Towards Building an Experiential Music Visualizer

Suranga Chandima Nanayakkara*, Elizabeth Taylor *, Lonce Wyse† and S.H. Ong‡

*Marine Mammal Research Laboratory, Tropical Marine Science Institute,
National University of Singapore, 14 Kent Ridge Road, Singapore 119223
scn@nus.edu.sg, mdcohe@leonis.nus.edu.sg

†Communications And New Media Programme,Faculty of Arts and Social Sciences
National University of Singapore,
Block AS3 #04-09, 3 Arts Link, Singapore 117570
lonce.wyse@nus.edu.sg

‡Department of Electrical and Computer Engineering, Division of Bioengineering
National University of Singapore, 9 Engineering Drive 1,
Singapore 117576
eleongsh@nus.edu.sg

Abstract— **There have been many attempts to represent music using a wide variety of different music visualisation schemes. In this paper, we propose a novel system architecture which combines Max/MSP™ with Flash™ that can be used to build real-time music visualisations rapidly. In addition, we have used this proposed architecture to develop a music visualisation scheme that uses individual notes from a MIDI keyboard or from a standard MIDI file and creates novel displays that reflect pitch, note duration, characteristics such as how hard a key is struck, and which instruments are playing at any one time. The proposed music visualization scheme is a first step towards developing a music visualization that alone can provide a sense of musical experience to the user.**

I. Introduction

Music is the time-based art of sound. Listeners bring a host of cultural and personal musical experience to bear on their experience during a piece of music.

Statistical regularities among a set of twelve tones are the fundamental blocks on which the structural regularities in western tonal music is based. A chord is defined as the simultaneous playing of three tones, while a subset of seven tones and chords generated from them defines a key [1]. Conventions of melodic patterns, chord sequences and key changes are exploited to create an intellectual and emotional response that we call the musical experience. A question central to our research is whether or not the musical experience can be conveyed by sensory channels other than sound. In this paper we will describe a system that transcodes musical sequences of information into a visual sequence. The proposed system architecture allows us to build different types of displays rapidly allowing us to experiment in finding a suitable mapping between the musical data and the visual data.

It is not just musical 'information' that we want to convey, but the musical 'experience'. For example, rhythm is physical and related to a sense of movement or dance. Can the different experience we have when listening to a march compared to when listening to a waltz be conveyed using a visual method of presentation? Are there relationships between sound and graphics so that some mappings work better than others? It is practically impossible to consider all the possible one-to-one, one-to-many and many-to-one mapping of features in the audio domain to those in the visual domain. Therefore, some guidelines from pervious works and some of our own ideas have been used in this work. The validity of such a mapping can then be verified through a detailed user study, which is currently in development.

As a starting point, we decided to focus on very basic musical features such as note (pitch and duration), volume (loudness), instrument (timbre) and key changes. Extracting note and instrument information from an audio stream is an extremely difficult problem [2] and this is not the main objective of this paper. Hence, we decided to use MIDI (Musical Instrument Digital Interface) data as the main source of information instead of live audio stream. Using MIDI makes note and instrument identification much easier.

II. Related work

The visual representation of music has a long and colourful history. Among the earliest researchers, Mitroo [3] used musical attributes such as pitch, notes, chords, velocity, loudness, etc., to create colour compositions and moving objects. Cardle *et al.* [4], extracted musical features from MIDI and audio sources to modify motion of animated objects in an ad-hoc way. Ferguson *et al.* [5] proposed a real time visualisation system that communicates data extracted from real time acoustic analysis to musicians. Moreover, commercial products such as visualisation plug-ins for windows Media Player™, as well as visualisations for helping to train singers are now available. One of the earliest attempts at visualisation for training singers was 'SINGAD' [6] and more recently 'sing and see' [7] has been available for commercial purposes. The latter uses real time spectral displays, metering and traditional notations to provide visual feedback for singing pedagogy. However, almost all of these music visualisation schemes are intended to complement the music in some way. In other words, the visualisation alone does not provide any musical experience. Our goal is to explore the possibilities of building a display generated by the music being played in as spontaneous a way as possible, and which provides a musical experience that ideally would reflect the experience derived from listening to the piece. If used with a live music stream, a successful visualisation would also be different each time a piece of music is played, thereby reflecting an individual performance.

