# Exploration of the correspondence between visual and acoustic parameter spaces

David Gerhard Daryl Hepting Matthew McKague

{gerhard, dhh, mckaguem}@cs.uregina.ca
University of Regina, 3737 Wascana Parkway
Regina, SK S4S 0A2 CANADA

## ABSTRACT

This paper describes an approach to match visual and acoustic parameters to produce an animated musical expression. Music may be generated to correspond to animation, as described here; imagery may be created to correspond to music; or both may be developed simultaneously. This approach is intended to provide new tools to facilitate both collaboration between visual artists and musicians and examination of perceptual issues between visual and acoustic media. As a proof-of-concept, a complete example is developed with linear fractals as a basis for the animation, and arranged rhythmic loops for the music. Since both visual and acoustic elements in the example are generated from concise specifications, the potential of this approach to create new works through parameter space exploration is accentuated, however, there are opportunities for application to a wide variety of source material. These additional applications are also discussed, along with issues encountered in development of the example.

## Keywords

Multimedia creation and interaction, parameter space, visualization, sonification.

## 1. INTRODUCTION

A discussion of parameters with respect to images or music is largely a product of the modern computer era, especially in the domain of new media. In this case, the definition of a parameter as “one of a set of independent variables that express the coordinates of a point”, is appropriate. Parameters such as tone and timbre describe particular musical notes and locate them within a larger space of all musical possibilities. The same is true for images, which can be parameterized by colour and shape. Adding a temporal dimension introduces the notion of animations and musical compositions as paths through parameter space. Manovich [8] discusses programmability as a key feature of new media. A parametric representation of visual and acoustical elements enables an algorithmic manipulation of the aesthetic space.

Considering the variety of music available, it is clear that the size of the parameter space involved is huge. Because of the number of possibilities, it may be difficult for a user to navigate—to find an aesthetically pleasing path—through the space, which would result in a work of art. In his book called *Digital Mantras*, Holtzman [5] gave many examples of artists exploring parameter spaces algorithmically. The term *algorist* has been used to describe an artist whose work is defined by the algorithms they employ. Composers make use of constraints to limit their exploration to a reasonable number of possibilities. In the same way, a system designed to explore the common parameter space between sound and vision must allow the user to constrain the possibilities to a well-defined subspace of particular interest.

While there have been many attempts to generate either graphical or acoustic works through the exploration of parameters, relatively little has been done to explore or establish correspondences *between* the two media. As an example, we are investigating the mapping of graphical parameters into musical constructs, specifically, examining whether an exploration of parameters in one medium that leads to a desired expression can be used as the basis for an exploration of parameters in the other medium. This work extends the *cogito* system [3].

## 2. MOTIVATION

More than visualization of a sound, more than synchronization of acoustic and visual objects, we are interested in creating novel visual and acoustic effects which are related to a central set of parameters under the control of the composer.

The concept of sound visualization is well understood, and many research and commercial applications are available. Typically, acoustic parameters are extracted from the waveform and corresponding graphics are generated. There are a couple of problems with this form of visualization: There is often a slight but perceptual lag between the perceived audio and the corresponding visualization; and visual components are often generated that are unrelated to the sound. Additionally, the data flow in these visualizations is inherently unidirectional—that is, the generated visuals have no effect on the sounds used to generate them. This is reasonable, since the source sounds are pre-recorded and not intended to be manipulated.

Wayne Lytle has created a series of animations based on musical pieces. He says about his work [7]: “Our technique differs significantly from reactive sound visualization technology, as made popular by music player plug-ins. Rather than reacting to sound with undulating shapes, our animation is correlated to the music at a note-for-note granularity, based on a non-real-time analysis pre-process.”

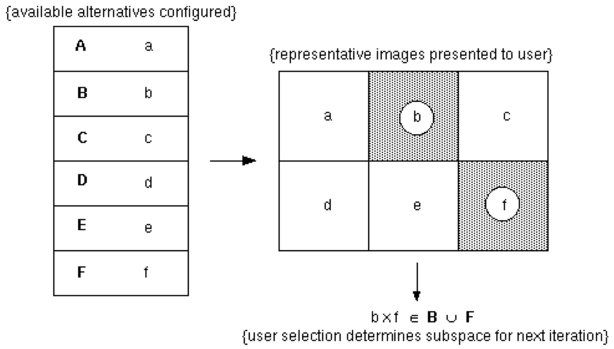

**Figure 1: Schematic look at the *cogito* interface.**

Sonification is a complementary technique, where sound is generated based on numerical or visual parameters [10, 6]. An example is the musical score: commercial systems exist which will scan a piece of sheet music and translate the resulting image into a musical representation such as MIDI or MusicXML [2]. A more general application is composition by time-frequency diagrams, where a composer "draws" the score on a piece of paper in a time-frequency representation, which is scanned and translated into music by computer. This method is not real-time and requires a development–evaluation cycle where a figure is drawn first, and after the figure is complete the new composition can be heard.

