Painting as an Interface for Timbre Design

Michael Bylstra[1], Haruhiro Katayose[2]

[1] Kwansei Gakuin University
Gakuen, Sanda, 669-1337 JAPAN
mbylstra@digitalfeast.com.au

[2] Kwansei Gakuin University
Gakuen, Sanda, 669-1337 JAPAN
http://ist.ksc.kwansei.ac.jp/~katayose/

Abstract. There is a challenge in designing a system for timbre design that is engaging for new users and enables experienced users to intuitively design a diverse range of complex timbres. This paper discusses some of the issues involved in achieving these aims and proposes that a timbre can be intuitively represented as an image. The design of TimbrePainter, a system that uses images painted with a mouse to specify the parameters of a harmonic additive synthesizer, is described.

1 Introduction

For a timbre design system to appeal to a wide range of musicians, it must be engaging for new users and enable experienced users to intuitively design a diverse range of complex timbres[1]. Synthesizer timbre design has generally been considered a special technical skill and new users have been encouraged to select from a set of predefined timbres rather than design their own.

A synthesizer can be evaluated in terms of power and usability. Synthesis methods exist that offer an extremely high level of power in that almost any sound possible can be created. However, interfaces for these synthesis methods usually suffer from a very low level of usability. On the other hand, synthesizers that employ simple synthesis methods and simple interfaces often do not offer much potential for exploration. There is a challenge in designing a flexible system that is both powerful and easy-to-use.

Another factor that should be considered is the degree in which a timbre can be predicted by observing the current state of the system. J.O.Smith III has referred to this concept as predictability [1]. The more predictable a system is, the more easily users can find desired timbres. In order for systems to be predictable for new users, predictability must be achievable without the need of any knowledge of the particular synthesis method employed. Predictability is likely to be an important factor in the design of a synthesizer that is both powerful and easy-to-use.

[1] Timbre design is commonly referred to as synthesizer “patch editing” or “programming”.

Although synthesizer interface design in the commercial sector has been rather stagnant, there have been a number of promising developments in the academic community in recent years. Notable developments include Tristimulus Synthesis [2], Interactive Evolution [3] and Scanned Synthesis [4]. In each of these examples usability is paramount but substantial amounts of power and predictability are also achievable to varying degrees.

With similar goals in mind, we have attempted to design a system that is strong in usability, power and predictability. We will present TimbrePainter, a system that demonstrates a method of timbre design which uses painting as an interface for a harmonic additive synthesizer. The following section is a review of synthesizer interface research that is relevant to the design of our system. Section 3 describes the design of TimbrePainter. Section 4 discusses and evaluates the design of TimbrePainter and puts forward some ideas for improvement and future research.

2 A Review of Previous Research

2.1 Recent Developments in the Performance Control of Timbre

Timbre can be designed in a non-linear editing fashion, where synthesis parameters are manipulated until a desired timbre has been achieved, or it can be modified during a performance. Seago refers to the former as ‘fixed synthesis’ and the latter as ‘real-time synthesis’[5]. The performance control of timbre is an important area of research as the manipulation of timbre is crucial for the expressive control of a musical instrument. However, the focus of this paper is on issues solely related to timbre design rather than expressive performance. Nonetheless, there have been many recent developments in the area of ‘real-time synthesis’ that are relevant to this study.

The Importance of Mapping. Hunt found that mapping strategies which are not one-to-one can be more engaging to users than one-to-one mappings [6]. Synthesizers have generally employed a one-to-one mapping, such as slider to filter cutoff frequency, whereas acoustic instruments, generally considered far more intuitive and expressive, often have a complex and non-linear relationship between the control surface and the sound generation mechanism.

Abstracting Large Parameter Spaces. Synthesizers capable of producing a wide range of complex sounds generally require a large number of parameters to be controlled. Mulder proposes two ways of reducing this problem [7]. The number of parameters needed to be directly controlled can be minimised by controlling an intermediate virtual interface that abstracts the synthesis interface. The number of parameters capable of being controlled simultaneously can be maximised by exploiting the human motor system’s ability to effortlessly control many degrees of freedom.