Musical visualisation schemes can be categorized into two groups - augmented score visualisations and performance visualisations [8]. The first group focuses on generating computer graphics similar to a conventional staff-like notation and the target audience is people with good musical background. Malinowski [9] has done a lot of work on this type of visualization especially in developing a system called MAM (Music Animation Machine). Other examples incudes the works of Hiraga *et al.* [8] and Foote [10]. The above mentioned music visualisations are informative, for example they can be used in music analysis, rather than experiential and hence are not of direct relevance.

The second group of music visualisation schemes deal with different musical characteristics like volume, pitch, mood, melody, instruments, tempo etc. that can be extracted from an audio stream. Such features have been mapped to visual properties of target rendered scene. The scene typically consists of objects with various colors, positions and other attributes. Ondrej *et al.* [11] have produced a basic scheme that can map music characteristics to the parameters of visual objects. The application consists of three parts - sound analyzer, visualisation module, and scene editor. The sound analyser extracts musical parameters from audio files and the visualization module uses a particle animation scheme to create real-time animations. This sound analyzer scheme only evaluates the music volume and balance (current music stereo position). There were no attempts to analyse and extract the most effective musical features so that the visualisation is made more meaningful. However, the basic idea of a particle animation system can be used to create a sophisticated and meaningful display, provided that we are able to extract the most appropriate audio features and develop a suitable audio-visual maping.

Smith [12] has performed a music visualisation, which maps music data to 3-D space. His programme takes in MIDI data files, and tones generated by individual instruments are represented by distinct colored spheres in the visualisation. The characteristics of each sphere are dependent on three properties that describe musical tones: volume, timbre and pitch. Each note is represented as a sphere where the relative size of the sphere corresponds to the loudness, colour corresponds to the timbre of the tone and relative vertical location of the sphere corresponds to the pitch of the tone. Individual instruments are mapped to particular values along the horizontal axis. Although this music display is totally generated by the music, Smith's aim was to present an alternative method for visualising music instead of the conventional music notation. Hence Smith's visualisation only provides information about the music being played. Similar work has been done by McLeod and Wyvill [13] using real-time audio stream as the input.

On the other hand, there have been attempts to extract meaningful musical features from live performances and map them to the behavior of an animated human character in such a way that the musician's performance elicits a response from the virtual character [14] [15]. DiPaola and Arya have developed a music-driven emotionally expressive face animation system called *MusicFace* [16], to extract the affective data from a piece of music to control facial expressions. Although the analysis attempted to extract affective features, it was not their intention to elicit emotional responses in the viewers. Furthermore, most of the music visualisation schemes reported in the literature have not target hearing-impaired people. However, there have been attempts to build displays capable of providing information to the hearing-impaired about sounds in their environment. For example, Matthews *et al.* [17] have proposed a small desktop screen display with icons and spectrographs that can keep the deaf person informed about sounds viscinity. Similar work has been done by Ho-Ching *et al.* [18] where they implemented two prototypes to provide awareness of environmental sounds to deaf people. However when it comes to experiencing sounds, Ho-Ching writes:

> ... *there is still a gap between the sound experience of a hearing person and the experience of a deaf person. For example, although there are several methods used to provide awareness of certain notification sounds, there is little effective support for monitoring.*

This quotation seems especially apt for musical sounds, where experiencing the music is more important than just knowing the acoustic signal attributes.

III. Seeing Sound

Seeing colours when listening to sounds is one type of 'Synesthesia', a condition in which the real information of one sense is accompanied by a perception in another sense. This interesting phenomenon is important for our research because it can give us clues to sound-to-visual mappings that are naturally meaningful and therefore potentially useful for conveying musical experiences. Individuals who have music-to-colour synesthesia, experience colours in response to tones or other aspects of musical stimuli (e.g., timbre or key). These synesthetic experiences might be useful as guidelines to map audio data to visual features since specific audio features tend to be associated with specific visual features.

