A comparative evaluation of auditory-visual mappings for sound visualisation

KOSTAS GIANNAKIS

P.O. Box 60572, Athens 153 05, Greece
E-mail: kgiannakis@mixedupsenses.com

1. INTRODUCTION

The significant role of visual communication in modern computer applications is indisputable. In the case of music, various attempts have been made from time to time to translate non-visual ideas into visual codes (see Walters 1997 for a collection of graphic scores from the late computer music pioneer Iannis Xenakis, John Cage, Karlheinz Stockhausen, and others). In computer music research, most current sound design tools allow the direct manipulation of visual representations of sound such as time-domain and frequency-domain representations, with the most notable examples being the UPIC system (Xenakis 1992), Phonogramme (Lesbros 1996), Lemur (Fitz and Haken 1997), and MetaSynth (Wenger 1998), among others. Associations between auditory and visual dimensions have also been extensively studied in other scientific domains such as visual perception and cognitive psychology, as well as inspired new forms of artistic expression (see, for example, Wells 1980; Goldberg and Schrack 1986; Whitney 1991).

Although there has been such a cross-disciplinary interest in the investigation and application of visual metaphors for musical or related purposes, the quite distinct research methodologies incorporated in the above disciplines have not facilitated interdisciplinary attempts and coordination of research efforts. The proposed auditory-visual associations are primarily based on subjective judgements rather than on empirical evidence. Furthermore, the evaluation of auditory-visual mappings is an issue that has generally been disregarded. This situation makes it rather difficult to find well-established methods to evaluate a particular mapping or to compare different mappings against the same evaluation criteria. This article presents efforts to set the necessary groundwork for future investigations in this area.

2. PUTTING AUDITORY-VISUAL MAPPINGS INTO CONTEXT

In recent years, there has been a growing interest in the use of visualisation for the design of interactive systems (see Shneiderman 1998; Ware 2000). The goal here is to develop visual representations that are effective in the communication of information to users. To this end, Narayanan and Hübscher (1998) assert that 'three central issues of information representation are *what* is to be represented, *how* to represent it, and how to *associate* the representation with the represented'. The remainder of this section discusses these issues and their implications for the design of cognitively useful auditory-visual mappings for sound visualisation.

2.1. Perceptual vs physical

The first issue refers to what aspects or characteristics of sound we are interested in representing and a distinction can be made here between physical and perceptual characteristics of sound. Currently, visual representations of sound such as time-domain and frequency-domain representations function at a low level of abstraction representing physical or low-level perceptual parameters of sound. For example, a waveform representation can be useful in examining the pattern of amplitude variation over time; however, there is a lack of direct relationship between the depicted amplitude variation and the perceptual attributes of sound. As a result, it requires great expertise to describe a waveform in perceptual terms or to tell how a waveform will sound. In addition, two waveforms may look different but sound identical, i.e. they have different phase relationships among their frequency components but exactly the same spectral content (Roads 1996). On the other hand, although frequency-domain representations allow us to examine various spectrum characteristics and therefore give at a theoretical level a better picture of various sound attributes than time-domain representations, no attempt has been made to incorporate models of auditory perception into the representation in a manner that supports the intuitive control of perceptual dimensions by users.

2.2. Sensory vs arbitrary

The second issue refers to what visual dimensions should be used for the visualisation of auditory information. According to Ware, the visualisation framework that underlies a representation can be

Organised Sound **11**(3): 297–307 © 2006 Cambridge University Press. Printed in the United Kingdom. doi: 10.1017/S1355771806001531

https://doi.org/10.1017/S1355771806001531 Published online by Cambridge University Press

arbitrary, sensory, or a mix of arbitrary and sensory mappings. As he points out, '… the word *sensory* is used to refer to symbols and aspects of visualisations that derive their expressive power from their ability to use the perceptual processing power of the brain without learning. The word *arbitrary* is used to define aspects of representation that must be learned, having no perceptual basis' (Ware 2000: 10).

Usually, sensory mappings are based on visual percepts such as colour, texture, space and motion. In the case of sound visualisation, common sensory mappings include the use of height and colour brightness to represent changes in frequency and amplitude, respectively, as in *sonogram* representations.

Caivano (1994) suggested a model for sensory auditory-visual mappings that is based on correspondences that may exist between physical properties of sound and colour. For example, colour hue is associated with pitch since both these dimensions are closely related to the dominant wavelengths in colour and sound spectra, respectively. However, related research in cognitive psychology has shown that such an analogy is not valid (see Kubovy 1981). 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 wavelengths) are associated with complex tones and noise. Finally, colour brightness is associated with loudness (black and white represent silence and maximum loudness, respectively, with the greyscale representing intermediate levels of loudness).

Barrass (1997) experimented with various mappings between auditory and colour dimensions and proposed a three-dimensional sound space where the auditory dimensions of timbre, brightness and pitch are associated with colour hue, saturation and brightness, respectively.

It becomes clear that an important limitation of the above attempts to derive sensory auditory-visual mappings is the lack of empirical evidence to support or validate the proposed correlations between auditory and visual dimensions. In addition, prominent dimensions of sound such as timbre have been oversimplified, thus further limiting the scope of those approaches.

2.3. Hard-coded auditory-visual mappings

The third issue in the development of effective visual representations refers to how the representation and what is being represented are associated. It can be argued that the visualisation of auditory information poses a different challenge than ordinary data (e.g. a series of stock market prices) in the sense that it could be directly related to a different modality. As such, there might be perceptually based correspondences between different modalities, in this case hearing and vision, which are hard to break and need to be identified. Partial evidence for such auditory-visual associations comes from the investigations of synaesthesia and cross-modal associations in the field of cognitive psychology.