Scanned Synthesis. Scanned Synthesis is a recent, innovative approach to sound synthesis [7]. A dynamic Wavetable is determined by the state of an arbitrary, slowly moving physical system which may be real or virtual. Although any interface can be used, this technique is notable because there is a high degree of amplification between input and output complexity. This means that simple, low parameter interfaces can yield highly complex and interesting results, making for instantly accessible and engaging instruments for new users. However, as the synthesis technique operates in the time-domain, the results are very unpredictable. As a system for ‘fixed synthesis’ it may be unsuitable as experienced users may feel a lack of control and personal authorship of the sounds they have made.

Tristimulus Synthesis. The Tristimulus Synthesizer [2] is a very recent and innovative synthesizer design inspired by the Tristimulus model of timbre perception proposed by Pollard and Jansson [8]. Similarly to the way that just three types of colour receptors are necessary to generate sensations of a wide range of colours, the Tristimulus model proposes that the relative loudness of three bands of partials plays a large part in the sensation of timbre. Using just three parameters that control the loudness of three bands of partials, a wide range of timbres can be produced. Controls are mapped to an abstract layer based on the perception of timbre that is entirely independent of the synthesis method used. Although there are more powerful synthesis methods available, the Tristimulus Synthesizer is currently perhaps the best example of a synthesizer that is all at once strong in usability, power and predictability.

Tristimulus synthesis is successful as a ‘real-time synthesis’ technique and it is also promising as a fixed synthesis technique. However, although Tristimulus parameters may play a large part in the perception of timbre, they cannot describe all of the perceptual features of timbre. Research into other perceptual features of timbre could lead to a very powerful, easy-to-use and predictable method of ‘fixed-synthesis’.

2.2 Interfaces for ‘Fixed-Synthesis’

Conventional synthesizer interfaces for fixed synthesis have consisted of either an array of one dimensional controls (usually knobs or sliders) that control one parameter each or a single control that can be used to control any parameter. In the second case, a menu displayed on a small LCD screen must be navigated in order to select the current parameter to be edited. It has been widely acknowledged that these conventional interfaces suffer from poor usability. In an effort to find the root of the problem and propose a solution, Seago has conducted usability tests and has recommended that controls should map to perceptual space rather than the parameters of a particular synthesis method [5].

Additive synthesis is perhaps the most powerful synthesis method possible. According to Fourier Theory, any possible signal can be represented as a summation of sine functions. However, there is little or no amplification between input complexity and output complexity. An enormous number of parameters and a large amount of computational resources are required to generate interesting sounds. However, in recent years a promising application of additive synthesis has been Analysis-Resynthesis [9, 10].

Recordings of real instruments (sources of complex timbre) are analysed and data at a high level of abstraction such as the envelope of each partial is extracted. This data is used to determine the parameters of an additive synthesizer. This technique offers the potential for a much higher level of timbral control than with the commercially popular sample-playback synthesis technique. Ircam's AudioSculpt enables the user to study an analysed sound's spectrum with a visual representation [11]. The image and resulting sound can be edited by selecting an area with the mouse and applying a transformation. Analysis-Resynthesis makes it possible to realistically emulate musical instruments, but the possibility for exploring new timbres is usually limited to morphing between known timbres[2].

Evolutionary algorithms have been used to search large parameter spaces. Dahlstedt's systems use Interactive Evolution, where user evaluation of automatically generated timbres guides the evolution of a timbre, to search for new and interesting sounds [3]. This interface is very easy-to-use, as deciding whether or not you like a timbre is an extremely intuitive act. Interactive Evolution makes it easy for users with no experience in timbre design to discover sounds that are customized to their own personal preference. However, there are difficulties in steering the search to find a desired timbre and a large number of user evaluations is often necessary.

2.3 Painting as an Interface for Synthesizer Parameter Setting

Synthesizer designers have for a long time realised the potential of designing sound using drawing or painting as an interface. In the 1930s, some composers experimented with drawing sound waves directly onto film. Although some interesting results were possible, painting sound waves in the time-domain was a very laborious process and the results were extremely unpredictable. This, Hunt writes, is because, "human beings are not naturally equipped with the knowledge of how to draw sound" [12]. This seems to be true for sound representations in the time-domain but may not necessarily be true for other sound representations. For painting to be effective, it should map to a higher level of abstraction than time-domain data.

Daphne Oram created the 'Oramics' machine in the 1960s where hand drawn lines were used to define the envelopes of subtractive synthesis parameters such as oscillator frequency or filter cutoff frequency [13]. With this higher level of abstraction, results could be obtained much more quickly, but only with a good knowledge of the synthesis method could the user predict the output.

