highly textured music.

3.1 Colour Filtering and Performance

It is possible that the raw image from the webcam may contain a number of dark areas and shadows, and may lack overall definition. To address this, a series of tuning mechanisms have been implemented. This affords the user the ability to clean up the image on the fly. Any part of the image in the application can be selected for processing, making it possible to eliminate the darker regions surrounding the edges as well as limiting the number of blocks examined at any one time. Additionally, the brightness and contrast can also be adjusted in real-time. The figure below shows a series of screen captures taken from the test application.

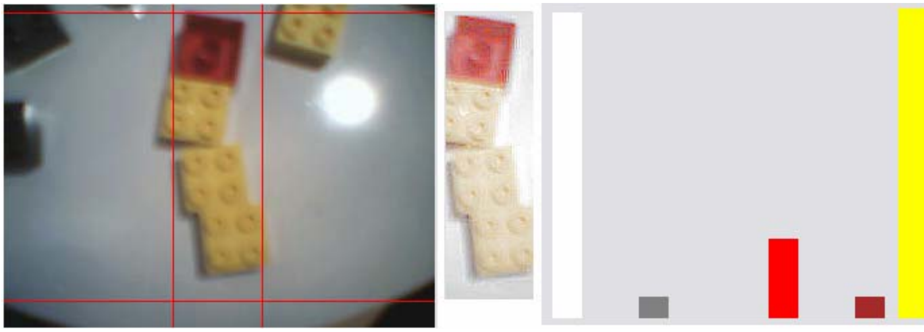

Figure 2 Screen taken before and after colour extraction

The idea of the colour filtering is to allocate pixels based on their RGB values into a discrete colour set. The method employed works by passing each RGB pixel value through a simple framework of logical operations, and after rational deduction, the system selects the colour to which it matches closest. At present, the system is able to distinguish between 10 different colours; the levels of these correspond to the volumes of the individual sound clips, and are illustrated on a dynamically changing bar chart on the interface. Adjusting the volume this way helps to make transitions smoother, between objects and areas with no objects.

3.2 Background masking

Currently MusiCam is programmed to nominate a single colour to ignore so that this can be used as the background mask. This allows the system to easily distinguish between moving or foreign objects. Without this mask, MusiCam would continue to play music with or without user interactions. Whilst this might be appropriate for a static ambient display it limits greatly the number of possible webcam arrangements. For this reason a further background masking module is introduced. The concept of this is to capture the image without the foreign objects present, and then subtract this image from all subsequent frames. By using this process it is possible to display all the pixels that have significant change (new objects in the frame) and to replace non-changing pixels with white respectively. This new image can then be processed and filtered for its levels of colour as before.

Since even the smallest changes of intensity could be detected, a thresholding equation similar to that developed in (Foyle 2004) was introduced. The values were tweaked accordingly so that only the larger changes will be displayed. The modified equation is presented below, where change in intensity is the difference in intensity between a pixel in the background mask and the corresponding pixel in the captured frame, and the new intensity refers to the intensity of the pixel in the captured fame.

$$Output = \begin{cases} DrawPixel, & if \ \frac{ChangeInIntensity^2}{(NewIntensity + 1)} > 1.02 \\ MakeWhite, & otherwise \end{cases} \quad (1)$$

Although this background removal technique works well with simple backgrounds, in more demanding and busier backgrounds the image resulting from the background removal process often does not exclusively contain the foreign blocks, it will also contain noise representing shadows, small changes in the scene and changes in lighting. The existing process is refined with the addition of another algorithm, which works on the principle that a pixel in a densely populated (non-white) row and column is more likely to belong to a group of pixels making up a foreign object than that of noise. Simply, the method works by firstly totalling the number of non-white pixels for each column and row, after this each suspect pixel is ranked by multiplying the corresponding column and row value. By using this method it is possible to eliminate some of the noise in the image. In Figure 3 a coloured block is introduced into the picture, the background is removed using the thresholding equation and then with the added noise filter to remove the light nuances.

Figure 3 Background masking and noise filtering

3.3 Multiple voices

A straightforward way of achieving much richer and interesting sound from MusiCam is to increase the number of voices being used at a time. The image can be divided horizontally into equal parts according to the number of voices selected. Each of these segments can then be assigned to a different voice or sound bank, leading to multiple colour-to-sound mappings. For instance, if the picture were to be divided into two parts, it would be entirely possible to use the same colour to trigger off two independent sounds by moving it from one half of the image to the other.