Harrison and Baron-Cohen define synaesthesia as '… occurring when stimulation of one sensory modality automatically triggers a perception in a second modality, in the absence of any direct stimulation to this second modality' (Harrison and Baron-Cohen 1997: 3). Synaesthesia is a rare phenomenon with recent estimates ranging from 1/25,000 to 1/1,000,000 adults and is more common among women than men (Dann 1998). Although associations between various sensory modalities have been reported (for example gustatory hearing, tactile hearing, coloured smell), the association between visual and sonic stimuli is the most common synaesthetic condition. Usually, synaesthetes experience different musical notes or vowels as different colours. For example, a C note may be red, D may be blue, the vowel u may be perceived as green, and so on. In addition, high-pitched stimuli evoke lighter colours while low-pitched stimuli are experienced as darker colours. These experiences, however, are rarely in agreement among different synaesthetes.

Studies of cross-modal associations have shown similar associations for non-synaesthete subjects (see Hubbard 1996) and it can be argued that a possible generalisation of auditory-visual mappings can be made to people without the condition of synaesthesia. However, further research is needed in order to achieve conclusive results in this area.

2.4. Synopsis

Based on the above discussion, it can be argued that the limitations of current research in the area of auditory-visual mappings for sound visualisation are a function of two factors. First, current visual representations of sound are based on low-level characteristics of sound, which bear no direct relationship to perceptual experiences. Second, the associations between visual and auditory dimensions, i.e. the visualisation frameworks underlying those representations have not been empirically supported and validated.

As described next, our current knowledge of auditory and visual perception mechanisms could support efforts to incorporate high-level perceptual dimensions into the design process of interactive systems that support sound visualisation. For example, the perception of timbre has been shown to depend on spectrum characteristics and many studies suggest that there is a limited number of perceptual dimensions on which every sound can be given a value (e.g. Bismarck 1974a; Grey 1975; Plomp 1976; McAdams 1999). Visualisation can play a significant role in providing users of computer music or related tools with explicit control of high-level perceptual dimensions of sound; however, the proposed

https://doi.org/10.1017/S1355771806001531 Published online by Cambridge University Press

auditory-visual mappings should be based more on empirical support and less on arbitrary design choices.

3. SOUND MOSAICS

Sound Mosaics (Giannakis 2001) is a prototype graphical user interface (GUI) for sound synthesis based on a set of sensory auditory-visual mappings that has been formulated upon the findings of existing related studies and a series of empirical investigations that were performed for the first time in the area of computer music research.

An important aspect of Sound Mosaics is an attempt to incorporate models of auditory and visual perception in the process of sound design. Since the formulation of a complete model of auditory perception is a formidable task, currently Sound Mosaics is confined to the auditory dimensions of pitch, loudness and dimensions of timbre that pertain to the steady-state portions of sounds. Similarly, a limited, yet open-ended, set of visual dimensions, namely dimensions of colour and visual texture, is used for the control and manipulation of auditory parameters, as shown in Figure 1.

The exact auditory-visual mappings have been identified through a series of controlled experiments, where subjects from both music and non-music backgrounds were asked to perform a series of sound-image association tasks for the above sets of perceptual dimensions (see Giannakis and Smith 2001; Giannakis 2001).

3.1. Pitch, loudness and colour

Sound Mosaics treats pitch as a two-dimensional attribute of sound with the two dimensions being *pitch height* and *tone chroma*. This is primarily based on the concept of *octave equivalence*, i.e. the perceptual similarity between tones whose frequencies are separated by octaves, i.e. intervals in the ratio or a power of 2:1. The results of various studies of musical systems used in different cultures suggest that octave equivalence is culturally universal (Shepard 1964; Deutsch 1999). An octave interval can be subdivided into any number of sub-intervals (for example, in Western music the octave is usually divided in twelve equal steps as in twelve-tone equal temperament) and tone chroma refers to the interval position within the octave.



Figure 1. The Sound Mosaics synthesis environment.

In Sound Mosaics, tone chroma is associated with colour hue, whereas pitch height is associated with colour brightness in a way that it represents octave intervals for a particular tone chroma in a scale from low to high. For example, if we assume that the tone chroma of a 110 Hz tone is red, then a 220 Hz tone is also red but lighter. In addition, Sound Mosaics uses colour saturation, i.e. how weak or strong a colour is, to visualise the perceptual dimension of loudness on a scale from quiet to loud.

3.2. Timbre and visual texture

The current version of Sound Mosaics includes an explicit model of timbre perception comprising the steady-state perceptual dimensions of *sharpness*, *compactness* and *sensory dissonance*:

- **Sharpness** (other terms include auditory brightness, spectral centroid, etc.) is the most prominent dimension of timbre suggested in the related literature. For pure tones, sharpness is determined by the fundamental frequency, i.e. the higher the fundamental frequency, the greater the sharpness. In the case of complex tones, the determining factors for sharpness are the upper limiting frequency and the way energy is distributed over the frequency spectrum, i.e. the higher the frequency location of the spectral envelope centroid, the greater the sharpness (Bismarck 1974b).
- **Compactness** is a measure of a sound on a scale between complex tone and noise, i.e. the difference between discrete and continuous spectra. However, the formulation of such a scale has been proven difficult (see Bismarck 1974a). Compactness is also related to the concept of periodicity. An ideal periodic (or harmonic) spectrum contains energy only on exact integer multiples of the tone's fundamental frequency.
- **Sensory dissonance** (or roughness) occurs when two pure tones with very small difference in frequency are sounded together. In a series of experiments with pairs of pure tones, Plomp (1976) found that sensory dissonance reaches its maximal point at approximately 1/4 of the relative critical bandwidth. For complex tones, sensory dissonance can be estimated as the sum of all the

https://doi.org/10.1017/S1355771806001531 Published online by Cambridge University Press

dissonances between all pairs of frequency partials (see Sethares 1999).