2.4 Painting as an Interface for Additive Synthesis

A number of systems have implemented the idea of using a painting interface to control an additive synthesizer. The huge number of parameters required makes it impractical to use conventional interfaces such as knobs and sliders or a push-button LCD interface. Painting allows many parameters to be manipulated with just one brush stroke.

---

[2] It should be noted that the painting interface described in this paper could be used to manipulate analysed timbres.

The Pattern Playback. The Pattern Playback was a machine created in the late 1940s that was capable of transforming a painted spectrogram into a tone consisting of 50 harmonics [14]. The loudness of the oscillators was controlled by a row of light sensors that sensed light from a moving, painted spectrogram. It was very useful for speech research, the intended application, but limited as a musical instrument as the fundamental frequency was fixed at 120 Hz.

The vOICe. Maijer has developed a system for the blind that uses video images to generate sound from an additive synthesizer [15]. A painting interface can be used to create custom images and sounds. The y-axis of the image represents pitch, the x-axis represents time and the brightness of each sound pixel represents oscillator amplitude. In this case the frequency of each oscillator is not a harmonic. Instead, the frequencies increase exponentially. What is heard could be described as a cacophony of individual tones rather than a single coherent timbre.

Metasynth. Metasynth [16] is a similar system to the vOICe. In Metasynth the frequencies can be set to correspond to a musical scale such as the diatonic scale or a microtonal scale. By using musical scales the boundary between sound design and composition is blurred. Although it is possible to create interesting, unconventional music and sounds with this software, harmonic timbres, like the timbres of most conventional instruments, cannot be created.

Yellowtail. Golan Levin's Yellowtail is an innovative system that uses interactive animation, rather than still images, to control an additive synthesizer [17]. Animations can be created on-the-fly using painterly mouse gestures. This leads to interesting and constantly evolving sounds. Like the vOICe and Metasynth, only unconventional, inharmonic sounds can be produced.

2.5 The Potential of Painting as an Interface to Timbre Design.

A computer painting interface is instantly accessible for new users as computer painting programs are very familiar. Unlike actions required by conventional interfaces such as entering numbers into small LCD screen, painting happens to be an enjoyable activity by itself. Painting offers a good combination of ease-of-use and power in that paintings can be arbitrarily simple (a smiley face) or arbitrarily complex (a detailed depiction of a piano timbre).

Although the relationship between image and sound is counter-intuitive with time-domain representations there is a much more natural relationship with frequency-domain representations or spectrograms. Helmholtz has likened the cochlea to a spectral analyzer which resonates at specific locations along the basilar membrane [18]. There is a direct mapping between additive synthesis parameters and this stage of the human auditory perception system. Rodet and Depalle write that there is a simple mapping of frequency and amplitude parameters into the human perceptual space and that these parameters are meaningful to musicians [9].

In theory, with a large enough array of sine wave oscillators, a digital additive synthesizer could produce any perceptible sound, including harmonic, inharmonic and noisy sounds [1]. However, an almost infinite number of oscillators would be necessary to produce an arbitrary sound. Due to the high computation costs of additive synthesis, we are limited in the number of oscillators available. We can choose to distribute the frequencies inharmonically (as in The vOICe, Metasynth, or Yellowtail) or distribute them as harmonics of a single fundamental frequency (as in The Pattern Playback). Inharmonic additive synthesis has been explored by several systems and is limited to a subset of inorganic sounds that can only appeal to a particular niche group of musicians and listeners. There is room for exploration in the design of a harmonic additive synthesizer that uses painting as an interface.

3 TimbrePainter

TimbrePainter is a system that features a painting interface for manipulating the parameters of a harmonic additive synthesizer. Painting can be used to paint a new picture or to edit an image captured from a web cam or imported from a file. Its intended application is to enable new users to explore timbre design through painting. While designing sounds, melodies generated from a step-sequencer give constant feedback. Alternatively, a MIDI controller can be used to trigger notes. A simple, self-explanatory GUI with few elements has been employed to make it easy for new users to instantly begin creating sounds. A video demonstration of the system can be viewed at *http://ist.ksc.kwansei.ac.jp/~katayose/TimbrePainter/*.

