

Real-Time Gesture-Controlled Physical Modelling Music Synthesis with Tactile Feedback

David M. Howard

*Media Engineering Research Group, Department of Electronics, University of York, Heslington, York, YO10 5DD, UK
Email: dh@ohm.york.ac.uk*

Stuart Rimell

Media Engineering Research Group, Department of Electronics, University of York, Heslington, York, YO10 5DD, UK

Received 30 June 2003; Revised 13 November 2003

Electronic sound synthesis continues to offer huge potential possibilities for the creation of new musical instruments. The traditional approach is, however, seriously limited in that it incorporates only auditory feedback and it will typically make use of a sound synthesis model (e.g., additive, subtractive, wavetable, and sampling) that is inherently limited and very often nonintuitive to the musician. In a direct attempt to challenge these issues, this paper describes a system that provides tactile as well as acoustic feedback, with real-time synthesis that invokes a more intuitive response from players since it is based upon mass-spring physical modelling. Virtual instruments are set up via a graphical user interface in terms of the physical properties of basic well-understood sounding objects such as strings, membranes, and solids. These can be interconnected to form complex integrated structures. Acoustic excitation can be applied at any point mass via virtual bowing, plucking, striking, specified waveform, or from any external sound source. Virtual microphones can be placed at any point masses to deliver the acoustic output. These aspects of the instrument are described along with the nature of the resulting acoustic output.

Keywords and phrases: physical modelling, music synthesis, haptic interface, force feedback, gestural control.

1. INTRODUCTION

Musicians are always searching for new sounds and new ways of producing sounds in their compositions and performances. The availability of modern computer systems has enabled considerable processing power to be made available on the desktop and such machines have the capability of enabling sound synthesis techniques to be employed in real-time, that would have required large dedicated computer systems just a few decades ago. Despite the increased incorporation of computer technology in electronic musical instruments, the search is still on for virtual instruments that are closer in terms of how they are played to their physical acoustic counterparts.

The system described in this paper aims to integrate music synthesis by physical modelling with novel control interfaces for real-time use in composition and live performances. Traditionally, sound synthesis has relied on techniques involving oscillators, wavetables, filters, time envelope shapers, and digital sampling of natural sounds (e.g., [1]). More recently, physical models of musical instruments have been used to generate sounds which have more natural qualities and have control parameters which are less abstract and

more closely related to musicians' experiences with acoustic instruments [2, 3, 4, 5]. Professional electroacoustic musicians require control over all aspects of the sounds with which they are working, in much the same way as a conductor is in control of the sound produced by an orchestra. Such control is not usually available from traditional synthesis techniques, since user adjustment of available synthesis parameters rarely leads to obviously predictable acoustic results. Physical modelling, on the other hand, offers the potential of more intuitive control, because the underlying technique is related directly to the physical vibrating properties of objects, such as strings and membranes with which the user can interact through inference relating to expectation.

The acoustic output from traditional electronic musical instruments is often described as "cold" or "lifeless" by players and audience alike. Indeed, many report that such sounds become less interesting with extended exposure. The acoustic output from acoustic musical instruments, on the other hand, is often described as "warm," "intimate" or "organic." The application of physical modelling for sound synthesis produces output sounds that resemble much more closely their physical counterparts.

The success of a user interface for an electronic musical instrument might be judged on its ability to enable the user to experience the illusion of directly manipulating objects, and one approach might be the use of virtual reality interfaces. However, this is not necessarily the best way to achieve such a goal in the context of a musical instrument, since a performing musician needs to be actively in touch visually and acoustically not only with other players, but also with the audience. This is summed up by Shneiderman [6]: “virtual reality is a lively new direction for those who seek the immersion experience, where they block out the real world by having goggles on their heads.” In any case, traditionally trained musicians rely less on visual feedback with their instrument and more on tactile and sonic feedback as they become increasingly accustomed to playing it. For example, Hunt and Kirk [7] note that “observation of competent pianists will quickly reveal that they do not need to look at their fingers, let alone any annotation (e.g., sticky labels with the names of the notes on) which beginners commonly use. Graphics are a useful way of presenting information (especially to beginners), but are not the primary channel which humans use when fully accustomed to a system.”