Sound Mosaics provides explicit control of the above timbral dimensions using high-level perceptual dimensions of visual texture. Recently, visual texture has been proven effective when used in the visualisation of multidimensional data sets (e.g. Healey and Enns 1998; Ware 2000). In Sound Mosaics, sharp sounds are associated with coarse textures, i.e. textures comprised by large elements, whereas smaller elements are associated with dull sounds. Tone-like sounds are associated with non-granular simple textures, whereas granular and more complex textures correspond to noise-like sounds. Finally, texture repetitiveness is associated with sensory dissonance. In this case, regular textures correspond to sounds with harmonically related frequency partials, whereas the increasing deviation from the harmonic series is associated with textures increasing in their irregularity.

Visual textures produced with Sound Mosaics belong to a class of textures called synthetic due to their non-resemblance to natural textures. Synthetic textures can represent perceptual dimensions of natural textures in a controlled manner and they have been used extensively in visual texture perception studies (Heaps and Handel 1999).

As a result of this approach, Sound Mosaics allows users to create and manipulate sound in perceptual terms, thus making the process of sound synthesis a less complex and more intuitive activity than the manipulation of low-level sound characteristics. However, note should be made that Sound Mosaics is limited in the sense that it does not take into account other important auditory dimensions such as duration, spatial characteristics, and temporal dimensions of timbre. These have been left for further work leading from the research described in this article.

4. CHALLENGING THE FRQUENCY-DOMAIN PARADIGM

This section presents the design and results of a comparative evaluation study of current auditory-visual mappings for sound visualisation. Evaluation studies of auditory-visual mappings do not appear to have been previously performed in the related literature. This is a first attempt to define a set of appropriate criteria in order to investigate the strengths and limitations of current visual representations of sound.

For the purposes of this study, the sensory visual mappings used in Sound Mosaics for the auditory dimensions of pitch height, loudness, sharpness, compactness, and sensory dissonance were compared with corresponding mappings in sonogram-like representations of sound, a widely used type of frequency-domain sound visualisation. Tone chroma has been excluded from this study, as sonogram-like representations treat pitch as a one-dimensional attribute of sound and there does not seem to be an explicit visual mapping for tone chroma. Moreover, the association between tone chroma and colour hue in Sound Mosaics has not been empirically derived or validated and therefore further empirical work is needed, although colour hue is usually used to visualise nominal (i.e. categorical and non-ordered) data such as tone chroma (Travis 1991).

4.1. Experimental design

The experiment was based on a within-subjects design, where each participant performed a series of image-sound association tasks for both visualisation frameworks under investigation. The visualisation frameworks were presented in a random order for each subject in order to control possible ordering effects. The study was conducted with one group of eight subjects with no musical experience, assessed by a screening questionnaire prior to participation.

4.2. Apparatus and stimuli

A prototype computer application was designed for use in this experiment comprising an image palette and fifteen sound sequences (see Figure 2). The content of the image palette varied according to the visualisation framework presented for each task as described later in this section. The sound stimuli were sequences of five sounds based on three modes of variation for each perceptual dimension of sound under investigation, (i) ascending, (ii) descending, and (iii) non-ordered. However, since all tasks had to be performed for both visualisation frameworks, a sound stimuli set could contain only one ordered (either ascending or descending but not both) sequence for each dimension, in order to keep the experiment within reasonable time limits.

The sound stimuli were designed using a simple additive synthesis model built in Sound Mosaics that is capable of producing spectral components that do not change in their frequency or amplitude during the evolution of a sound object, an approach that is similar to what is known as fixed-waveform synthesis (Roads 1996). Although this is a limiting factor for the kinds of sounds that can be produced with Sound Mosaics, this design choice is based on the steady-state model of timbre incorporated at this stage of the research presented here. In addition, all sound stimuli for the purposes of this study had a fixed duration of 1.5 seconds and an amplitude envelope with 0.1 seconds attack and decay parts in order to eliminate the abrupt start and end of the sound.

The sound stimuli for pitch height were pure tones varying in their fundamental frequency at octave intervals (55 Hz, 110 Hz, 220 Hz, 440 Hz, 880 Hz, and 1,760 Hz) with constant amplitude levels (90 dB). In a

https://doi.org/10.1017/S1355771806001531 Published online by Cambridge University Press