3.1 Implementation and Features

The parameters of a harmonic additive synthesizer can be described as an array of amplitude envelopes - one for each harmonic or sine wave oscillator [20]. The frequency of the fundamental is arbitrary and the frequency of each harmonic is determined by the equation:

*frequency of harmonic = fundamental frequency * (harmonic number + 1)*

This is best visualized as a set of curves in 3D space as in Fig.1

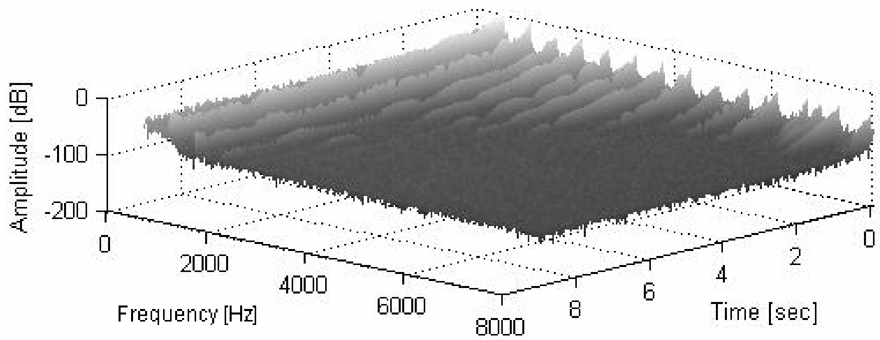

Fig. 1. A 3D representation of a piano timbre

This 3D space can be represented in a 2D view if the z-axis is represented by colour intensity rather than height (Fig. 2).

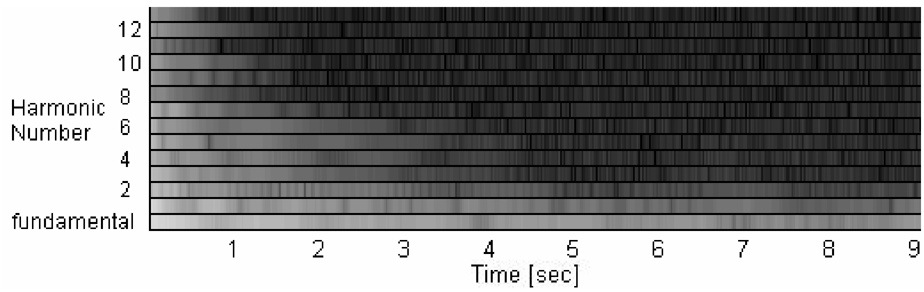

Fig. 2. A 2D representation of additive synthesis parameters

This representation is essentially a 2D image with a single colour channel. If additive synthesizer parameters can be represented as an image, then the reverse is also possible - an image can be used to specify additive synthesizer parameters.

TimbrePainter features a painting area or 'canvas' for creating and editing images. Fig.4. shows an example of an image created in TimbrePainter. White represents the highest amplitudes and black represents zero amplitude. The first row represents the fundamental. Its frequency is determined by midi input or the step-sequencer.

The GUI of TimbrePainter (Fig.5) is divided into four sections: ‘Timbre Canvas’, ‘Brush Options’, ‘Step Sequencer’ and ‘Spectral Filter.’ Drawing or painting is used throughout the system.

The ‘Timbre Canvas’ is used for painting timbres with the mouse. If a USB WebCam is connected, clicking on the ‘Take Photo’ button captures the current frame from the web cam. A JPEG or TIFF file can be imported by clicking the ‘Import Image’ button. WebCam photos or imported images can be further edited with the paintbrush. Images can be saved as a JPEG or TIFF file.

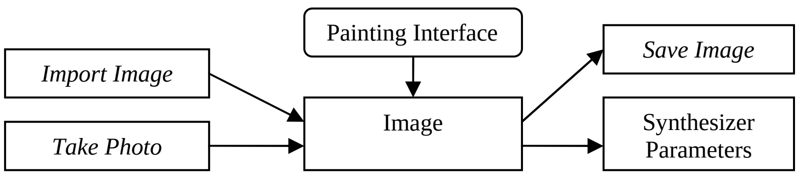

Fig. 3. Flow diagram of image creation and editing methods

With the ‘Step Sequencer’ users can draw a looping melody which is used for feedback while designing a sound. In the ‘Spectral Filter’ section users can draw an arbitrary spectral filter which can be used to control timbre during performance. In the ‘Brush Options’ section users can choose the colour intensity of the paintbrush, change the brush size and select the brush mode. ‘Burn’ mode gradually increases the intensity at the location of the mouse, and ‘Dodge’ gradually decreases it, allowing for smoother gradients.

