TUTORIAL ARTICLE

Physically based sound modelling

GIOVANNI DE POLI and DAVIDE ROCCHESSO

Dipartimento di Elettronica e Informatica, Università degli Studi di Padova, via Gradenigo, 6/A, 35131 Padova, Italy
E-mail: depoli@dei.unipd.it, roc@dei.unipd.it

In multimedia art and communication, sound models are needed which are versatile, responsive to users' expectations, and have high audio quality. Moreover, model flexibility for human–machine interaction is a major issue. Models based on the physics of actual or virtual objects can meet all of these requirements, thus allowing the user to rely on high-level descriptions of the sounding entities. As long as the sound description is based on the physics of sounding objects and not only on the characteristics of human hearing, an integration with physics-based graphic models becomes possible.

1. INTRODUCTION

In the last few years, non-speech sounds have been gaining importance in multimedia systems and computer-based communication, for various reasons. On the one hand, audio cards are now a standard component in popular personal computers, and people are getting used to sophisticated sound effects added to games, multimedia presentations, and human–machine interfaces. On the other hand, research in virtual reality and multimedia has demonstrated how auditory information can be much more effective than visual information for specific tasks (Begault 1994, Madhyastha and Reed 1995). The importance of audio communication is expected to be increased as soon as *ubiquitous computing* becomes part of everyday life (MacIntyre and Feiner 1996). In the short-term future we will be surrounded by numerous-yet-unremarkable computers, so that acoustic signals will be used to draw the user's attention and give her quick access to important information. In this respect, the fundamental advantage of hearing over sight is that our ears are two open channels, providing a whole-around 'view' of acoustic events.

In order to generate, manipulate and think about sounds, it is useful to organise our intuitive sound abstractions into objects, in the same way as abstract categories are needed for defining visual objects. To this end, we believe that great advantage can be taken from the experience of sound and music researchers and artists, thus avoiding the need to rediscover early results. Sound objects have been extensively analysed by Schaeffer from a perceptual viewpoint (Schaeffer 1966), and his classification has been extended by Schafer (1977), who also introduced a catalogue of sounds organised according to referential attributes. Nowadays, a common terminology is available for describing sound objects both from a phenomenological or a referential viewpoint, and for describing collections of such objects (i.e. *soundscapes*) (Risset 1969, Truax 1978, McAdams 1987).

For effective generation and manipulation of the sound objects it is necessary to define models for sound synthesis, processing and composition. Identifying models, either visual or acoustic, is equivalent to making high-level constructive interpretations, built up from the zero level (i.e. pixels or sound samples). It is important for the model to be associated with a semantic interpretation, in such a way that an intuitive action on model parameters becomes possible. While the activity of music composition pertains to the musician, who has developed the skills for organising sounds with respect to time, the design and use of models for sound generation and processing pertain to sound researchers and engineers. Real-world applications are also exhibiting the need of an intermediate, multidisciplinary professional role, which is that of the 'New Technologies Orchestrator' (Kendall 1991). The penetration of sound models into multimedia technology is likely to be boosted as soon as MPEG-4, a new standard in audiovisual communication, is approved. This standard includes a description format for structuring audio, in such a way that sound models can be represented, integrated and exchanged (Vercoe, Gardner and Scheirer 1998).

This paper surveys a family of sound models which are based on the physical description of sound objects. One of the motivations comes from the observation that physically based modelling is being successfully introduced into the domain of visual communication (Barzel 1992, Rich, Waters, Strohecker, Schabes, Freeman, Torrence, Goldin and Roth 1994a, b), especially where interactivity is a major issue. On the other hand, the application of physically based techniques to sound modelling for multimedia is still largely unexplored, even though it is likely to provide significant improvements to multimodal communication and human–computer interaction. Our expectation is strengthened by the mass of

Organised Sound 3(1): 61–76 © 1998 Cambridge University Press. Printed in the United Kingdom.