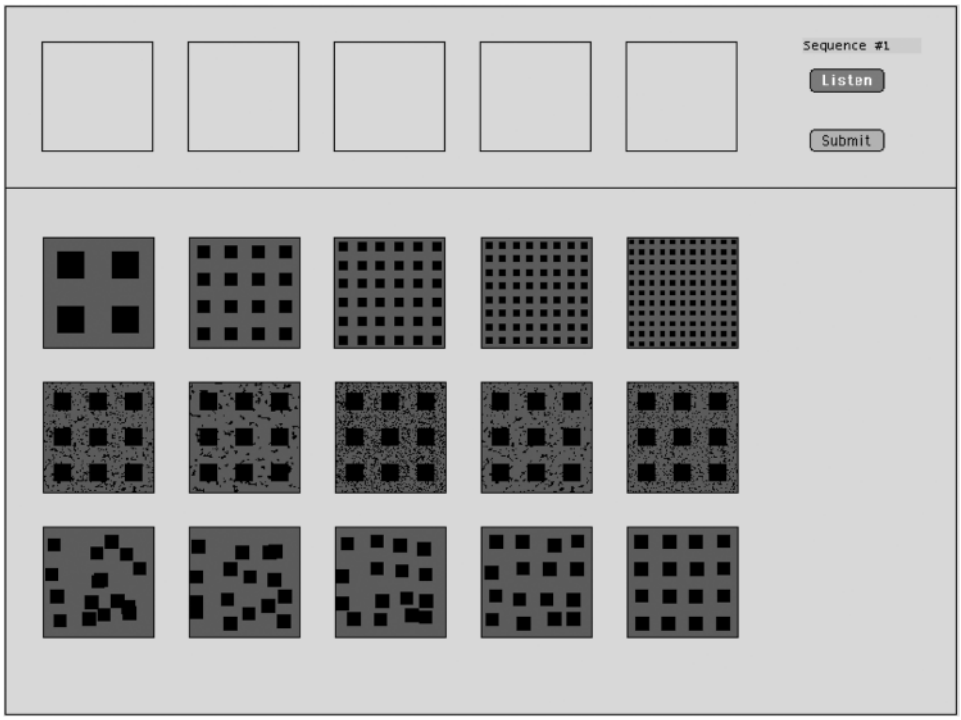


Figure 2. The prototype application used in this experiment. Subjects could drag images onto the empty display area (see top of image) after listening to a sound sequence. Images were alternating according to each visualisation framework and auditory dimension under investigation.

similar manner, pure tones with the same fundamental frequency (110 Hz) at varying levels of sound intensity were used as loudness stimuli covering a 10 dB – 90 dB range at intervals of 20 dB. For dimensions of timbre, the stimuli were complex tones varying in one timbral dimension while the remaining two were kept constant. All complex tones had the same fundamental frequency (110 Hz) and overall sound intensity levels (90 dB) in order to approximately equalise pitch height and loudness. Note should be made that although perceptual scales for the subjective quantities of pitch and loudness have been suggested in the literature (e.g. the *mel* and *sone* scales), this research has been based on the physical quantities of frequency and sound intensity as rough indications of pitch and loudness, respectively (see also Rasch and Plomp 1999). The levels of sharpness and sensory dissonance were measured using the formulae proposed in Kendall and Carterette (1996), and Hutchinson and Knopoff (1978), respectively. Compactness is a dimension of timbre related to the differences between tone-like and noise-like sounds. However, the formulation of a measurement scale for compactness has been proven difficult (Bismarck 1974). In Sound Mosaics, once the spectrum is computed for the desired sharpness and sensory dissonance levels, the compactness algorithm produces noise bands centred at the partial frequencies in order to achieve a tone-noise morphing effect (a similar approach can be found in Barrass 1997).

Figure 3 shows the visual stimuli used for Sound Mosaics. The visual representations of the sound stimuli were either colour patches that corresponded with sounds changing either in pitch height or loudness, or texture images that corresponded with sounds changing in any of the three dimensions of timbre.

The visual stimuli for the frequency-domain framework (see Figure 4) were produced with MetaSynth (Wenger 1998), a graphic sound synthesis and music composition environment that makes it possible to synthesise sounds directly from an image coming from any source. Sonogram representations created with MetaSynth are drawn on a two-dimensional plane with the vertical and horizontal axes representing frequency and time, respectively. In addition, a black-white scale is used to represent the amplitude (soft-loud) of the individual frequency components. Therefore, for pure tones, height and brightness are the visual dimensions associated with pitch height and loudness, respectively. The visual stimuli for pitch height and loudness were drawn directly on MetaSynth's *Image Synth* window for the same fundamental frequencies and amplitude levels used in Sound Mosaics.

For sounds changing in any of the three dimensions of timbre, the sound stimuli from Sound Mosaics were analysed using MetaSynth's *Filter* function. The latter performs a fast Fourier transform analysis of sound and produces a sonogram representation composed of a series of line components, where *line addition*, *pixelation*

https://doi.org/10.1017/S1355771806001531 Published online by Cambridge University Press