There is evidence to suggest that the limited information available from the conventional screen and mouse interface is certainly limiting and potentially detrimental for creating electroacoustic music. Buxton [8] suggests that the visual senses are overstimulated, whilst the others are under-stimulated. In particular, he suggests that tactile input devices also provide output to enable the user to relate to the system as an object rather than an abstract system, “every haptic input device can also be considered to provide output. This would be through the tactile or kinaesthetic feedback that it provides to the user Some devices actually provide force feedback, as with some special joysticks.” Fitzmaurice [9] proposes “*graspable user interfaces*” as real objects which can be held and manipulated, positioned, and conjoined in order to make interfaces which are more akin to the way a human interacts with the real world. It has further been noted that the haptic senses provide the second most important means (after the audio output) by which users observe and interact with the behaviour of musical instruments [10], and that complex and realistic musical expression can only result when both tactile (vibrational and textural) and proprioceptive cues are available in combination with aural feedback [11].

Considerable activity exists on capturing human gesture <http://www.media.mit.edu/hyperins/> and <http://www.megaproject.org/> [12]. Specific to the control of musical instruments is the provision of tactile feedback [13], electronic keyboards that have a *feel* close to a real piano [14], haptic feedback bows that simulate the feel and forces of real bows [15], and the use of finger-fitted vibrational devices in open air gestural musical instruments [16]. Such haptic control devices are generally one-off, relatively expensive, and designed to operate linked with specific computer systems, and as such, they are essentially inaccessible to the musical masses. A key feature of our instrument is its potential for wide applicability, and therefore inexpensive and widely

available PC force feedback gaming devices are employed to provide its real-time gestural control and haptic feedback.

The instrument described in this paper, known as Cymatic [17], took its inspiration from the fact that traditional acoustic instruments are controlled by direct physical gesture, whilst providing both aural and tactile feedback. Cymatic has been designed to provide players with an immersive, easy to understand, as well as tactile musical experience that is more commonly associated with acoustic instruments but rarely found with computer-based instruments. The audio output from Cymatic is derived from a physical modelling synthesis engine which has its origins in TAO [3]. It shares some common approaches with other physical modelling sound synthesis environments such as Mosaic in [4] and Cordis-Anima in [5]. Cymatic makes use of the more intuitive approach to sound synthesis offered by physical modelling, to provide a building block approach to the creation of virtual instruments, based on elemental structures in one (string), two (sheet), three (block), or more dimensions that can be interconnected to form complex virtual acoustically resonant structures. Such instruments can be excited acoustically, controlled in real-time via gestural devices that incorporate force feedback to provide a tactile response in addition to the acoustic output, and heard after placing one or more virtual microphones at user-specified positions within the instrument.

2. DESIGNING AND PLAYING CYMATIC INSTRUMENTS

Cymatic is a physical modelling synthesis system that makes use of a mass-spring paradigm with which it synthesises resonating structures in real-time. It is implemented on a Windows-based PC machine in C++, and it incorporates support for standard force feedback PC gaming controllers to provide gestural control and tactile feedback. Acoustic output is realised via a sound card that provides support for ASIO audio drivers. Operation of Cymatic is a two-stage process: (1) virtual instrument design and (2) real-time sound synthesis.

Virtual instrument design is accomplished via a graphical interface, with which individual building block resonating elements including strings, sheets, and solids can be incorporated in the instrument and interconnected on a user-specified mass to mass basis. The ends of strings and edges of sheets and blocks can be locked as desired. The tension and mass parameters of the masses and springs within each building block element can be user defined in value and either left fixed or placed under dynamical control using a gestural controller during synthesis. Virtual instruments can be customised in shape to enable arbitrary structures to be realised by deleting or locking any of the individual masses. Each building block resonating element will behave as a vibrating structure. The individual axial resonant frequencies will be determined by the number of masses along the given axis, the sampling rate, and the specified mass and tension values. Standard relationships hold in terms of the relative values of resonant frequency between building blocks,

