Auditory-Visual Associations for Music Compositional Processes: A Survey

Kostas Giannakis & Matt Smith
School of Computing Science, Middlesex University
Bounds Green Rd, London N11 2NQ, United Kingdom
{k.giannakis; m.r.smith}@mdx.ac.uk

ABSTRACT

In this paper we give a review of auditory-visual associations as these have been investigated in computer music research and related areas. Furthermore, we reflect on the main issues that need to be addressed in order to advance further our knowledge in this area of research investigation.

1. INTRODUCTION

Music composition can be seen as the product of two distinct but complementary processes:

i) The design of individual sounds (i.e. the micro-compositional level).

ii) The arrangement of the designed sounds into a musical score (i.e. the macro-compositional level).

One of the most prominent contributions that computer music research has accomplished during the past fifty years was to provide the means for the design and control of sounds with novel timbral properties. A large number of synthesis techniques have been developed and thoroughly explored (e.g. additive synthesis, subtractive synthesis, physical modelling, granular synthesis). A common characteristic of synthesis techniques is that a sound is represented as an object consisting of a large number (hundreds or thousands) of short sub-events that can be controlled by numerous time-varying parameters. Inevitably, a vast amount of musical data must be defined and modified (ideally in real time) making the process of creating a sound object a very complex, non-musical and tedious task. Therefore, although it is theoretically possible to create almost any sound using one or more techniques, the main problem with current computer sound synthesis is how the composer interacts with the system in order to create and manipulate sound. In this paper we focus on research attempts to provide a visual metaphor for the design and control of musical sound.

There is a large number of studies and attempts to correlate musical and visual elements. We can classify these studies in three main categories:

i) Studies that are the product of computer music research (e.g. graphic sound synthesis).

ii) Studies that take place in colour vision research.

iii) Investigations of the phenomenon of *synaesthesia,* and cross-modal associations.

Associations between auditory and visual elements (e.g. Isaac Newton's colour-pitch associations) have also inspired a new artistic movement under the title of *visual music* (e.g. Wells, 1980; Goldberg and Schrack, 1986; Peacock, 1988; Whitney, 1980; Pocock-Williams, 1992). Although there has been such a cross-disciplinary interest in the investigation and application of visual metaphors for musical purposes, the quite distinct research methodologies incorporated in the above scientific and artistic domains have not facilitated interdisciplinary attempts and co-ordination of research efforts. Finally, we argue here that the above studies have mainly focused on the macro-compositional level, overlooking the design of new timbres.

2. GRAPHIC SOUND SYNTHESIS

The significant role of visual communication in computer applications is indisputable. In the case of music it seems that it is very natural for musicians to translate non-visual ideas into visual codes (see Walters (1997) for examples of graphic scores from J. Cage, K. Stockhausen, I. Xenakis, and others). In computer music research, traditional text-based synthesis (e.g. Csound) has been improved with the addition of Graphical User Interfaces (GUIs). Modern systems incorporate graphical editors (e.g. for the drawing of waveforms) and on-screen interconnections of graphical objects (e.g. oscillators, filters, etc.).

Iannis Xenakis, a pioneer of computer music, soon realised the great potential of using computers in the process of making music. In 1977 his team of engineers and programmers at CEMAMu (Centre d'Etudes de Mathematique et Automatique Musicales) developed the first UPIC system (Xenakis, 1992), a system that allowed the user to design a sound by drawing its parameters. For the purposes of this paper we will focus on UPIC's interaction with the user and sound representation scheme. At the micro-compositional level, UPIC is built on time-domain representations of sound that depict the variation of air pressure over time (i.e. the waveform) in the form of a two-dimensional graph with air pressure and time on the vertical and horizontal axes respectively. Any waveform can be drawn from simple sinusoidal waves to more complex ones (noise). The user can either draw continuously or set points that are then automatically connected by the computer. The initial version of UPIC utilised a graphics tablet (positioned in such a way as to allow natural drawing from users) as an input device on which