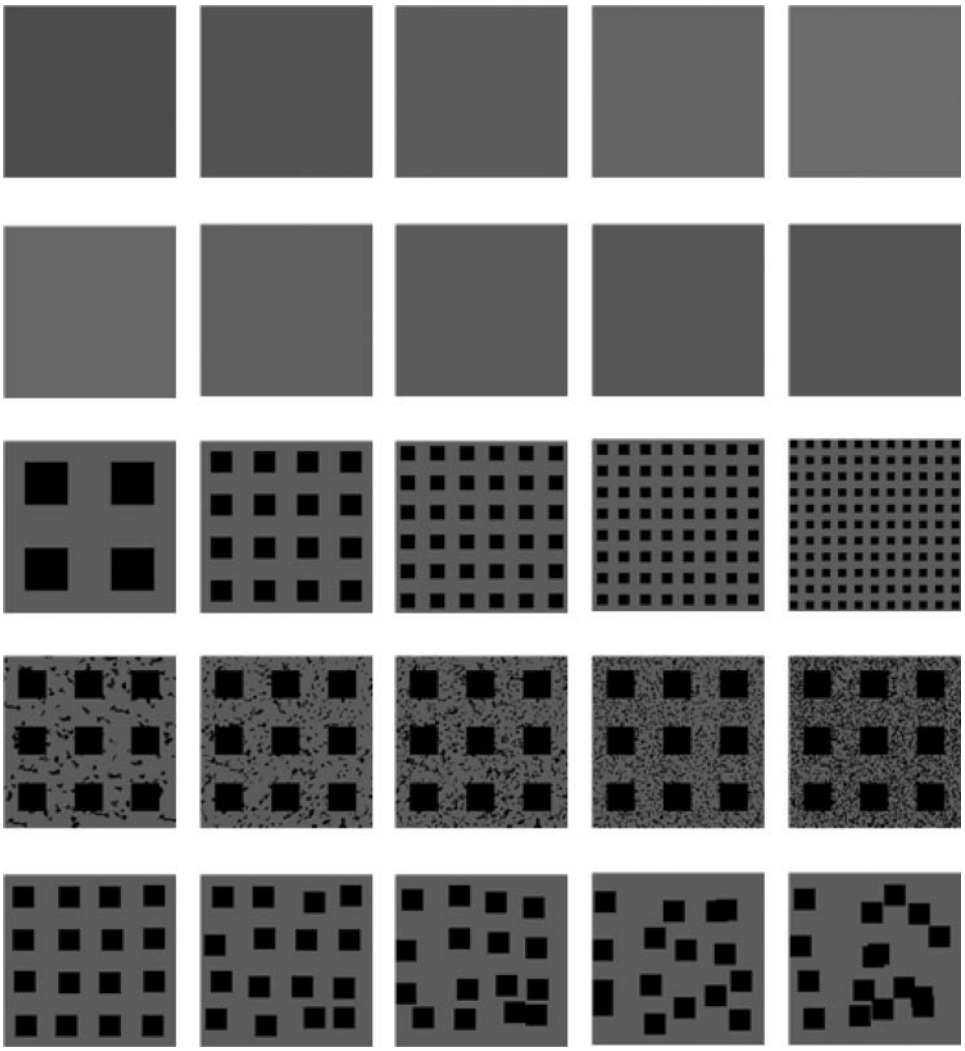


Figure 3. The visual stimuli used to form the content of the image palette when subjects were presented with the Sound Mosaics visualisation framework. From top to bottom: Brightness (dark-light), Saturation (weak-strong), Coarseness (coarse-fine), Granularity (non-granular-granular), and Periodicity (regular-irregular).

and *density* correspond to auditory sharpness, compactness, and sensory dissonance, respectively. Note that these visual terms are based on a somewhat arbitrary interpretation of sonogram-like representations. For example, since sharpness increases with the addition of higher-frequency partials, line addition seems an appropriate term to use for this dimension of timbre. Similarly, sensory dissonance increases when frequency components are closer to each other, thus it is the density of the line components that determines the degree of sensory dissonance.

4.3. Experimental task

At the beginning of each session, the experimenter demonstrated how to use the computer-based application and there was a short practice period for subjects to familiarise themselves with the task and the image and sound stimuli incorporated in this experiment. The experimental task was: for the current sequence of five sounds to create a sequence of five corresponding images selected from the current image palette. Before proceeding to the next sound sequence, subjects were instructed to specify their level of confidence for their current sound-image associations by filling a short on-screen questionnaire. During the experiment, both image and sound stimuli were introduced in a random order for each subject. Each subject completed the task for twenty sound sequences (ten for each visualisation framework). The experimenter was present throughout the experiment recording observations that formed the basis for post-experiment interviews with subjects.

4.4. Experimental environment

The experiment was conducted in a room with normal 'office' lighting and sounds were presented binaurally through headphones. The experiment was designed and

https://doi.org/10.1017/S1355771806001531 Published online by Cambridge University Press