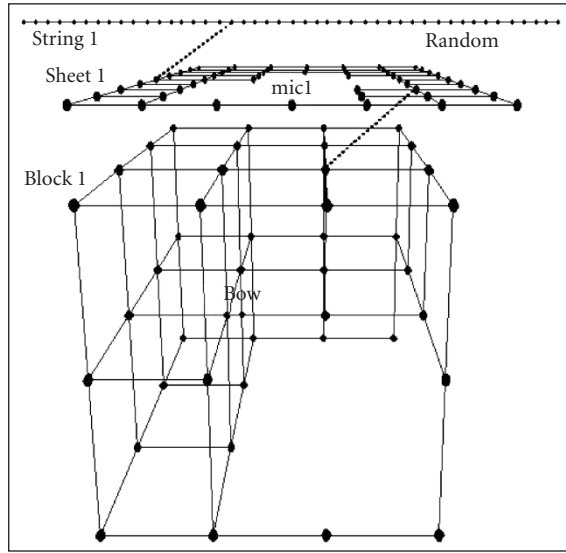


FIGURE 1: Example build-up of a Cymatic virtual instrument starting with a string with 45 masses (top left), then adding a sheet of 7 by 9 masses (bottom left), then a block of 4 by 4 by 3 masses (top right), and finally the completed instrument (bottom right). Mic1: audio output virtual microphone on the sheet at mass (4, 1). Random: random excitation at mass 33 of the string. Bow: bowed excitation at mass (2, 2, 2) of the block. Joins (dotted line) between string mass 18 and sheet mass (1, 5). Join (dotted line) between sheet mass (6, 3) and block mass (3, 2, 1).

for example, a string twice the length of another will have a fundamental frequency that is one octave lower.

An excitation function, selected from the following list, can be placed on any mass within the virtual instrument: pluck, bow, random, sine wave, square wave, triangular wave, or live audio. Parameters relating to the selected excitation, including excitation force and its velocity and time of application where appropriate can be specified by the user. Multiple excitations can be specified on the basis that each is applied to its own individual mass element. Monophonic audio output to the sound card is achieved via a virtual microphone placed on any individual mass within the instrument. Stereophonic output is available either from two individual microphones or from any number of microphones greater than two, where the output from each is panned between the left and right channels as desired. Cymatic supports whatever range of sampling rates that is available on the sound card. For example, when used with an Eridol UA-5 USB audio interface, the following are available: 8 kHz, 9.6 kHz, 11.025 kHz, 12 kHz, 16 kHz, 22.05 kHz, 24 kHz, 32 kHz, 44.1 kHz, 48 kHz, 88.2 kHz, and 96 kHz.

Figure 1 illustrates the process of building up a virtual instrument. The instrument has been built up from a string of 45 masses, a sheet of 7 by 9 masses, and a block of 4 by 4 by 3 masses. There is an interconnection between the string (mass 18 from the left) and the sheet (mass (1, 5) as well as the sheet (mass (6, 3) and the block (mass (3, 2, 1) as indicated by the dotted lines (a simple process based on clicking on the relevant masses). Two excitations have been in-

cluded: a random input to the string at mass 33 and a bowed excitation to the block at mass (2, 2, 2). The basic sheet and block have been edited. Masses have been removed from both the sheet and the block as indicated by the gaps in their structure and the masses on the back surface of the block have all been locked. The audio output is derived from a virtual microphone placed on the sheet at mass (4, 1). These are indicated on the figure as *random*, *bow*, and *mic1*, respectively. Individual components, excitations, and microphones can be added, edited, or deleted as desired.

The instrument is controlled in real-time using a Microsoft Sidewinder Force Feedback Pro Joystick and a Logitech iFeel mouse found on <http://www.immersion.com>. The various gestures that can be captured by these devices can be mapped to any of the parameters that are associated with the physical modelling process on an element-by-element basis. The joystick offers four degrees of freedom (x, y, z-twist movement and a rotary “throttle” controller) and eight buttons. The mouse has two degrees of freedom (X, Y) and three buttons. Cymatic parameters that can be controlled include the mass or tension of any of the basic elements that make up the instrument and the parameters associated with the chosen excitation, such as bowing pressure, excitation force, or excitation velocity. The buttons can be configured to suppress the effect of any of the gestural movements to enable the user to move to a new position while making no change and then the change can be made instantaneously by releasing the button. In this way, step variations can be accommodated.

The force feedback capability of the joystick allows for the provision of tactile feedback with a high degree of customisability. It receives its force instructions via MIDI through the combined MIDI/joystick port on most PC sound cards, and Cymatic outputs the appropriate MIDI messages to control its force feedback devices. The Logitech iFeel mouse is an optical mouse which implements Immersion’s iFeel technology (<http://www.immersion.com>). It contains a vibrotactile device to produce tactile feedback over a range of frequencies and amplitudes via the “*Immersion Touchsense Entertainment*” software, which converts any audio signal to tactile sensations. The force feedback amplitude is controlled by the acoustic amplitude of the signal from a user-specified virtual microphone, which might be involved in the provision of the main acoustic output, or it could solely be responsible for the control of tactile feedback.