There have been a number of synesthetic composers and musicians. Among the earliest reported was Kircher, who around 1646 developed a system of correspondences between musical intervals and colors. Similar work was done by Marin Cureau de la Chambre in 1650. Sir Isaac Newton was able to show the parallel between the colour spectrum and notes on the musical scale. His work was based on Aristotle's theories and was more elaborative and mathematical. In 1742, the French Jesuit monk, Louis Castel, developed a 'light-keyboard' which would simultaneously produce both sound and what he believed to be the 'correct' associated colour for each note. Many such tone-to-color mappings have been proposed and Fred Collopy's collection of 'color scales' [19] shows some of them. Table I lists some of the better known synesthetic

TABLE I
LIST OF SYNESTHETIC COMPOSERS, MUSICIANS AND ARTISTS

Artist	**Synesthetic Experience**
Amy Beach, Nikolai Rimsky-Korsakov	Association of certain colours with certain keys
Sean Day, Joachim Raff	Association of timbres with colours
Tony DeCaprio, Brooks Kerr, Franz Liszt, Olivier Messiaen	Association of each tone as a different colour
Duke Ellington	Association of each tone as a different colour
Harley Gittleman American composer	Claims: *each tone I hear is a certain color, creating a cornucopia of compelling melodies and harmonies for which to visually merge.*
Gyrgy Ligeti	Association of chords with colours and shapes
Jennifer Paull	Association of letters and numbers with colour
Jean Sibelius	Association of every impression of sound with colour and register in his memory
Henrik Wiese	Experience of coloured music and coloured letters.
David Hockney	Association of synesthetic colours with musical stimuli
Jane Mackay	Association of musical sounds with images in her mind.
Marcia Smilack	Experience of colour-sound synesthesia in both directions. (i.e. sees colours when hears sounds and hears sounds when sees colours)

composers, musicians and artists. There is rarely agreement amongst synesthetes that a given tone will be a certain colour. However, consistent trends can be found, such that higher pitched notes are experienced as being more brightly colored [20]. In 1974, Marks [21] conducted a study in which subjects were asked to match pure tones with increasing pitch to brightnesses of gray surfaces. He found that most people would match increasing pitch with increasing brightness, while some would match increasing loudness with increasing brightness. We hope these trends can be used as a rational basis for finding a 'good' map between musical features and visual features.

IV. MUSIC VISUALISER

A. Audio-visual mapping

The main challenge in building a music generated display is to choose a suitable way of mapping musical features to visual effects. In this section we discuss a few possible audio-visual mappings and how we have used them in the display. The smaller a physical object is, the higher the frequencies it tends to produce when resonating. Hence, we will map higher pitched notes of a piece of music to smaller sized visual objects. We believe that a visualisation that maps high notes to small shapes and low notes to large shapes would be more 'natural' and intuitive than the reverse because it is consistent with our experience of the physical world. Similarly, there is a rational basis for amplitude being mapped to brightness. This is because both amplitude and brightness are measures of intensity in the audio and visual domains respectively as justified experimentally by Marks in 1974 [21].

As far as colour is concerned, although a number of 'colour scales' have been proposed [19], there is no basis for the universality of any of them. In our visualisation, we have used the 'colour scale' proposed by Belmont (1944) to represent the colours of notes. Each note produced by a non-percussive instrument is mapped to a star-like object to emphasise the note onset, and notes are arranged left-to-right in order of increasing pitch, mainly because this method mirrors the piano keyboard and allows chord structure to be visualised.

General MIDI can specify a maximum of 47 different percussive instruments. Currently we have implemented a visual effect for the sounds of 8 different percussion instruments. Most of these visual effects were designed by trial and error until we found a result that was considered satisfactory by the research team. Table II shows this visual mapping of percussion instruments.