Although the translation of data from one medium to another is a well-established research area, one medium usually has a strictly defined format from which parameters are extracted to generate the second medium. Little has been done in the way of perceptual linkages between these spaces, or the simultaneous or bidirectional generation of visual and acoustic objects from a common set of parameters.

The *cogito* system illustrated in Figure 1 has proven to be a useful tool in the exploration of parameter spaces [4]. In this figure, the space of available alternatives is grouped according to user-specified criteria. Each group (A – F) has a representative element (a – f) which is displayed to the user. The subspace for the next search iteration is based on the user selection (b and f).

## 3. FRACTALS AND RHYTHMIC LOOPS

To illustrate our approach to establishing correspondences between visual and acoustic parameter spaces, we investigate the concurrent sonification of fractal transitions. Fractals have the property of database amplification [9]: from very concise descriptions come complex shapes. In this paper, we use linear fractals [1] because of their flexibility and the huge parameter space. Rhythmic loops are used to explore a varied musical parameter space while remaining rhythmically centered. The video which will be described throughout this paper is available at:
`www2.cs.uregina.ca/~gerhard/research/vaps.mov`

We begin with the Sierpiński gasket (Figure 2), generated by iteratively applying the transformations in Table 1.

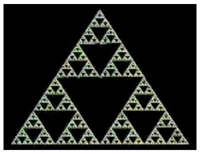

**Figure 2: The Sierpiński gasket.**

$$T_1 = \begin{bmatrix} 0.500 & 0.000 & 0.000 \\ 0.000 & 0.500 & 0.000 \\ 0.000 & 0.000 & 1.000 \end{bmatrix}$$

$$T_2 = \begin{bmatrix} 0.500 & 0.000 & 0.000 \\ 0.000 & 0.500 & 0.000 \\ 0.250 & 0.433 & 1.000 \end{bmatrix}$$

$$T_3 = \begin{bmatrix} 0.500 & 0.000 & 0.000 \\ 0.000 & 0.500 & 0.000 \\ 0.500 & 0.000 & 1.000 \end{bmatrix}$$

**Table 1: Transformations for the Sierpiński gasket**

The top left $2 \times 2$ portion of each matrix controls the rotation and scaling. The lower left $2 \times 1$ portion controls the position. Although the parameter space can be explored by modifying these matrices directly, the connection to the perceptual parameters of scaling and rotation is lost.

The fractal patterns are constructed using a set of parameters. Each transformation can be scaled $(-1..1)$ in $x$ and $y$ dimensions and rotated $(-\pi..\pi]$, giving a set of 9 parameters. Each frame is generated from the set of parameters, without keyframe generation. Parameter tracks for the sample video are shown in Figure 6. The planning of these fractal transitions is done using the *cogito* system illustrated in Figure 1. Figure 3 shows the root position of the exploration of the parameter space, with samples of the fractals generated by each type of rotation. Depicted here are rotations in one, two, and three of the transformations given in Table 1. From the eight samples shown, two are selected (framed in white) for further exploration (see Figure 4).

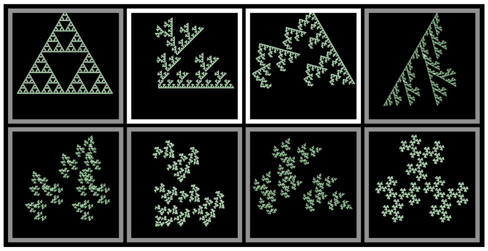

**Figure 3: Initial level of the *cogito* exploration of the parameter space, partitioned by rotation.**

## 4. PARAMETER TRANSLATION

The key to the translation between visual and acoustic parameter spaces is the two-level parameter translation system, shown in Figure 5. Here, Medium 2 is generated both

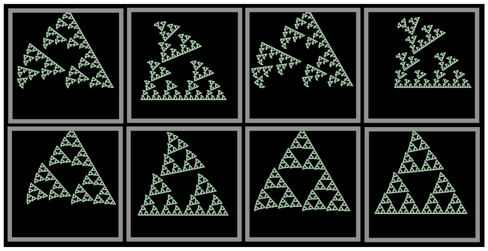

**Figure 4: Subsequent level (see Figure 3)**

from a mapping of the user control parameters and from parameters extracted from Medium 1. In our current example, this corresponds to generating music from a mapping of the user-controlled visual parameters and from parameters extracted from the graphics. The user has direct control over Medium 1, in this case the graphics, and there are two levels of control over the generated music (Medium 2): the mapping from the graphics parameters, and the extraction of the graphics parameters.