music professionals who are becoming aware of physical modelling, and by the growing number of physics-based sound synthesis engines which are being commercialised (Yamaha Corp. 1994, Metlay 1996). Since knowledge and technology for physically based modelling are available, we don't see why we should remain stuck with inexpressive sound samples. Moreover, we believe that synchronisation and integration of aural and visual information are better achieved starting from a referential object description and developing physics-based models (Takala and Hahn 1992, Takala, Hahn, Gritz, Geigel and Lee 1993). On the contrary, we might start from phenomenological descriptions of sounds and images and develop perception-based models. However, this latter approach does not seem to be as effective for synchronisation and integration due to the large differences found in the two kinds of human perception.

Section 2 is intended to be introductory and discusses the issue of sound modelling. We propose an organisation of sound manipulations into generative and processing models. We will see how different modelling paradigms can be used for both categories, and we will divide these paradigms into signal models and physics-based models. Section 3 is the kernel of our paper and illustrates physically based modelling in some detail. Several techniques are presented for modelling sound sources and general acoustic systems. Section 4 is a physics-based view of sound processing models, such as reverberation and spatialisation techniques. Finally, section 5 points to examples of integration of physically based modelling into multimedia presentations.

2. SOUND MODELLING

Much of the communicative power of a picture or a sound is in the possibility of associating it to 'objects'. Such objects can be described either with a phenomenological or a referential vocabulary, and the semantic interpretation can be different according to the description that is adopted (Schafer 1977). Models are constructive descriptions of objects which can be used for generating new sound samples or for modifying existing ones. For example, in the case of visual communication, models allow image transformations in order to get many different views. In the same way, for sounds we feel the necessity to build models allowing interaction with the sound object and overcoming the slavery to 'frozen' sounds. It is also reasonable to expect that if the models for visual and acoustic 'views' are alike, the communicative potential of multimedia systems is increased. This statement holds for communication 'to' a spectator, but is even more important when we are dealing with communication 'with' an agent. For instance, a slider on a screen can be thought of as a static image, but if we are planning that the user will act on the slider, then we have to make an abstraction of the slider object, and associate it to a model in order to get a context-dependent view. By analogy, the acoustic event which is produced by friction between two objects can be represented by a prerecorded sound. But if we want the user to be able to influence the contact pressure between the two surfaces, then we have to conceive a 'friction' object and a model accounting for possible friction configurations.

In the sound domain, we define *generative models* as those models which give computational form to abstract objects, thus representing a sound generation mechanism. Sound fruition requires a further processing step, which accounts for sound propagation in enclosures and for the listener's position relative to the sound source. Modifications such as these, which intervene on the attributes of sound objects, are controllable by means of *processing models*.

Generative models can represent the dynamics of real or virtual generating objects (*physics-based models*), or they can represent the physical quantities as they arrive at human senses (*signal models*) (Smith 1991) (see figure 1). In our terminology, signal models are models of signals as they arrive at the ears, rather than models of the human perception mechanisms. The connection with human perception is better understood when considering the evaluation criteria of the generative models. The evaluation of a signal model should be done according to certain perceptual cues. On the other hand, physics-based models are better evaluated according to the physical behaviour involved in the sound production process.

Processing models can also be classified with respect to their commmitment to model the causes or the effects of sound propagation from the source to the ears. For example, a reverberation system can be built from an abstract signal processing algorithm where its parameters are mapped to perceptual cues (e.g. warmth or brilliance) or to physical attributes (e.g. wall absorption or diffusion).

In classic sound synthesis, signal models dominate the scene, due to the availability of very efficient and widely applicable algorithms (e.g. frequency modulation). Moreover, signal models allow us to design sounds as objects *per se* without having to rely on actual pieces of material which act as a sound source.