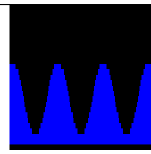
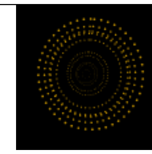
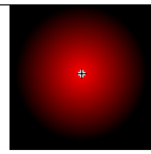
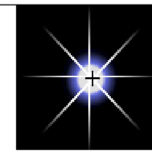
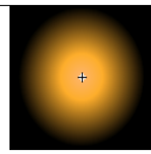
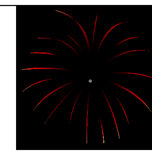
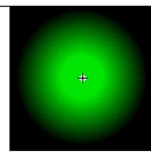
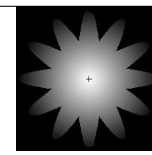
We have also introduced visual effects for the key changes that might occur during a song. Since many synesthetic artists, for example Amy Beach and Nikolai Rimsky-Korsakov, have associated musical key with colour, we decide to visualise key changes by using different backgrounds instead of representing the absolute 'value' or precise sound frequencies in a key by the height of an icon as we did with notes. Different keys function as a kind of background context for chords and notes without changing the harmonic relationship between them. This fact also supports the idea of mapping key to the background color. However, key changes are not trivially extractable from the MIDI note stream and, for the work presented in this paper, we used manually marked-up scores for key identification. In future work we intend to use a computer algorithm optimised for key extraction in real time.

B. System Architecture

The proposed music visualisation scheme consists of three main components: processing layer, server and display. The processing layer can either take in a MIDI data stream coming from an external MIDI keyboard, or read in a standard MIDI file, or generate a random MIDI stream. In this layer, Max™ [22], a graphical environment for music, audio, and multimedia has been used to process the incoming MIDI information and extract the note, velocity, instrument (whether it is a percussion instrument or not) and possible key changes. Max *midiin* and *midiparse* objects were used to capture raw MIDI data coming from a MIDI keyboard and process them, and the *detonate* object was used to deal with the standard single track MIDI files [23]. Note and velocity data can be read directly from the processed MIDI. Percussive and non-

TABLE II

AUDIO-VISUAL MAPPING OF DIFFERENT INSTRUMENTS

Instrument	Visual Effect	Instrument	Visual Effect
Bass Drum Effect: Wave (Pulsating up and down) Colour: Blue Screen position: Bottom edge of the screen		**Closed Hi-Hat** Effect: Bursting effect Colour: Yellow Screen position: Lower-Left of the screen	
Snare Drum Effect: Sphere fading away Colour: Red Screen position: Lower-Middle of the screen		**Ride Cymbal** Effect: Star falling down Colour: Silver Screen position: Top-Middle of the screen	
Hi Tom Effect: Sphere fading away Colour: Gold Screen position: Lower-Right of the screen		**Crash Cymbal** Effect: Firework effect Colour: Red Screen position: High-Middle of the screen	
Hi Bongo Effect: Sphere fading away Colour: Green Screen position: Lower-left of the screen		**Piano note** Effect: Star-like object fading away Colour and screen position: depends on the note class (C, C# ect)	

percussive sounds were separated by considering the MIDI channel number [24]. Key changes were identified using the manually marked-up scores.

The extracted musical information is passed to Adobe's Flash player™ via a max *flashsever* external [25] object. The basic functionality of *flashserver* is to establish a connection between Flash and Max/MSP. The TCP/IP socket connection that gets created allows exchange of data between both programmes in either direction over a local network or via the internet. It thus allows building interactive Max-controlled animations in Flash.

The information received from *flashserver* is used to drive action-script controlled flash animations. Each received data point will trigger a flash sub-movie clip inside the main flash movie. The shape, color and screen position of the sub-movie clips depend on the recived musical information. The basic system architecture is shown in Figure1 and a sample output of the flash display is shown in Figure2.

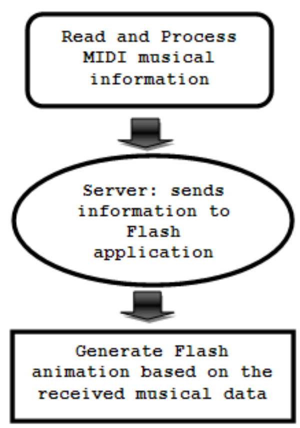


Fig. 1. System architecture of the Experiential music visualiser.