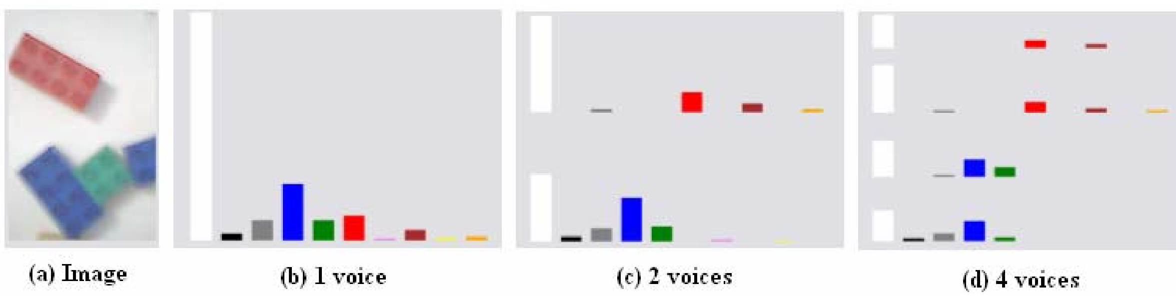

Figure 4 A demonstration of image segmentation

By using this function it is not only possible to increase the number of different instruments, but also to extend the range of a single instrument, for example by having a higher register of notes in the upper segments and vice versa.

3.3 User modes

For demonstrative purposes there are three modes which aim to exploit the diversity of the system. Each mode encourages a different type of interaction and consequently has different application potential. Demonstrations of these modes can be found in (Yau 2006).

In the first mode, collections of ambient sounds are used. This encourages minimal or moderate exchanges and suits therapeutic purposes. It is also attuned to other ambient based applications, for example by positioning the webcam in a hallway or lobby it could serve as an ambient device, playing music to people passing by.

Much different to the ambient mode is that of the synthesiser. Funky drum, guitar and bass loops make up this mode, and it promotes almost aggressive interactions as users cope with mixing multiple tracks in real-time. It is much preferred to have the platform stationary in this mode and not continuously rotating. Areas of interest can be set up in different parts of the circle and activated by spinning the plate manually, the action similar to that of a DJ spinning a turntable deck. A user can also wave the webcam like a stylus over the areas of interest, zooming in and out to gain expression.

A collection of musical notes make up the sound-banks in the last mode. Aiming to play virtual instruments much like a keyboard would, this mode has as much appeal as a toy xylophone, and thus benefits as an introductory music aid or educational toy.

4. User Feedback and Further Plans

To date, more than a dozen users have experimented informally with MusiCam and feedback has been very encouraging. As well as giving users an insight into synesthesia, users have frequently remarked on the quirky originality of the concept, and have positively given ideas, prompting the future implementation of extra features and functionality, which could heighten the user experience. An idea currently under consideration is to include sound panning. So if an object was to enter the frame on the left and leave on the right, the system would emulate this by panning the sound from left to right accordingly. The rate of object movement could also be studied and used to influence the pitch of a sound. However, both of these new additions require some form of object tracking, and hence the system would need to take into account previous states.

Indeed the novelty of using colour as a musical controller is not restricted to it being an ambient, interactive musical platform. MusiCam also has considerable scope for developing into a tool for music education, much like the Hyperscore system (Farbood 2004); a graphical compositional software utility which interprets gestures and strokes as musical ideas.

There has also been much interest in the area of therapy and assistive technology for children with special needs. One expert who specialises in building interactive rooms and spaces for children who are disabled or have sensory impairments wishes to use the technology to enhance the experience in these spaces. They consider the technology could be extremely rewarding especially for children that suffer from blindness, as colour based gestures could be used to manipulate appropriate sounds, for example blues would indicate water based clips and greens associated with fauna and meadows, and a fun interaction would encourage those previously unable to distinguish or see the beauty in natural colour. Also, studies have shown that music therapy is especially of value for pupils who through differing forms of autism find it difficult to communicate in more traditional ways (Alvin 1991). Another area of research therefore is to establish whether this type of HCI serves as a useful interaction mechanism for these individuals, as well as an interactive toy

Modern music artists nowadays use computers to create and manipulate sound. It is not unusual for a band to bring on the stage a laptop as a supporting musician. Triggering these sound clips and tracks can be to the audience visually dull and un-stimulating however using the MusiCam software as a real-time musical controller instead of the typical keyboard would allow more freedom and expression to the performer and also enhance the live music experience for the audience.