3. PHYSICAL MODELLING SYNTHESIS IN CYMATIC

Physical modelling audio synthesis in Cymatic is carried out by solving for the mechanical interaction between the masses and springs that make up the virtual instrument on a sample-by-sample basis. The central difference method of numerical integration is employed as follows:

$$\begin{aligned} x(t + dt) &= x(t) + v\left(t + \frac{dt}{2}\right)dt, \\ v\left(t + \frac{dt}{2}\right) &= v\left(t - \frac{dt}{2}\right) + a(t)dt, \end{aligned} \quad (1)$$

where x = mass position, v = mass velocity, a = mass acceleration, t = time, and dt = sampling interval.

The mass velocity is calculated half a time step ahead of its position, which results in a more stable model than an implementation of the Euler approximation. The acceleration at time t of a cell is calculated by the classical equation

$$a = \frac{F}{m}, \quad (2)$$

where F = the sum of all the forces on the cell and m = cell mass.

Three forces are acting on the cell:

$$F_{\text{total}} = F_{\text{spring}} + F_{\text{damping}} + F_{\text{external}}, \quad (3)$$

where F_{spring} = the force on the cell from springs connected to neighbouring cells, F_{damping} = the frictional damping force on the cell due to the viscosity of the medium, F_{external} = the force on the cell from external excitations.

F_{spring} is calculated by summing the force on the cell from the springs connecting it to its neighbours, calculated via Hooke's law:

$$F_{\text{spring}} = k \sum (p_n - p_0), \quad (4)$$

where k = spring constant, p_n = the position of the n th neighbour, and p_0 = the position of the current cell.

F_{damping} is the frictional force on the cell caused by the viscosity of the medium in which the cell is contained. It is proportional to the cell velocity, where the constant of proportionality is the damping parameter of the cell.

$$F_{\text{damping}} = -\rho v(t), \quad (5)$$

where ρ = the damping parameter of the cell, $v(t)$ = the velocity of the cell at time t .

The acceleration of a particular cell at any instant can be established by combining these forces into (2)

$$a(t) = (1/m) \left(k \sum (p_n - p_0) - \rho v(t) + F_{\text{external}} \right). \quad (6)$$

The position, velocity, and acceleration are calculated once per sampling interval for each cell in the virtual instrument. Any virtual microphones in the instrument output their cell positions to provide an output audio waveform.

4. CYMATIC OUTPUTS

Audio spectrograms provide a representation that enables the detailed nature of the acoustic output from Cymatic to be observed visually. Figure 2 shows a virtual Cymatic instrument consisting of a string and a modified sheet which are joined together between mass 30 (from the left) on the string to mass (6, 3) on the sheet. A random excitation is applied at mass 10 of the string and a virtual microphone (mic1) is located at mass (4, 3) of the sheet. Figure 3 shows the force feedback joystick settings dialog used to control the virtual instrument and it can be seen that the component mass of the string, the component tension, and damping and mass of the sheet are controlled by the X, Y, Z and slider (throttle)

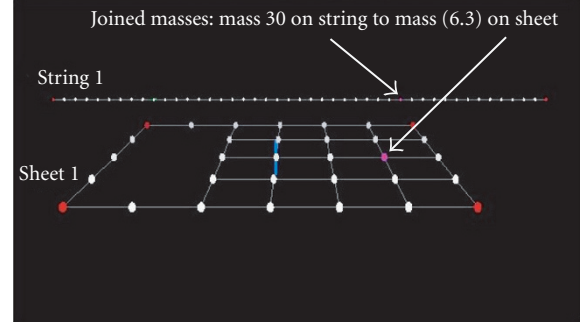


FIGURE 2: Cymatic virtual instrument consisting of a string and modified sheet. They are joined together between mass 30 (from the left) on the string to mass (6, 3) on the sheet. A random excitation is applied at point 10 of the string and the virtual microphone is located at mass (6, 3) of the sheet.