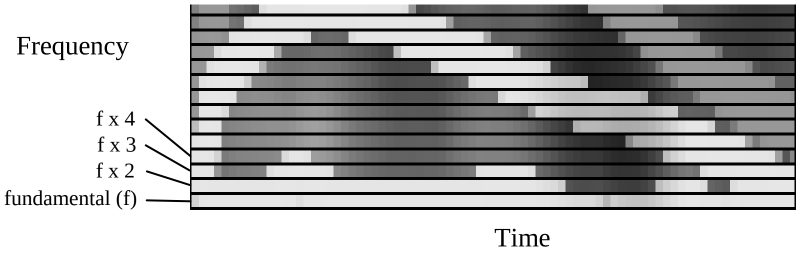

Fig. 4. A description of an image created in TimbrePainter

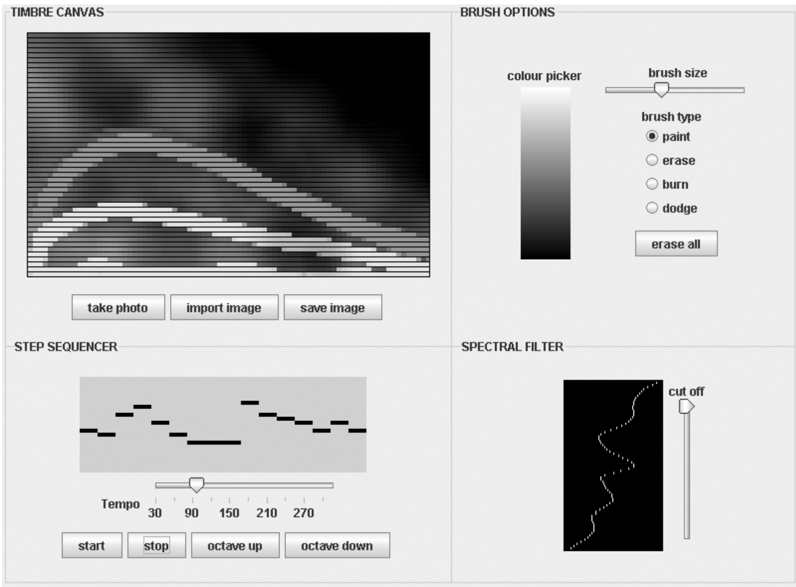

Fig. 5. A screenshot of the GUI used in TimbrePainter

3.2 The Mappings of the Painting Interface

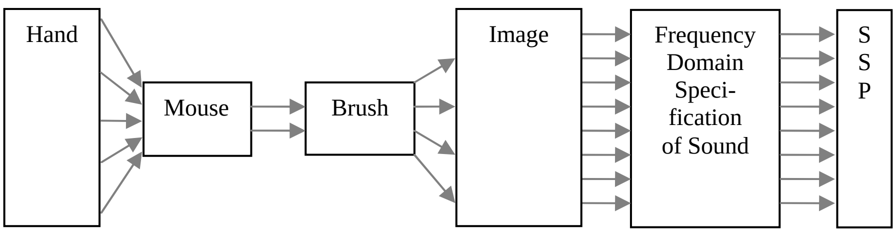

Fig.6. Diagram of hand movement to sound synthesis parameter mapping

TimbrePainter can be viewed as a system with several mapping layers. One hand, which has many degrees of freedom, is used to control two mouse parameters (X and Y position) which simply map to the on-screen location of the paintbrush. The mapping between the mouse and the image can be seen as a two-to-many mapping where the number of image parameters simultaneously controlled depends on the size of the brush. The image is mapped to frequency-domain layer of the perceptual system which is then directly mapped to additive synthesis parameters. The way the properties of the brush influence the mapping between mouse and image is where this interface derives its flexibility. By changing the size of the brush users can choose to control the interaction level from macroscopic (coarse control of timbre) to microscopic (precise control of timbre) [6].

4 Discussion

So far the testing of the system has been limited to observing user behaviour and collecting verbal feedback from a selection of users ranging from synthesizer enthusiasts to non-musicians. Most users found the system to be engaging and very easy-to-use. They were able to generate a reasonably wide range of interesting timbres which were predictable to a certain degree. However, there is much room for improvement to the system and several common comments and observations have been very useful in thinking of improvements.