However, many people are becoming convinced of the fact that physics-based models are closer to the users/designers' needs of interacting with sound objects. The semantic power of these models seems to make them preferable for this purpose. The computational complexity of physically based algorithms is becoming affordable with modern-day technology, even for realtime applications. We keep in mind that the advantage we gain in model expressivity comes at

https://doi.org/10.1017/S1355771898009182 Published online by Cambridge University Press

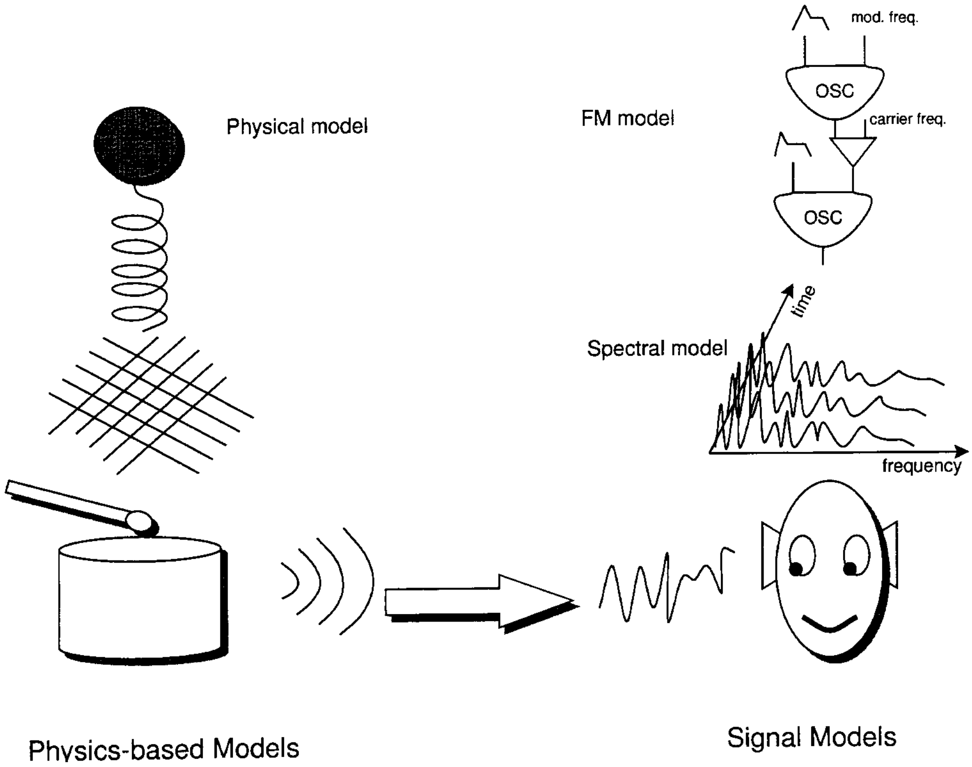

Figure 1. Physics-based models and signal models.

the expense of the flexibility of several general-purpose signal models. For this reason, signal models keep being the model of choice in many applications, especially for music composition.

In the perspective of a multisensorial unification under common models, physics-based models offer an evident advantage over signal models. In fact, the mechanisms of perception for sight and hearing are very different, and a unification at this level looks difficult. Even though analogies based on perception are possible, an authentic sensorial coherence seems to be ensured only by physics-based models. The interaction among various perceptions seems to be an essential feature if we want to maximise the amount of information conveyed to the spectator/actor. The unification of visual and aural cues is more properly done at the level of abstractions, where the cultural and experiential aspects become fundamental. Thus, building models closer to the abstract object, as it is conceived by the designer, is a fundamental step in the direction of this unification.

The importance of physics-based models in multimedia communication is also emphasised by use of new devices for human–machine interaction. Several gestural controllers with force feedback have already been developed (MacIntyre and Feiner 1996), and the association of these devices with icons, menus and other user-interface objects is being explored. As an example of application, imagine a blind user who searches an icon by means of his gestural controller, finds it by feeling the opposite force exerted by the edge, and then overcomes this force and gets into the icon. If interface objects are associated with sound events that naturally follows the physical constraints of the environment being explored, then interaction is expected to become more intuitive. This expectation is reinforced by recent experiments which have showed that the sense of presence is increased even by the insertion of sounds uncorrelated with the objects of a virtual environment, provided that these sounds are properly spatialised (Hendrix and Barfield 1996).