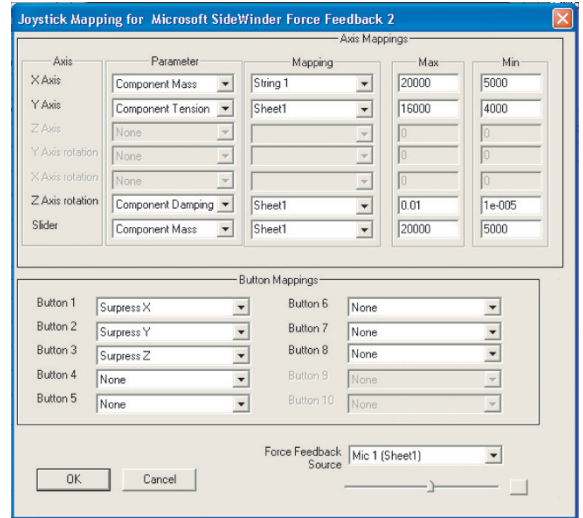


FIGURE 3: Force feedback joystick settings dialog.

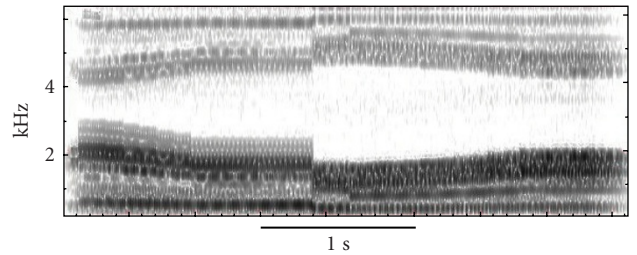


FIGURE 4: Spectrogram of output from the Cymatic virtual instrument, shown in Figure 2, consisting of a string and modified sheet.

functions of the joystick. Three of the buttons have been set to suppress X, Y, and Z; a feature which enables a new setting to be jumped to as desired, for example, by pressing button 1, moving the joystick in the X axis and then releasing button 1. Force feedback is applied based on the output amplitude level from mic1.

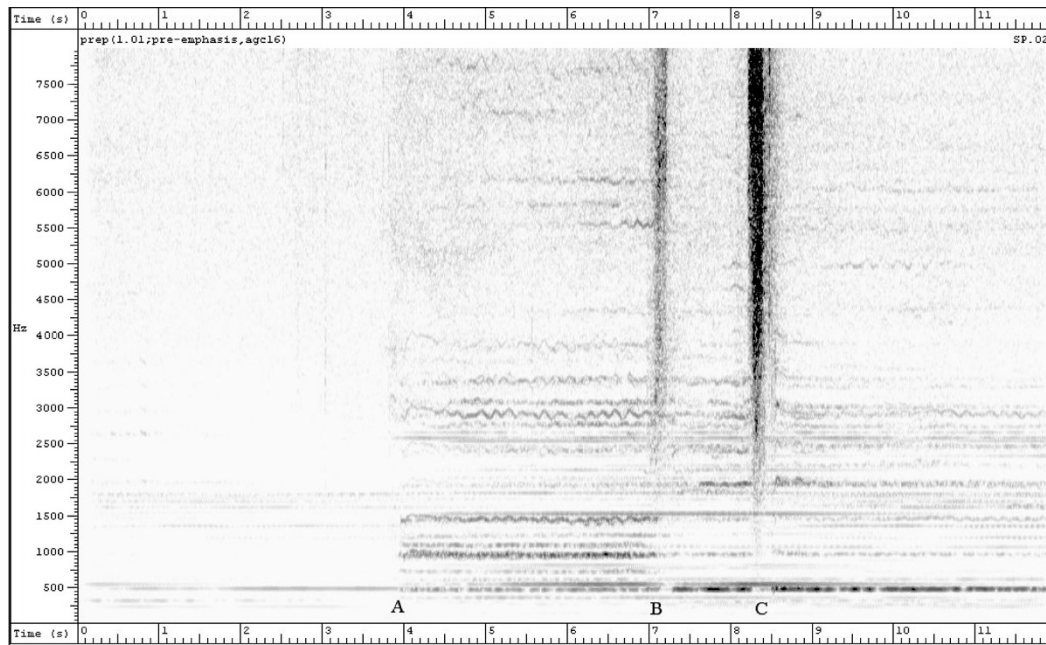


FIGURE 5: Spectrogram of a section of “the child is sleeping” by Stuart Rimell showing Cymatic alone (from the start to A), the word “hush” sung by the four-part choir (A to B) and the “st” of “still” at C.