4.1 Limitations of Using a Mouse as a Controller

The most noticeable short-coming of TimbrePainter was in the lack of complexity or quality of the sounds created by the users. Although a reasonably wide range of interesting timbres were created, they all belonged to a certain class particular to this system. Although in theory, using the current system, it should be possible to create harmonic sounds of an unlimited degree of complexity, such as a piano sound, the results rarely achieved this aim.

Natural timbres often have complex amplitude variations and this is difficult to achieve with TimbrePainter's paint brush, as it can only paint flat areas of colour. The use of Ryokai's "I/O Brush" would be a fun and engaging way to create interesting paint brush textures [21]. The mouse is limited as a device for painting as it can only specify x or y coordinates. During painting it is not capable of specifying a z coordinate which could be used to specify colour intensity (amplitude). Relative sound levels of harmonics play a large part in the perceived timbre of a sound and even subtle changes in sound level can be very noticeable. Therefore, it is very important to have very accurate control of colour intensity.

A digitizing tablet would be preferable to use as a controller for TimbrePainter as the pen pressure could be used to specify colour intensity. However, although a high level of accuracy can be obtained in the x and y coordinates, only a relatively small level of accuracy can be obtained in the z coordinate (pen pressure). An alternative idea, the use of a clay model, will be presented in section 4.3.

4.2 Limitations of the Perceptual Model used

The current system operates on the frequency-domain level of abstraction of timbre. Each harmonic is treated as equally important and each time slice is treated as equally important. However, psycho-acoustics research has found that certain bands of harmonics and certain time stages are more important than others in the perception of timbre.

High level perceptual features include the relative levels of the Tristimulus bands, inharmonicity, roughness and others. The user should be able to begin by designing in a high level of abstraction and then work on finer detail in the frequency domain. High level perceptual features should draw from psycho-acoustics research. An idea could be to increase the size of the pixels in the regions of more perceptual importance so the user has more accuracy in these areas. Verbal descriptions such as 'bright' describe timbre at a very high level of abstraction and could be useful for performance control.

Time domain	Frequency Domain	High level perceptual features	Verbal description

Fig. 7. Levels of abstraction in describing timbre

4.3 Using Live Video to specify Additive Synthesis Parameters

Live video of light reflected from a shallow pool of water has been used by Dannenberg to specify the parameters of an additive synthesizer [20]. Three vertical lines correspond to three voices and the parameters of one voice are controlled by a single vertical strip of video data. One of the limitations of the current implementation of TimbrePainter is that timbre is the same for every pitch. By assigning different pitches to different vertical lines, timbre would change dramatically depending on pitch and interesting results may be achieved.

4.4 Using Modeling Clay as an Interface

As 3D coordinates can be used to describe additive synthesis parameters, a very intuitive way to specify the parameters is to sculpt a real 3D model using your hands. As the amplitude envelopes of partials can be best visualized as a 3D model, the visual feedback is more intuitive. Inexpensive modeling clay such as PlayDoh is ideal. Using modeling clay, models can be quickly made and reworked. This is a both fun and a powerful method for timbre design. In line with Mulder's aims, the hands can be exploited to their full potential of gestural control [7]. There is a many-to-many mapping between the hands and the clay model and the bottle-neck in the hand-to-mouse-to-brush-to-image mapping is removed.

We have experimented with digitizing model information by simply using a web cam and mapping the pixel data to additive synthesis parameters using the 'take photo' feature. However, no meaningful height information from the model is retained, so outcomes are largely unpredictable. For this interface to be effective, depth perception techniques would need to be used to gather height information from the model. A commercial 3D scanner or computer vision techniques used to analyse images from a stereoscopic pair of digital cameras could be used.

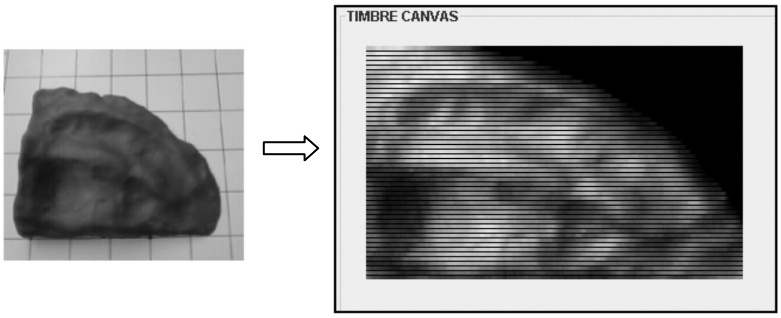

Fig. 8. A photo of model converted to image parameters

Mulder's approach of using a Dataglove to sculpt a virtual clay model could also be used [7]. This approach would have the advantage that the process of extracting height information from the model would be relatively much simpler and models could be saved and recalled. However, this would be at the expense of a severe reduction in the quality of visual feedback and a total loss of tactile feedback.