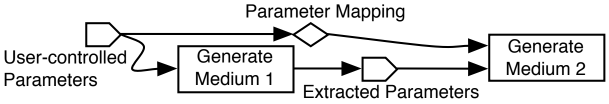

**Figure 5: Parameter mapping.**

This proposed method is different from standard visualization, sonification or combined approaches because in traditional methods, the visuals and sounds are generated entirely from a static data set. Often the results are combined for a visual-acoustic representation, but in most cases this is done to increase the dimensionality available for human interpretation. In our proposed system, it is the interactions between the visual and the acoustic representations of the same data which is interesting, as well as how the human interaction with one medium affects the other.

Rhythm is a feature of both vision and sound that is relatively consistent between sensory domains. People sing and dance to the same music, and music videos add a visual dimension to an inherently acoustic domain. It is the rhythmic nature of the music that allows a connection between these domains. For this reason we have chosen rhythm as an underlying parameter which will drive both the visual and acoustic generation at the lowest level. The visual characteristics of the fractal are set to move in cycles, and the underlying beat of the music follows the same cyclic patterns. The parameters can be aligned to hold a single value for a cycle, to vary linearly (for rotations) through a whole number multiple of periods, or to vary according to a beat pulse. Although the two media diverge at higher levels of detail, appropriate parameter extraction can be used to maintain perceptual linkages between them. In this example, the individual transforms of the gasket are rotated through a full cycle to generate one "bar" of the visual piece. The audio piece is generated to match this visualization, with one bar of music relating to a whole number multiple of complete rotations. Perceptual parameters are then extracted from the visual piece and used to generate corresponding audio events.

Figure 6 shows a the parameter tracks used to generate the sample animation. $R(\{T1, T2, T3\})$ are the rotation coefficients for each transformation. Since $-\pi = \pi$, there is no actual discontinuity at frame 32, and the rotation proceeds smoothly around the complete cycle. $S(\text{fig})$ is a scale factor for the overall figure, which is increased to 115% briefly every 32 frames, giving an overall pulse which synchronizes with the on-beat in the audio track. $S(T)$ is a scale factor applied to each transformation, taking place every 32 frames starting at frame 16. This provides a secondary pulse relating to the off-beat.

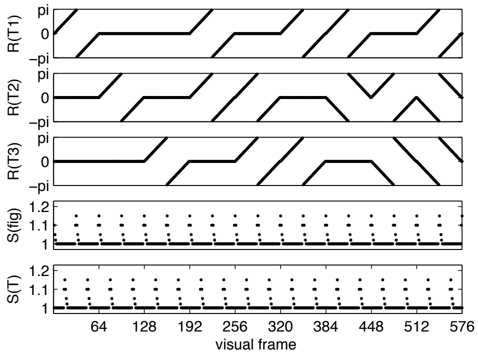

**Figure 6: Value tracks for 5 parameters**

A set of parameters is now extracted from the visualization, and musical events are generated corresponding to the visual events. Figure 7 shows a series of visual events that can be detected with image processing techniques. A straightforward visual parameter that can be extracted is the *connectedness* of the visual. Figure 8 shows the difference between connected forms and "dusts" which are asymptotically disconnected.

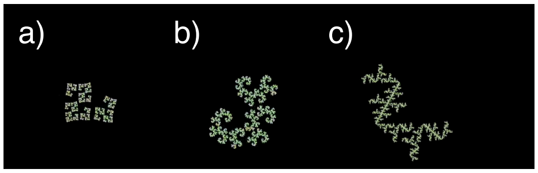

**Figure 7: Forms: a) squares, b) spirals, c) lines.**

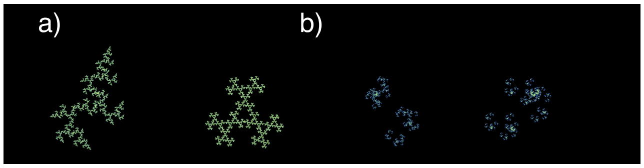

**Figure 8: a) connected forms, b) dusts.**

For this example, we have assigned high-frequency, spectrally sparse, events to occur when a dust forms, and low-frequency percussive events to occur whenever the gasket re-forms. Additionally, when spirals are detected, a mid-range "scratch" event is initiated, and when lines or boxes are detected, a low-frequency "bass" note is played.