Figure 4 shows a spectrogram of the output from *mic1* of the instrument. The tonality visible (horizontal banding in the spectrogram) is entirely due to the resonant properties of the string and sheet themselves, since the input excitation is random. Variations in the tonality are rendered through gestural control of the joystick, and the step change notable just before half way through is a result of using one of the “suppress” buttons.

Cymatic was used in a public live concert in December 2002, for which a new piece “*the child is sleeping*” was specially composed by Stuart Rimell for a capella choir and Cymatic (<http://www.users.york.ac.uk/~dmh>). It was performed by the Beningbrough Singers in York, conducted by David Howard. The composer performed the Cymatic part, which made use of three cymbal-like structures controlled by the mouse and joystick. The choir provided a backing in the form of a slow moving carol in four-part harmony, while Cymatic played an obligato solo line. The spectrogram in Figure 5 illustrates this with a section which has Cymatic alone (up to point A), and then the choir enters singing “hush be still,” with the “sh” of “hush” showing at point B and the “st” of “still” at point C. In this particular Cymatic example, the sound colours being used lie at the extremes of the vocal spectral range, but there are clearly tonal elements in the Cymatic output visible. Indeed, these were essential as a means of giving the choir their starting pitches.

5. DISCUSSION AND CONCLUSIONS

An instrument known as Cymatic has been described, which provides its players with an immersive, easy to understand,

as well as tactile musical experience that is rarely found with computer-based instruments, but commonly expected from acoustic musical instruments. The audio output from Cymatic is derived from a physical modelling synthesis engine, which enables virtual instruments with arbitrary shapes to be built up by interconnecting one (string), two (sheet), three (block), or more dimensional basic building blocks. An acoustic excitation chosen from bowing, plucking, striking, or waveform is applied at any mass element, and the output is derived from a virtual microphone placed at any other mass element. Cymatic is controlled via gestural controllers that incorporate force feedback to provide the player with tactile as well as acoustic feedback.

Cymatic has the potential to enable new musical instruments to be explored, that have the potential to produce original and inspiring new timbral palates, since virtual instruments that are not physically realizable can be implemented. In addition, interaction with these instruments can include aspects that cannot be used with their physical counterparts, such as deleting part of the instrument while it is sounding, or changing its physical properties in real-time during performance. The design of the user interface ensures that all of these activities can be carried out in a manner that is more intuitive than with traditional electronic instruments, since it is based on the resonant properties of physical structures. A user can therefore make sense of what she or he is doing through reference to the likely behaviour of strings, sheets, and blocks. Cymatic has the further potential in the future (as processing speed increases further) to move well away from the real physical world, while maintaining the link with this intuition, since the spatial dimensionality of the virtual

instruments can in principle be extended well beyond the three of the physical world.

Cymatic provides the player with an increased sense of immersion, which is particularly useful when developing performance skills since it reinforces the visual and aural feedback cues and helps the player internalise models of the instrument's response to gesture. Tactile feedback also has the potential to prove invaluable in group performance, where traditionally computer instruments have placed an over-reliance on visual feedback, thereby detracting from the player's visual attention which should be directed elsewhere in a group situation, for example, towards a conductor.

ACKNOWLEDGMENTS

The authors acknowledge the support of the Engineering and Physical Sciences Research Council, UK, under Grant number GR/M94137. They also thank the anonymous referees for their helpful and useful comments.