Conclusion

This paper has dealt with issues related to the challenge of designing a system for timbre design which is easy-to-use, powerful and predictable. We have attempted to meet these aims in our design of TimbrePainter, a system which uses painting as an interface for additive synthesis. Although this system offers a good trade-off between these three aims, there is room for improvement in each of these areas. As conventional synthesizer interfaces have been deemed inadequate for controlling complex synthesizers, there is a need to find alternative interfaces, such as painting, that make better use of the full potential of human motor control. 3D sculpting has been proposed as another alternative interface with a good potential. Of equal importance is establishing a meaningful way of describing timbre that does not depend on any particular synthesis method. There is a need for further research into the nature of timbre perception and this research should be used to guide the design of future systems.

References

1. Smith, J.O.: Viewpoints on the History of Digital Synthesis. Proceedings of the International Computer Conference. San Francisco (1991)

2. A. Riley, D. Howard.: A Real-Time Tristimulus Synthesizer. http://www-users.york.ac.uk/~dmh8/tristimulus.htm (2004)
3. Dahlstedt, P.: Creating and Exploring Huge Parameter Spaces: Interactive Evolution as a Tool for Sound Generation. Proceedings of the International Computer Music Conference. Habana, Cuba (2001)
4. Verplank, B., Mathews, M., Shaw, R.: Scanned Synthesis. Proceedings of the 2000 International Computer Music Conference. Zannos editor, ICMA, Berlin (2000) 368-371
5. Seago, A., Holland, S.,Mulholland, P.: A Critical Analysis of Synthesizer User Interfaces for Timbre. HCI 2004: Design for Life. British HCI Group, Leeds (2004)
6. Hunt A., Wanderley M., Kirk R.: Towards a Model for Instrumental Mapping in Expert Musical Interaction. Proceedings of the 2000 International Computer Music Conference. Berlin, Germany (2000)
7. Mulder, A., Fels, S., Mase, K.: Mapping virtual object manipulation to sound variation. IPSJ SIG notes Vol. 97, No. 122, 97-MUS-23 (USA/Japan intercollege computer music festival) (1997) 63-68.
8. Pollard, H.F., Jansson, E.V.: A tristimulus method for the specification of musical timbre. Acustica 51 (1982) 162-171.
9. Rodet, X., Depalle, P.: Spectral Envelopes and Inverse FFT Synthesis. Proc. 93rd AES Conv, San Francisco (1992)
10. Serra, X.: “Musical Sound Modeling with Sinusoids plus Noise”. Musical Signal Processing, Swets & Zeitlinger Publishers (1997)
11. Bogaards, N., Röbel, A., Rodet,X.: Sound Analysis and Processing with AudioSculpt 2, Proc. Int. Computer Music Conference (ICMC'04). Miami (2004)
12. Hunt, A: Radical User Interfaces for Real-time Musical Control. DPhil thesis. University of York, UK (1999)
13. Daphne Oram and 'Oramics' (1959), http://www.obsolete.com/120_years/machines/oramics/
14. Rubin,P., Goldstein,L.: The Pattern Playback. http://www.haskins.yale.edu/Haskins/MISC/PP/pp.html
15. Meijer, P.: An Experimental System for Auditory Image Representations. IEEE Transactions on Biomedical Engineering (1992) 112-121
16. UI Software, Metasynth, 1998.
17. Levin, G.: Painterly Interfaces for Audiovisual Performance. (2000) Chapter 3, Section 2.1.
18. Hong,T.: Salient feature extraction of musical instrument signals. M.A. Thesis. Dartmouth College (2000)
19. Dannenberg, Neuendorffer: Sound Synthesis from Real-Time Video Images. Proceedings of the 2003 International Computer Music Conference. International Computer Music Association, San Francisco (2003) 385-388
20. Roads, C.: The Computer Music Tutorial. MIT Press, Cambridge (1996).
21. Ryokai, K., Marti, S., Ishii, H.: Designing the World as Your Palette. Proceedings of the Conference on Human Factors in Computing Systems (CHI '05). Portland (2005)