C. Discussion

In our visualisation scheme, we have used a 'movie' presentation and not a 'piano role' presentation. The 'piano role' presentation refers to a display that scrolls from left to right where events corresponding to a given time window are displayed in a single column, and past events and future events are displayed on the left side and right side, respectively, of the current time. However, we believe that this type of presentation is very different from the way people listen. When listening, people only hear the instantaneous events; hence a 'movie' presentation is more appropriate for an experiential visualization. In a 'movie'-type presentation, the entire display is used to show the instantaneous events which also gives more freedom of expression. The visual effect for a particular audio feature is visible on the screen as long as that audio feature is audible, and fades away into the screen as the audio feature fades away. This method of presentation is much more natural and makes the display experiential rather than simply informative.

The rational underlying the use of a Max-controlled flash based animation is to build different visualisations quickly. Ideally, each design has to go through a formal user evaluation before deciding what needs attention/change/improvement. However, the current design was not based on formal use

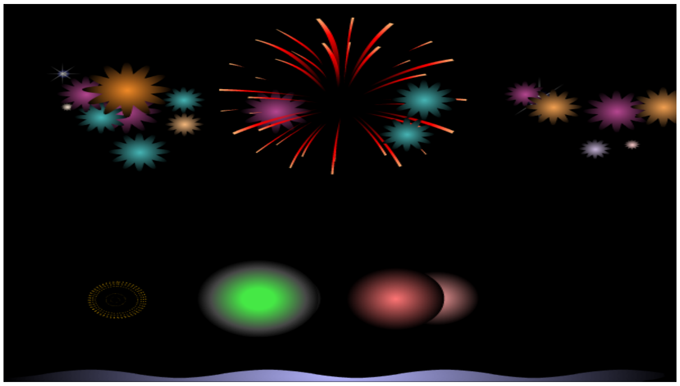


Fig. 2. Sample output of the Experiential music visualiser.

evaluation. Nevertheless, informal testing was carried out to determine what would work in general and what would not. Although this is a basic visualisation scheme, we believe that it is an useful starting point.

V. Conclusion and Future Directions

We have reviewed some of the selected music visualisation schemes from the literature and discussed possible mapping from the audio domain to the visual domain. In addition, a novel system architecture has been proposed to build interactive music visualisations.

One obvious extension to the current visualisation scheme is to incorporate more musical features. Attempts will be made to display high level musical features such as rhythm, minor versus major key, melodic contours and other qualitative aspects of the music. Making this visualisation system work for live audio streams is desirable but not yet technically feasible, and is not the focus of the current phase of the work. Key changes are not trivially extractable from the MIDI note stream and currently we are working on automatic key identification using a method developed by Chew [26] [27] based on a mathematical model for tonality called the 'Spiral Array Model'.

A survey is being prepared to gather information from people with hearing impairment about how the proposed music-driven visual display might increase their musical appreciation. Once the survey is completed, the results will serve as a basis for conceptualising approaches that move us towards understanding how best to provide musical sensory enhancement for the deaf. We also hope this approach will provide a pleasing complementary sensory experience for people with normal hearing, and that it might be useful as an aid in creating and playing music.

References

[1] B. Tillmann, S. Koelsch, N. Escoffier, E. Bigand, P. Lalitte, A. Friederici, and D. von Cramon, "Cognitive priming in sung and instrumental music: Activation of inferior frontal cortex," *NeuroImage*, vol. 31, pp. 1771–1782, 2006.

[2] E. Scheirer, "Music listening systems," Ph.D. dissertation, Massachusetts Institute of Technology, 2000.

[3] J. B. Mitroo, N. Herman, and N. I. Badler, "Movies from music: Visualizing musical compositions," *ACM SIGGRAPH*, 1979.

[4] M. Cardle, L. Barthe, S. Brooks, and P. Robinson, "Music-driven motion editing: Local motion transformations guided by music analysis," in *EGUK '02: Proceedings of the 20th UK Conference on Eurographics*. Washington, DC, USA: IEEE Computer Society, 2002, p. 38.

[5] S. Ferguson, A. V. Moere, and D. Cabrera, "Seeing sound: Real-time sound visualisation in visual feedback loops used for training musicians," in *IV '05: Proceedings of the Ninth International Conference on Information Visualisation (IV'05)*. Washington, DC, USA: IEEE Computer Society, 2005, pp. 97–102.