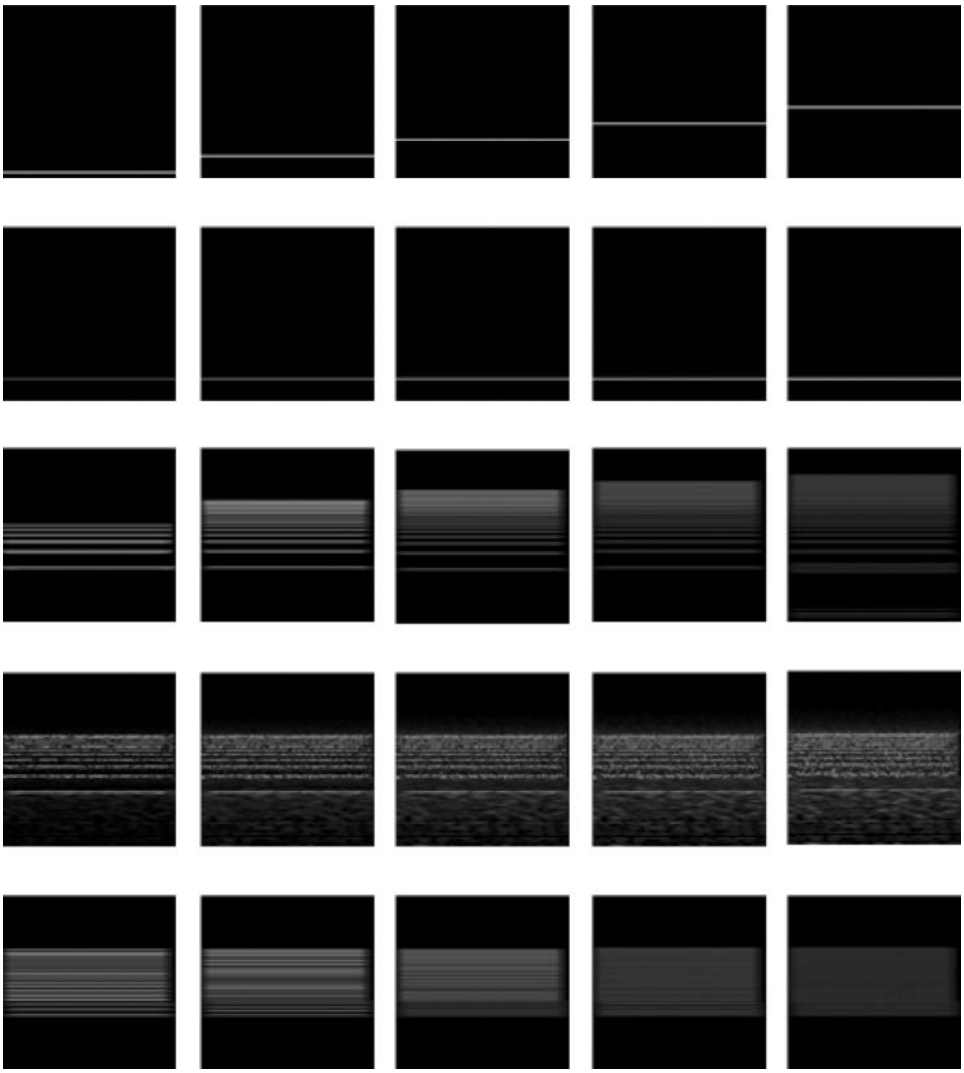


Figure 4. The visual stimuli used to form the content of the image palette when subjects were presented with the frequency-domain visualisation framework. From top to bottom: Height (low-high), Brightness (dark-light), Line addition (less-more harmonic partials), Pixelation (non-granular-granular), and Density (sparse-dense).

run on an Apple Power Macintosh G3 personal computer. Subjects sat approximately 80 cm away from the computer screen and the components of the interface were sized for comfortable viewing and manipulation at that distance.

4.5. Evaluation criteria

The main criteria used in this study to evaluate the Sound Mosaics and frequency-domain visualisation frameworks were comprehensibility and learnability. Comprehensibility was related to the accuracy of the auditory-visual mappings performed by subjects. An auditory-visual mapping was considered accurate if the subject had selected at least three images from the correct visual dimension as a response to the varying auditory dimension in terms of the corresponding visualisation framework. Table 1 shows the target auditory-visual mappings for both visualisation frameworks. Learnability was related to task completion times and levels of confidence as logged by the application for each subject.

4.6. Results

This section presents the results obtained from the above-described experiment for each evaluation criterion. For simplicity of presentation, the Sound Mosaics and frequency-domain frameworks are abbreviated as SM and FD, respectively.

4.6.1. Comprehensibility

Tables 2 and 3 show the overall results obtained for the SM and FD visualisation frameworks, respectively. In the case of visual mappings for pitch height, accuracy

https://doi.org/10.1017/S1355771806001531 Published online by Cambridge University Press

Table 1. The target auditory-visual mappings for the Sound Mosaics and frequency-domain visualisation frameworks.

	Visual	
Auditory	Sound mosaics	Frequency-domain
Pitch height	Color brightness	Line height
Loudness	Color saturation	Color brightness
Sharpness	Texture coarseness	Line addition
Compactness	Texture granularity	Pixelation
Sensory dissonance	Texture repetitiveness	Line density

Table 2. Overall results for the Sound Mosaics framework obtained by non-music subjects.

Parameter	Pitch height	Loudness	Sharpness	Compactness	S. Dissonance
P	Associations with *p* (%)	Associations with *p* (%)	Associations with *p* (%)	Associations with *p* (%)	Associations with *p* (%)
Brightness	**50**	18.75	–	–	–
Saturation	50	**81.25**	–	–	–
Coarseness	–	–	**75**	6.25	18.75
Granularity	–	–	6.25	**87.50**	6.25
Repetitiveness	–	–	18.75	6.25	**68.75**
Mixed	–	–	–	–	6.25
None	–	–	–	–	–
Totals	100	100	100	100	100

Table 3. Overall results for the frequency-domain framework obtained by non-music subjects.

Parameter	Pitch height	Loudness	Sharpness	Compactness	S. Dissonance
P	Associations with *p* (%)	Associations with *p* (%)	Associations with *p* (%)	Associations with *p* (%)	Associations with *p* (%)
Height	**81.25**	25	–	–	–
Brightness	18.75	**75**	–	–	–
Line addition	–	–	**50**	6.25	31.25
Pixelation	–	–	12.50	**81.25**	31.25
Density	–	–	37.50	–	**31.25**
Mixed	–	–	–	12.50	6.25
None	–	–	–	–	–
Totals	100	100	100	100	100

levels for SM reached 50%, whereas FD representations scored 81.25%. The SM mapping for loudness reached 81.25% with FD representations following at 75%.

In more detail, the use of height in FD representations to represent pitch height appears to be very strong and outperformed the corresponding SM mapping, a result that was expected since the low-high metaphor for pitch representation in music (e.g. the musical staff) is very strong in cultural terms. However, it can be argued that the association between pitch height and brightness is based more on perceptual reality than on cultural factors and as such has cross-cultural validity (see Ware 2000).

The accuracy levels for loudness strongly suggest that the auditory-visual mappings used in both frameworks to represent this auditory dimension are very comprehensible. However, a very important paradox can be observed for the sequences varying in loudness. Although subjects associated brightness with loudness when presented with FD representations, they associated loudness with saturation instead of brightness when presented with representations from SM.

As discussed earlier, the association between loudness and brightness in FD representations is primarily based on the physical correspondence between the two dimensions. Furthermore, brightness as a sensory channel can represent data in an ordered manner and thus it is suitable for the visualisation of ordinal data such as loudness values. However, in the SM framework, both brightness and saturation are used for the visualisation of ordered information, and therefore it can be argued that the subjects' preference of a

https://doi.org/10.1017/S1355771806001531 Published online by Cambridge University Press

saturation-loudness association can be explained either as a preference of saturation over brightness or as a direct association between saturation and loudness. The higher accuracy levels achieved with SM provide further support for this latter case.