The key to this method is to have appropriate and justifiable linkages between the visual and auditory parameters. Some parameters can be easily mapped, and some appear to be unique to their sensory domain. User studies will evaluate current mappings and discover other relevant mappings. It is expected that individuals may prefer different mappings based on their experience and context.

## 5. DISCUSSION

When composing multimedia artwork, the various media are often sufficiently dissimilar that it is hard for an individual to have expertiese in all relevant areas. A musician may not have the graphics programming ability to build a rendering engine, and an animator may not have the musical expertise to generate a complementary musical score. This system would allow experts in one medium to explore the range of possibilities in another medium, using the language and context of their area of expertise. In this way, each individual uses the control device with which they are most comfortable, and is offered the opportunity to control and develop artistic expression outside of their normal frame of experience, without having to learn techniques and fingerings. A musician can compose a piece and map the parameters to a visual space, generating relevant and connected artwork to complement the piece, or to produce a conversational foundation for further collaborative development with a visual artist. Artists who are expert in both graphics and acoustics may find a new freedom of expression by composing one exploration and then examining the corresponding exploration in the other media, *without* combining them.

The concept we are describing is quite variable in the amount of processing required. Because of this, both real-time interactive prototyping systems and full-quality rendering engines can be developed which will respond to the same set of parameters. Composers and artists can then collaborate on the initial visualization/sonification in real-time, and the resulting script of parameter manipulations can be fed to a rendering engine to generate a high-quality piece of algorist artwork. This is another advantage of using the *cogito* system, since the artists and musicians are not required to understand, at a syntactic level, the parameters used or the parameter space available to them – they can explore the space interactively using exemplars.

The extraction of parameters from animation to sound is done in a way analogous to plug-in visualizers: information is extracted from the generated frame in the animation and used to determine acoustic parameters. the approach described in this paper allows various levels of detail, because a near real-time display can be generated from very little data. Additionally, one can go back and add subtle nuances, based on increasing amounts of data as both the visual and acoustic parts are rendered with more detail.

The parameters can be specified in many ways. The method presented above describes the use of parameter paths, calculated using a set of user controls. Another specification method is to build a set of key states and interpolate the parameter curves between these key states.

## 6. CONCLUSIONS AND FUTURE WORK

Through the development of the sonification of fractal transitions, we have shown the feasability of creating a correspondence between visual and acoustic parameter spaces. The two types of parameter linkages, mapping from user input and extracting from a generated medium, provide a synchronization and a perceptual connection between the resulting media which is impossible to obtain using current unidirectional visualization or sonification techniques.

To turn this technique into a true musical interface, we will allow users to play a musical instrument to drive the parameter changes. The real-time interactive system will be developed to use MIDI messages to define transitions and modifications of the media objects. The MIDI control device (keyboard, guitar, wind instrument) will manipulate the direct parameters, and the indirect parameters will be found using the extraction techniques described above.

## 7. ACKNOWLEDGEMENTS

The authors wish to acknowledge the Natural Sciences and Engineering Research Council of Canada Discovery Grant and Undergraduate Student Research Award programs.

## 8. REFERENCES

[1] B. Burch and J. C. Hart. Linear fractal shape interpolation. In W. A. Davis, M. Mantei, and R. V. Klassen, editors, *Graphics Interface '97*, pages 155–162. Canadian Human-Computer Communications Society, 1997.

[2] M. Good. MusicXML for notation and analysis. In W. Hewlett and E. Selfridge-Field, editors, *The Virtual Score*, pages 113–124. MIT Press, Cambridge, 2001.

[3] D. H. Hepting. *A New Paradigm for Exploration in Computer-Aided Visualization*. PhD thesis, Simon Fraser University, 1999. Ph.D. Dissertation.

[4] D. H. Hepting. Interactive evolution for systematic exploration of a parameter space. In C. H. Dagli, editor, *Proceedings of ANNIE 2003*, page to appear. American Society of Mechanical Engineers, 2003.

[5] S. R. Holtzman. *Digital Mantras*. MIT Press, 1994.

[6] J. W. John G. Neuhoff, Gregory Kramer. Sonification and the interaction of perceptual dimensions. In *ICAD*, 2000.

[7] W. Lytle. *Computer Animated Music*. Online: [http://www.animusic.com/software.html], Retrieved January 29,2004.

[8] L. Manovich. *The Language of New Media*. MIT Press, 2002.

[9] A. R. Smith. Plants, fractals, and formal languages. *Computer Graphics*, 18(3):1–10, 1984.

[10] R. Storms. Auditory-visual cross-modal perception. In *ICAD*, 2000.