[6] D. M. Howard and J. A. Angus, "A comparison between singing pitching strategies of 8 to 11 year olds and trained adult singers," *Logopedics Phoniatrics Vocology*, vol. 22, no. 4, pp. 169–176, 1988.

[7] Cantare Systems, "Sing and see," http://www.singandsee.com.

[8] R. Hiraga, F. Watanabe, and I. Fujishiro, "Music learning through visualization," *Proceedings of the second international Conference on WEB Delivering of Music*, 2002.

[9] S. A. Malinowski, "Music animation machine," http://www.musanim.com/mam/mamhist.htm.

[10] J. Foote, "Visualizing music and audio using self-similarity," in *MULTIMEDIA '99: Proceedings of the seventh ACM international conference on Multimedia (Part 1)*. New York, NY, USA: ACM Press, 1999, pp. 77–80.

[11] O. Kubelka, "Interactive music visualization," Czech Technical University. [Online]. Available: http://www.cescg.org/CESCG-2000/OKubelka/

[12] S. Smith and G. Williams, "A visuaization of music," *IEEE*, 1997.

[13] P. McLeod and G. Wyvill, "Visualization of musical pitch," *Proceedings of the Computer Graphics International IEEE*, 2003.

[14] R. Taylor, P. Boulanger, and D. Torres, "Visualizing emotion in musical performance using a virtual character." in *Smart Graphics*, 2005, pp. 13–24.

[15] R. Taylor, D. Torres, and P. Boulanger, "Using music to interact with a virtual character," *International Conference on New Interfaces for Musical Expression(NIME05)*, pp. 220–223, 2005.

[16] S. DiPaola and A. Arya, "Affective Communication Remapping in Musicface System," 2005, Simon Fraser University, Surrey, BC, Canada.

[17] T. Matthews, J. Fong, and J. Mankoff, "Visualizing non-speech sounds for the deaf," *Proceedings of the ACM SIGACCESS Conference on Computers and Accessibility, ASSETS 2005, Baltimore, MD, USA*, pp. 52–59, October 2005.

[18] F. W. Ho-Ching, J. Mankoff, and J. A. Landay, "Can you see what I hear?: the design and evaluation of a peripheral sound display for the deaf." in *CHI*, G. Cockton and P. Korhonen, Eds. ACM, 2003, pp. 161–168.

[19] "Three centuries of color scales," http://rhythmiclight.com/archives/ideas/colorscales.html.

[20] J. Ward, B. Huckstep, and E. Tsakanikos, "Sound-colour synaesthesia: To what extent does it use cross-modal mechanisms common to us all?" *Cortex*, vol. 2, no. 42, pp. 264–280, 2006.

[21] L. E. Marks, "On associations of light and sound: the mediation of brightness, pitch, and loudness," *American Journal of Psychology*, vol. 87, no. 1-2, pp. 173–188, 1974.

[22] "Cycling '74 - tools for new media," http://www.cycling74.com.

[23] D. Zicarelli, G. Taylor, J. K. Clayton, Jhno, R. Dudas, and B. Nevile, "Max reference manual," 27 July 2005.

[24] "General MIDI1 Specification," http://www.midi.org/about-midi/gm/gm1_spec.shtml.

[25] O. Matthes, *flashserver External for Max/MSP*, version 1.1 ed., http://www.nullmedium.de/dev/flashserver/, 2002.

[26] E. Chew, "Towards a mathematical model of tonality," Ph.D. dissertation, Operations Research Center, Massachusetts Institute of Technology, Cambridge, MA, 2000.

[27] ——, "Modeling Tonality: Applications to Music Cognition," in *Proceedings of the 23rd Annual Meeting of the Cognitive Science Society*, J. D. Moore and K. Stenning, Eds. Edinburgh, Scotland, UK: Lawrence Erlbaum Assoc. Pub, Mahwah, NJ/London, August 1-4 2001, pp. 206–211. [Online]. Available: http://www.hcrc.ed.ac.uk/cogsci2001