REFERENCES

- [1] M. Russ, *Sound Synthesis and Sampling*, Focal Press, Oxford, UK, 1996.
- [2] J. O. Smith III, "Physical modelling synthesis update," *Computer Music Journal*, vol. 20, no. 2, pp. 44–56, 1996.
- [3] M. D. Pearson and D. M. Howard, "Recent developments with TAO physical modelling system," in *Proc. International Computer Music Conference*, pp. 97–99, Hong Kong, China, August 1996.
- [4] J. D. Morrison and J. M. Adrien, "MOSAIC: A framework for modal synthesis," *Computer Music Journal*, vol. 17, no. 1, pp. 45–56, 1993.
- [5] C. Cadoz, A. Luciani, and J. L. Florens, "CORDIS-ANIMA: A modelling system for sound and image synthesis, the general formalism," *Computer Music Journal*, vol. 17, no. 1, pp. 19–29, 1993.
- [6] J. Preece, "Interview with Ben Shneiderman," in *Human-Computer Interaction*, Y. Rogers, H. Sharp, D. Benyon, S. Holland, and J. Preece, Eds., Addison Wesley, Reading, Mass, USA, 1994.
- [7] A. D. Hunt and P. R. Kirk, *Digital Sound Processing for Music and Multimedia*, Focal Press, Oxford, UK, 1999.
- [8] W. Buxton, "There is more to interaction than meets the eye: Some issues in manual input," in *User Centered System Design: New Perspectives on Human-Computer Interaction*, D. A. Norman and S. W. Draper, Eds., pp. 319–337, Lawrence Erlbaum Associates, Hillsdale, NJ, USA, 1986.
- [9] G. W. Fitzmaurice, *Graspable user interfaces*, Ph.D. thesis, University of Toronto, Ontario, Canada, 1998.
- [10] B. Gillespie, "Introduction haptics," in *Music, Cognition, and Computerized Sound: An Introduction to Psychoacoustics*, P. R. Cook, Ed., pp. 229–245, MIT Press, London, UK, 1999.
- [11] D. M. Howard, S. Rimell, A. D. Hunt, P. R. Kirk, and A. M. Tyrrell, "Tactile feedback in the control of a physical modelling music synthesiser," in *Proc. 7th International Conference on Music Perception and Cognition*, C. Stevens, D. Burnham, G. McPherson, E. Schubert, and J. Renwick, Eds., pp. 224–227, Casual Publications, Adelaide, Australia, 2002.
- [12] S. Kenji, H. Riku, and H. Shuji, "Development of an autonomous humanoid robot, iSHA, for harmonized human-machine environment," *Journal of Robotics and Mechatronics*, vol. 14, no. 5, pp. 324–332, 2002.
- [13] C. Cadoz, A. Luciani, and J. L. Florens, "Responsive input devices and sound synthesis by simulation of instrumental mechanisms: The Cordis system," *Computer Music Journal*, vol. 8, no. 3, pp. 60–73, 1984.
- [14] B. Gillespie, *Haptic display of systems with changing kinematic constraints: The virtual piano action*, Ph.d. dissertation, Stanford University, Stanford, Calif, USA, 1996.
- [15] C. Nichols, "The vBow: Development of a virtual violin Bow haptic human-computer interface," in *Proc. New Interfaces for Musical Expression Conference*, pp. 168–169, Dublin, Ireland, May 2002.
- [16] J. Rován and V. Hayward, "Typology of tactile sounds and their synthesis in gesture-driven computer music performance," in *Trends in Gestural Control of Music*, M. Wanderley and M. Battier, Eds., pp. 297–320, Editions IRCAM, Paris, France, 2000.
- [17] D. M. Howard, S. Rimell, and A. D. Hunt, "Force feedback gesture controlled physical modelling synthesis," in *Proc. Conference on New Musical Instruments for Musical Expression*, pp. 95–98, Montreal, Canada, May 2003.

David M. Howard holds a first-class B.S. degree in electrical and electronic engineering from University College London (1978), and a Ph.D. in human communication from the University of London (1985). His Ph.D. topic was the development of a signal processing unit for use with a single channel cochlear implant hearing aid. He is now with the Department of Electronics at the University of York, UK, teaching and researching in music technology. His specific research areas include the analysis and synthesis of music, singing, and speech. Current activities include the application of bio-inspired techniques for music synthesis, physical modelling synthesis for music, singing and speech, and real-time computer-based visual displays for professional voice development. David is a Chartered Engineer, a Fellow of the Institution of Electrical Engineers, and a Member of the Audio Engineering Society. Outside work, David finds time to conduct a local 12-strong choir from the tenor line and to play the pipe organ.



Stuart Rimell holds a B.S. in electronic music and psychology as well as an M.S. in digital music technology, both from the University of Keele, UK. He worked for 18 months with David Howard at the University of York on the development of the Cymatic system. There he studied electroacoustic composition for 3 years under Mike Vaughan and Rajmil Fischman. Stuart is interested in the exploration of new and fresh creative musical methods and their computer-based implementation for electronic music composition. Stuart is a guitarist and he also plays euphonium, trumpet, and piano and has been writing music for over 12 years. His compositions have been recognized internationally through prizes from the prestigious Bourge Festival of Electronic Music in 1999 and performances of his music worldwide.