Since the main focus of this paper is on physics-based models, we will not explain signal models any further, pointing to this end to several good tutorial articles and books on sound synthesis techniques (Moorer 1972, Moorer 1977, De Poli 1991, Moore 1990, De Poli, Piccialli and Roads 1991, Haus 1993, Roads 1996, Roads, Pope, Piccialli and De Poli 1997). Instead, section 3 will cover the most relevant paradigms in sound physically based modelling.

3. MODELLING SOUND SOURCES

The family of physics-based models consists of all the algorithms generating sounds as a side effect of a more general process of simulation of a physical phenomenon. Physics-based models can be classified

https://doi.org/10.1017/S1355771898009182 Published online by Cambridge University Press

according to the way of representing, simulating and discretising the physical reality. Hence, we can talk about cellular, finite-difference and waveguide models, thus intending that these categories are not disjoint but, in some cases, they represent different viewpoints on the same computational mechanism. Moreover, physics-based models do not necessarily have to be based on the physics of the real world, but they can, more generally, gain inspiration from it; in this case we will talk about pseudo-physical models.

In this paper, the approach to physically based synthesis is carried on with particular reference to realtime applications, and therefore the complexity of algorithms plays a key role. We can summarise the general objective of the presentation by saying that we want to obtain models for large families of sounding objects, and these models have to provide a satisfactory representation of the acoustic behaviour with the minimum computational effort.

3.1. Functional blocks

In real objects we can often outline functionally distinct parts, and express the overall behaviour of the system as the interaction of these parts. Outlining functional blocks helps the task of modelling, because for each block a different representation strategy can be chosen. In addition, the range of parameters can be better specified in isolated blocks, and the gain in semantic clearness is evident. Our analysis stems from musical instruments, and this is justified by the fact that the same generative mechanisms can be found in many other physical objects. In fact, we find it difficult to think about a physical process producing sound and having no analogy in some musical instrument; for instance, friction can be found in bowed string instruments, striking in percussion instruments, air turbulences in jet-driven instruments, etc. Generally speaking, we can think of musical instruments as a specialisation of natural dynamics for artistic purposes. Musical instruments are important for the whole area of sonification in multimedia environments because they constitute a testbed where the various simulation techniques can easily show their merits and pitfalls.

The first level of conceptual decomposition that we can devise for musical instruments is represented by the interaction scheme of figure 2, where two functional blocks are outlined: a resonator and an exciter. The resonator sustains and controls the oscillation, and is related with sound attributes like pitch and spectral envelope. The exciter is the place where energy is injected into the instrument, and it strongly affects the attack transient of sound, which is fundamental for timbre identification. The interaction of exciter and resonator is the main source of richness and variety of nuances which can be obtained from a

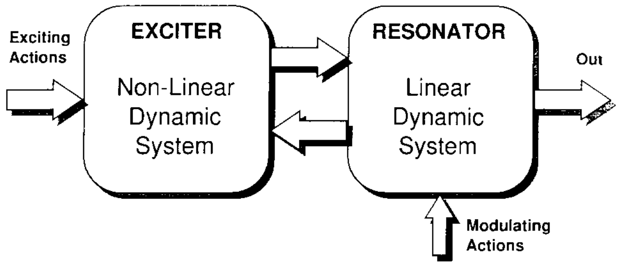

Figure 2. Exciter–resonator interaction scheme.