the user could draw all the desired sound information (parameters). However, there are two main problems with drawing waveforms. First, it requires great expertise in order to tell how a waveform will sound. In addition, two waveforms may look different but sound identical (Roads, 1996). Second, although the use of a graphics tablet is a very intuitive graphical input device, it is very hard to design a waveform precisely by hand (Nelson, 1997). The latest versions of UPIC for personal computers are based on a standard mouse which makes the task of drawing waveforms even more difficult (ibid.). The main strength of UPIC is at the macro-compositional level, where various musical structures (e.g. horizontal lines for sustained sounds, vertical lines for chords, diagonal lines for glissandi, etc.) can be created with great ease (Roads, 1996). Summarising, at the level of sound design, UPIC is based on time-domain representations of sound. In this light, UPIC takes a pure physical approach to sound that bears no connection with the perception of auditory dimensions by humans.

Phonogramme (Lesbros, 1996) is a graphic editor that translates images to sounds. The underlying image-to-sound representation is called a *phonogram* that resembles two-dimensional frequency-domain representations of sound. An interesting feature of Phonogramme is the way it handles amplitude envelopes. The starting point of a line segment gives the attack time; the grayscale levels of the pixels involved are used as amplitude coefficients (white is silence, black is maximum amplitude, levels of grey are intermediate amplitude levels); and the ending point indicates the beginning of the decay. However, no empirical evidence is given to support the above design strategy. Lesbros (1996) experimented with different kinds of physical drawing tools and techniques (e.g. ink, pencil, watercolour, etc.) to create drawings that were scanned and used as raw materials for graphic sound synthesis. However, we believe that these techniques require deeper interpretation. For example, let us suppose that we draw a line using a pencil and we get a resulting sound A. If we draw the same line using watercolour the resulting sound will be B. The purpose of drawing with watercolour was to create a certain visual effect. The question that arises is whether this visual effect is reflected by its acoustic counterpart.

Metasynth (Wenger, 1998) is a recent sound synthesis and music composition tool that translates static images (PICT files) to sound. At the level of sound design, Metasynth performs a Fast Fourier Transform (FFT) on a source sound (waveform, noise, sample, etc.) and produces a frequency-domain representation that can be altered and manipulated by applying any PICT file on it. In this sense, Metasynth functions as a subtractive synthesis tool, i.e. it starts from a spectrum rich in frequencies (e.g. noise) and uses pictures as filters to produce the desired sound result. On the macro-compositional level, a picture is scanned from left to right. Pitch and duration are represented on the vertical and horizontal axes respectively. What distinguishes Metasynth from UPIC and Phonogramme in this respect is that it also accounts for colour information. A red-yellow-green scale is used to determine the spatial position of sound while the greyscale level of pixels specifies the amplitude as in Phonogramme but in the opposite direction (white for maximum amplitude, black for silence). As in the case of UPIC and Phonogramme, the limitations of acoustic representations of sound and the lack of empirical results to support the auditory-visual associations that underlie the design strategies also apply to Metasynth.

3. COLOUR AND SOUND

In the colour vision domain, a significant number of studies (e.g. Caivano, 1994; Pridmore, 1992; Sebba, 1991), have dealt with correspondences between auditory dimensions (mainly pitch and loudness) and colour dimensions (such as hue, lightness, and saturation). However, these studies are based on correspondences that may exist between the physical characteristics of both sound and colour. For example, in Caivano's approach, hue is associated with pitch since both these dimensions are closely related to the dominant wavelengths in colour and sound spectra respectively. In the same manner, pure (or high-saturated) colours are associated with pure (or narrow bandwidth) tones whereas low-saturated colours (those that involve wider bandwidths of wavelength) are associated with complex tones and noise. Finally, colour lightness is associated with loudness (black and white represent silence and maximum loudness respectively with the greyscale representing intermediate levels of loudness). Sebba (1991) has taken a different approach. In a series of experiments, she investigated the structural correspondences that may exist between colour and music elements (as opposed to direct comparison of the elements themselves). The experiment results suggest that such structural correspondences between colour and music do exist (e.g. emotional expression, hierarchical organisation, contrast).