5. Conclusions

"The sound of colour is so definite that is would be hard to find anyone who would express yellow as a bass note or dark lake with treble…" - Wassily Kandinsky – Artist & Synesthete

The system developed during this project has given the chance for others to experience something quite unique. Aside from effectively demonstrating virtual synesthesia, testing and evaluation the system has demonstrated the realistic potential of MusiCam as an adaptable musical controller, that could easily grasped and learnt by a child as well as having the technical specifications required of such a controller by a professional musician.

A major benefit of the system developed is that it is easily accessible by others. Since the system is software based, and only requires the additional support of an off-the-shelf webcam and regular computer speakers, potentially any personal computer can take on the identity of an interactive musical device.

Just as Hyperscore (Farbood 2004) allows children to leap over the obstacle of musical notation and technique in order to express their compositional creativity, the MusiCam software also hopes to give people the opportunity to make instrumental music before they have achieved the mastery of the techniques of a conventional instrument.

Note 1: Colour organs are interactive musical controllers capable of creating abstract audio/visual compositions in real-time. A phrase coined by Rimington in 1893 (Peacock 1991) the earlier machines looked like typical instruments and when played controlled coloured gas lamps or coloured paper strips lit by candles.

References

Abbado, A. (1988) *Percepual correspondences of abstract animation and synthetic sounds,* Leonardo, Electronic Art Supplemental Issue, pp. 3-5

Al-Hammami, A.Z., Mehta, M., Gaumond, N. & Talapatra, A. (2000) *The design and development of musicBlox: a musical composition toy modelled for children*, Bachelors Thesis, McGill University

Alvin, J. and Warwick, A. (1991) *Music Therapy for the Autistic Child*, Oxford University Press, 2nd ed.

Cytowic, R.E. (2003) *The man who tasted shapes*, The MIT Press, 2nd ed.

Darwin, C.J., Turvey M.T. and Crowder, R.G. (1972) *An Auditory Analogue of the Sperling Partial Report Procedure: Evidence For Brief Auditory Storage*, Cognitive Psychology, 3, pp. 255-267

Farbood, M., Pasztor E. and Jennings K. (2004) *Hyperscore: A Graphical Sketchpad for Novice Composers*, IEEE Computer Graphics and Applications, 24(1), 50-54

Foyle, M. (2004) *iMo: Interaction via Motion Observation*, SCARP 2004, University of Reading

Goose, S. (1998) *Audio-Only Access to Internet Information in a Ca*r, Siemens Corporate Research, 755 College Road East, Princeton, New Jersey, USA 08540, 1998 Society of Automotive Engineers, Inc.

Machover, T. (1996) *The brain opera and active music*, Ars Electronica 96, Memesis, The Future of Evolution, New York: SpringerWien, pp. 300

MIT, *Virtual Synesthesia*, http://web.mit.edu/synesthesia/www/virtual.html, Accessed Jan 2006

Moritz, W. (1997) *The dream of colour music, and machines that made it possible*, Animation World Magazine, Issue 2.1

Peacock, K. (1991) *Famous early colour organs, experimental musical instruments*, Leonardo, pp. 397-406

Ramachandran, (2003) *Hearing Colours, Tasting Shapes, Scientific American*, May 2003, pp. 53-59

Weinberg, G. (2001) *Interconnected musical networks – Bringing Expression and Thoughtfulness to Collaborative Music Making*, Thesis proposal, Massachusetts Institute of Technology

Yau, D. (2006) *MusiCam*, http://www.musicam.co.uk/media Accessed July 1, 2006

Derek Yau was born in Swindon, England, in 1983. He has just completed a MEng degree in Cybernetics & Computer Science at The University of Reading, England. His current research interests lie in HCI technology and interactive musical environments.

Rachel McCrindle is Senior Lecturer in Multimedia and Head of the Ambient and Pervasive Intelligence Research Group at The University of Reading. Her current research interests span human computer interaction, multimedia and multimodal interfaces, inclusive design, and the social impact of science and technology on individuals and society.