The SM framework scored very high accuracy levels for all dimensions of timbre (75% for sharpness, 87.5% for compactness, and 68.75% for sensory dissonance) indicating that the underlying auditory-visual mappings were very comprehensible. In contrast, the results for FD representations show low levels of accuracy for the dimensions of sharpness and sensory dissonance (50% and 31.25%, respectively) and, although the results are better for compactness sequences (81.25%), they are lower than the corresponding SM results (87.5%). These results suggest that the SM framework was more comprehensible than the FD for all dimensions of timbre.

4.6.2. Learnability

Figures 5 and 6 show median completion times and median confidence levels per task as obtained for both visualisation frameworks. The results show that subjects were faster for SM representations in seven out of ten cases (stimuli 1, 2, 4, 5, 6, 7, 9) compared to two cases where FD was faster (stimuli 3 and 8) and one case where the two frameworks appear to be equal (stimulus

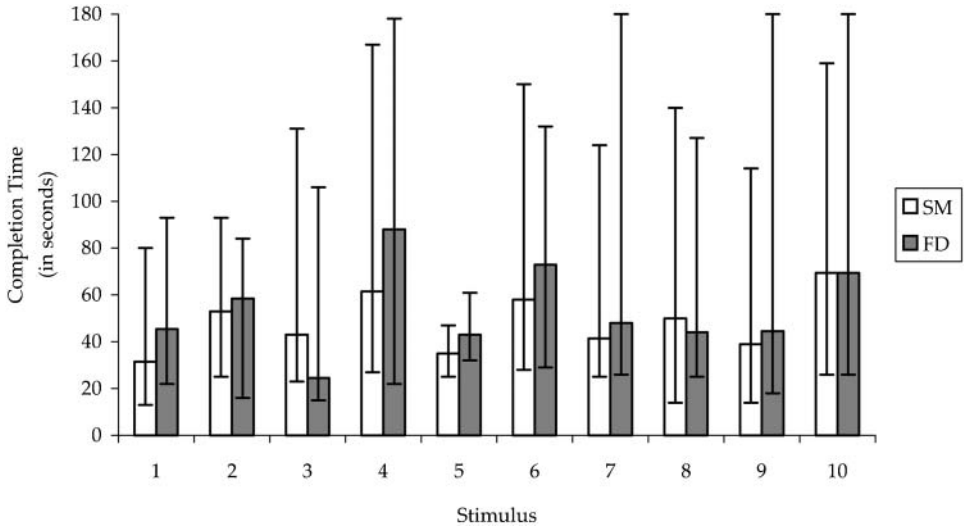


Figure 5. Maximum, minimum and median completion times per task obtained by non-music subjects for both visualisation frameworks.

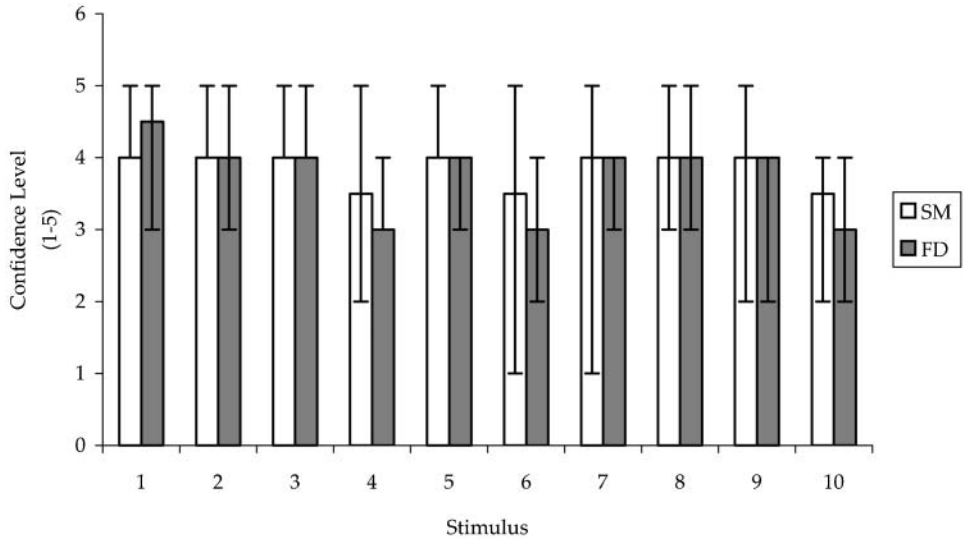


Figure 6. Maximum, minimum and median confidence levels per task obtained by non-music subjects for both visualisation frameworks.

https://doi.org/10.1017/S1355771806001531 Published online by Cambridge University Press

10). In addition, subjects felt more confident creating sequences with SM than FD visual representations in three out of ten cases (stimuli 4, 6, 10) compared to one case where FD was faster (stimulus 1), although in the remaining cases, confidence levels were equal for both visualisation frameworks.

5. CONCLUSION

Results of a comparative evaluation of sensory auditory-visual mappings have indicated that the visualisation of auditory information can be enhanced by using empirically derived high-level mappings as opposed to arbitrary associations or correspondences between the physical characteristics of sound and vision.

Visual representations created with Sound Mosaics were very comprehensible, especially for the dimensions of timbre, thus providing encouraging results for the design of computer music tools that support the intuitive exploration of sound-spaces without requiring expertise. In contrast, the evaluation of sonogram-like representations has yielded less satisfactory results in terms of comprehensibility and learnability when considering steady-state dimensions of timbre.