4. SYNAESTHESIA

One useful source of information for auditory-visual associations may be a closer investigation of the phenomenon of synaesthesia. In one of the most detailed accounts for synaesthesia to date, Marks (1997) defines synaesthesia as '...the translation of attributes of sensation from one sensory domain to another...'. The association between visual and sonic stimuli (i.e. coloured-hearing synaesthesia) is one of the most common synaesthetic conditions and manifests itself in two different but very related phenomena:

i) *Coloured vowels*, i.e. visual sensations produced by the sound of vowels,

ii) *Coloured music*, i.e. visual sensations produced by musical sound.

Marks examined a large number of reported synaesthesia studies related to coloured vowels and combined the results in order to identify general characteristics and consistencies among synaesthetes. The opponent colour model (see Fairchild, 1994) was used with the opponent colour axes being: black-white, red-green, and yellow-blue. Marks found that the black-white axis predicts vowel pitch and that the red-green axis predicts the ratio of the first two formants in the vowel spectra (the first two formants are considered to be the most important ones for vowel discrimination). In further studies (Marks, 1997:72) with musical tones, Marks reports experiments with non-synaesthete subjects that have shown associations between pitch and light intensity as well as loudness and light intensity. Although, these associations are in agreement with earlier synaesthesia studies, Marks'

overall conclusion was that it is neither pitch nor loudness that is related to light intensity, but auditory *brightness*. This conclusion is based on an assumption that auditory brightness is the same as auditory *density*, a dimension that increases when both pitch and loudness increase. However, auditory brightness has been shown to be a dimension of timbre that is determined by the upper limiting frequency and the way energy is distributed over the frequency spectrum of a sound (see Bismarck, 1974; Grey, 1975). Furthermore, a problem lies in the method behind the above-described experiments. Marks investigated only the dimension of light intensity, therefore colour hue and saturation were not considered. It is not very surprising to suggest that when people are asked to relate either pitch or loudness to a dark-light scale, they will succeed in both pitch and loudness. The question that arises is what happens when there are multiple visual and auditory dimensions for the subjects to associate.

In a recent study that involved all three colour dimensions, Giannakis & Smith (in press) suggested that pitch and loudness can be predicted by colour lightness and saturation respectively. This latter study was only concerned with pitch and loudness and involved the use of pure tones in order to neutralise the effect of timbral richness.

5. CONCLUSIONS

Although graphic sound synthesis supports the attempts for a visual metaphor to sound design, various weaknesses and limitations hamper current research in this area. Visual representations of sound such as time-domain and frequency-domain representations are based on physical approaches to sound understanding and cannot be used as intuitive conceptual metaphors for sound design. No attempt has been made to investigate the associations between abstract auditory and visual dimensions. Colour dimensions have been incorporated in a number of current computer music systems for graphic sound synthesis (e.g. Phonogramme, Metasynth). The most common association is between light intensity and loudness. The level of light intensity (how dark or light a colour is) specifies the loudness for a sound with black and white usually being the minimum and maximum values respectively. Hue and saturation have been neglected with the exception of Metasynth where a red-yellow-green scale is used to determine the spatial position of sound. However, none of the reviewed systems are based on empirical studies to support design strategies. This has resulted in a number of different approaches that in certain cases are very different and inconsistent.

Colour dimensions have been studied to a greater extent in the colour vision research area. However, none of the reviewed attempts to associate colour and auditory dimensions are based on perceptual associations that may exist between these dimensions and the physical associations that have been proposed have not been empirically investigated.

The majority of synaesthesia related studies have suggested an association between pitch and light intensity. Although this association is empirically supported, various methodological problems can be identified (e.g. other colour dimensions were not investigated in the reported experiments).

In general there is no theoretical framework for auditory-visual associations that is based on empirical studies and that can be used for intuitive sound descriptions. Most research effort to date has focused on the macro-compositional level, overlooking the design of new timbres. In fact, the dimension of timbre has been largely neglected and oversimplified. This is not surprising if we take into consideration the fact that traditional music compositional processes have focused mainly on pitch and treated timbre as a second-order attribute of sound (see Wishart, 1996). Furthermore, pitch and loudness are well understood auditory dimensions and both can be ordered on a single scale. In contrast, the perception of timbre is a more complex and multidimensional phenomenon. Many studies attempted to identify the prominent dimensions of timbre (e.g. Bismarck, 1974; Grey, 1975; Plomp, 1976; Slawson, 1985; McAdams, 1999). These studies suggest that there is a limited number of dimensions (e.g. sharpness, compactness, spectral smoothness) on which every sound can be given a value. However, there is no agreement on the dimensions of timbre that these studies proposed. This is mainly due to the different sets of sounds that were used as stimuli in the experiments (e.g. instrument tones as opposed to synthetic tones) and the different time portions of the sounds that were investigated (e.g. attack transients as opposed to steady states). As a result, our understanding of timbre is still inadequate in order to allow sound designers take full advantage of high-level sound specifications.

6. EPILOGUE

We believe that it is fundamentally important to investigate the potential contributions of visual metaphors on the micro-compositional level as well as incorporate more visual dimensions (e.g. shape, texture) in empirical investigations. Recently, we conducted a controlled experiment (for a description, see Giannakis & Smith 2000) in order to identify associations between musical timbre and dimensions of visual texture such as contrast, repetitiveness, and coarseness. The results of that experiment were very encouraging and further support our research efforts.

7. REFERENCES

Bismarck, G. von. 1974. “Sharpness as an Attribute of the Timbre of Steady Sounds”. *Acustica* 30:159-172.

Caivano, J. L. 1994. “Colour and Sound: Physical and Psychophysical relations”. *Colour Research and Application* 19(2): 126-132.

Fairchild, M. D. 1998. *Colour Appearance Models*. Addison Wesley.

Giannakis, K. and Smith, M. "Imaging Soundscapes: Identifying Cognitive Associations between Auditory and Visual Dimensions". (in press).

Giannakis, K. and Smith, M. 2000. "Towards a Theoretical Framework for Sound Synthesis based on Auditory-Visual Associations". Proceedings of the AISB'00 Symposium on Creative and Cultural Aspects and Applications of AI and Cognitive Science. University of Birmingham.

Goldberg, T. and Schrack, G. 1986. “Computer-Aided Correlation of Musical and Visual Structures.” *Leonardo* 19(1): 11-17.

Grey, J. M. 1975. *Exploration of Musical Timbre*. PhD Dissertation. Report No. STAN-M-2. CCRMA. Stanford University.

Lesbros, V. 1996. “From Images to Sounds: A Dual Representation”. *Computer Music Journal* 20(3): 59-69.

Marks, L. E. 1997. “On colored-hearing Synesthesia: Cross-modal Translations of Sensory Dimension.” In S. Baron-Cohen, J. E. Harrison, eds. *Synesthesia: Classic and contemporary readings*. Blackwell Publishers Ltd.

McAdams, S. 1999. "Perspectives on the Contribution of Timbre to Musical Structure". *Computer Music Journal*, 23(3):85-102.

Nelson, P. 1997. "The UPIC System as an Instrument of Learning". *Organised Sound* 2(1):35-42.

Peacock, K. 1988. “Instruments to Perform Colour-Music: Two Centuries of Technological Experimentation.” *Leonardo* 21(4): 397-406.

Plomp, R. 1976. *Aspects of Tone Sensation*. Academic Press.

Pocock-Williams, L. 1992. “Toward the Automatic Generation of Visual Music.” *Leonardo* 25(1): 445-452.

Pridmore, R. W. 1992. “Music and Color: Relations in the Psychophysical Perspective”. *Color Research and Application* 17(1): 57-61.

Roads, C. 1996. *The Computer Music Tutorial*. MIT Press.

Sebba, R. 1991. ”Structural Correspondence Between music and Color”. *Color Research and Application* 16(2): 81-88.

Slawson, W. 1985. *Sound Color*. University of California.

Walters, J. L. 1997. “Sound, Code, Image.” *EYE* magazine 7(26): 24-35.

Wells, A. 1980. “Music and Visual Colour: A proposed correlation.” *Leonardo* 13(1): 101-107.

Wenger 1998. *Metasynth.* Computer Music Application.

Whitney, John H. 1980. *Digital Harmony*. Byte Books.

Wishart, T. 1996. *On Sonic Art.* Revised Edition. Harwood Academic Publishers.

Xenakis, I. 1992. *Formalized Music*. Revised edition. Pendragon Press.