A shortcoming of the empirical investigation presented in this article is the limited statistical significance that can be placed on the obtained results as a consequence of having a small sample size that does not allow more in-depth statistical analysis and generalisation of the results to a larger population. However, the lack of empirical studies of auditory-visual mappings and the limitations of existing approaches in this area set the stage for more exploratory-oriented research rather than rigorous empirical investigations. Note should be made that although the incorporated models of auditory and visual perception are of course only partial, since various important perceptual dimensions (e.g. temporal characteristics) were not investigated, they are at the same time open-ended allowing further dimensions to be added through further work.

Finally, one goal of this study is to support the ever-increasing opinion that computer music research could be highly benefited from attempts to design effective interactive systems in other research fields, such as human-computer interaction. In this light, the evaluation of design choices should form an integral part of the design process.

REFERENCES

Barrass, S. 1997. *Auditory Information Design*. Ph.D. Thesis, The Australian National University.

Bismarck, G. von. 1974a. Timbre of steady sounds: a factorial investigation of its verbal attributes. *Acustica* **30**(3): 146–59.

Bismarck, G. von. 1974b. Sharpness as an attribute of the timbre of steady sounds. *Acustica* **30**(3): 159–72.

Caivano, J. L. 1994. Color and sound: physical and psychophysical relations. *Color Research and Application* **1**(2): 126–32.

Dann, K. T. 1998. *Bright Colors Falsely Seen: Synaesthesia and the Search for Transcendental Knowledge*. New Haven and London: Yale University Press.

Deutsch, D. 1999. The processing of pitch combinations. In D. Deutsch (ed.) *The Psychology of Music*. San Diego: Academic Press.

Fitz, K., and Haken, L. 1997. Sinusoidal modeling and manipulation using Lemur. *Computer Music Journal* **20**(4): 44–59.

Giannakis, K. 2001. *Sound Mosaics: A Graphical User Interface for Sound Synthesis based on Auditory-Visual Associations*. Ph.D. Thesis, Middlesex University.

Giannakis, K., and Smith, M. 2001. Imaging soundscapes: identifying cognitive associations between auditory and visual dimensions. In R. I. Godøy and H. Jørgensen (eds.) *Musical Imagery*. Lisse: Swets and Zeitlinger.

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*. Ph.D. Thesis, Stanford University.

Harrison, J. E., and Baron-Cohen, S. 1997. Synaesthesia: an introduction. In S. Baron-Cohen and J. E. Harrison (eds.) *Synaesthesia*. Oxford: Blackwell Publishers.

Healey, C., and Enns, J. 1998. Building perceptual textures to visualize multidimensional datasets. *Proc. of the 1998 IEEE Visualization Conf.*, pp. 111–18.

Heaps, C., and Handel, S. 1999. Similarity and features of natural textures. *Journal of Experimental Psychology: Human Perception and Performance* **25**(2): 299–320.

Hubbard, T. L. 1996. Synesthesia-like mappings of lightness, pitch, and melodic interval. *American Journal of Psychology* **109**(2): 219–38.

Hutchinson, W., and Knopoff, L. 1978. The acoustic component of Western consonance. *Interface* **7**(1): 1–29.

Kendall, R., and Carterette, E. C. 1996. Difference thresholds for timbre related to spectral centroid. *Proc. of the Fourth Int. Conf. on Music Perception and Cognition*, pp. 91–5.

Kubovy, M. 1981. Integral and separable dimensions and the theory of indispensable attributes. In M. Kubovy and J. Pomerantz (eds.) *Perceptual Organization*. Hillsdale: Lawrence Erlbaum.

Lesbros, V. 1996. From images to sounds: a dual representation. *Computer Music Journal* **20**(3): 59–69.

McAdams, S. 1999. Perspectives on the contribution of timbre to musical structure. *Computer Music Journal* **23**(3): 85–102.

Marks, L. E. 1997. On colored-hearing synesthesia: cross-modal translations of sensory dimension. In S. Baron-Cohen and J. E. Harrison (eds.) *Synaesthesia*. Oxford: Blackwell Publishers.

Narayanan, N. H., and Hübscher, R. 1998. Visual language theory: towards a human-computer interaction perspective. In K. Marriott and B. Meyer (eds.) *Visual Language Theory*. New York: Springer-Verlag.

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

https://doi.org/10.1017/S1355771806001531 Published online by Cambridge University Press

Pridmore, R. W. 1992. Music and color: relations in the psychophysical perspective. *Color Research and Application* **17**(1): 57–61.

Rasch, R., and Plomp, R. 1999. The perception of musical tones. In D. Deutsch (ed.) *The Psychology of Music*. San Diego: Academic Press.

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

Shepard, R. 1964. Circularity in judgements of relative pitch. *Journal of the Acoustical Society of America* **36**(12): 2,346–2,353.

Shneiderman, B. 1998. *Designing the User Interface: Strategies for Effective Human-Computer Interaction*. Reading: Addison-Wesley Longman.

Travis, D. 1991. *Effective Color Displays: Theory and Practice*. London: Academic Press.

Walters, J. L. 1997. Sound, code, image. *EYE* **7**(26): 24–35.

Ware, C. 2000. *Information Visualization: Perception for Design*. San Diego: Academic Press.

Wells, A. 1980. Music and visual color: a proposed correlation. *Leonardo* **13**(1): 101–7.

Wenger, E. 1998. *MetaSynth*. U & I Software, http://www.uisoftware.com/

Whitney, J. H. 1991. Fifty years of composing computer music and graphics: how time's new solid state tractability has changed audio-visual perspectives. *Leonardo* **24**(5): 597–9.

Xenakis, I. 1992. *Formalized Music*. Stuyvesant, New York: Pendragon Press.

https://doi.org/10.1017/S1355771806001531 Published online by Cambridge University Press