musical instrument. When translating the conceptual decomposition into a model, two dynamic system are found (Borin, De Poli and Sarti 1992b): the excitation block, which is strongly nonlinear, and the resonator, supposed to be linear to a great extent. The player controls the performance by means of inputs to the two blocks. The interaction can be 'feedfoward', when the exciter does not receive any information from the resonator, or 'feedback', when the two blocks exert a mutual information exchange. In this conceptual scheme, the radiating element (bell, resonating body, etc.) is implicitly enclosed within the resonator. In a clarinet, for instance, we have a feedback structure where the reed is the exciter and the bore with its bell acts as a resonator. The player exerts exciting actions such as controlling the mouth pressure and the embouchure, as well as modulating actions such as changing the bore effective length by opening and closing the holes. In a plucked string instrument, such as a guitar, the excitation is provided by plucking the string, the resonator is given by the strings and the body, and modulating actions take the form of fingering. The interaction is slightly feedback, so that a feedforward scheme can be adopted with a good approximation: the excitation adjusts the initial conditions and the resonator is then left free to vibrate.

In practical physical modelling, the block decomposition can be extended to finer levels of detail, as both the exciter and the resonator can be further decomposed into simpler functional components, e.g. the holes and the bell of a clarinet as a refinement of the resonator. At each stage of model decomposition, we are faced with the choice of expanding the blocks further (white-box modelling), or just considering the input–output behaviour of the basic components (black-box modelling). In particular, it is very tempting to model just the input–output behaviour of linear blocks, because in this case the problem reduces to a filter design. However, such an approach provides structures whose parameters are difficult to interpret and, therefore, to control. In any case, the decomposition of an instrument into blocks corresponds to a similar decomposition in digital structures which can lead to efficient algorithms. In fact, we can focus on functionally distinct parts and try to obtain the best results from each before coupling them together (Borin, De Poli and Sarti 1992a).

In digital implementations, a third block is often found in between the two blocks of exciter and resonator. This is an interaction block and it can convert the variables used in the exciter to the variables used in the resonator, or avoid possible anomalies introduced by the discretisation process. The idea is to have a sort of adaptor for connecting different blocks in a modular way. This adaptor might also serve to compensate for the simplifications introduced by the modelling process. To this end, a residual signal might be introduced in this block in order to improve the sound realism. Simplifications in the model of the exciter often lead to residues which take the form of pulsed noises, and are due to complex physical dynamics such as turbulence (Chafe 1990, Drioli and Rocchesso 1997). For example, the interaction block for a reed instrument should introduce a fluid-dynamic noise whose amplitude is modulated by the air flow (Rocchesso and Turra 1993). The limits of a detailed physical simulation are also found when we try to model the behaviour of a complex linear vibrating structure, such as a soundboard; in such cases it can be useful to record its impulse response and include it in the excitation signal as it is provided to a feedforward interaction scheme. Such a method is called commuted synthesis, since it makes use of commutativity of linear, time-invariant blocks (Smith 1993, Valimaki, Huopaniemi, Karjalainen and Janosy 1996).

It is interesting to notice that the integration of sampled noises or impulse responses into physical models is analogous to texture mapping in computer graphics (Blinn and Newell 1976). In both cases the realism of a synthetic scene is increased by insertion of snapshots of textures (either visual or aural) taken from actual objects and projected onto the model.

3.2. Cellular models

A possible approach to simulation of complex dynamical systems is their decomposition into a multitude of interacting particles. The dynamics of each of these particles are discretised and quantised in some way to produce a finite-state automaton (a cell), suitable for implementation on a processing element of a parallel computer. The discrete dynamical system consisting of a regular lattice of elementary cells is called a cellular automaton (von Neumann 1966, Wolfram 1984). The state of any cell is updated by a transition rule which is applied to the previous-step state of its neighbourhood. When the cellular automaton comes from the discretisation of a homogeneous and isotropic medium it is natural to assume functional homogeneity and isotropy, i.e. all the cells behave according to the same rules and are connected to all their immediate neighbours in the same way (von Neumann 1966). If the cellular automaton has to be representative of a physical system, the state of cells must be characterised by values of selected physical parameters, e.g. displacement, velocity, force.

Several approaches to physically based sound modelling can be recast in terms of cellular automata, the most notable being the CORDIS-ANIMA system introduced by Cadoz and his coworkers (Florens and Cadoz 1991, Cadoz, Luciani and Florens 1993, Incerti and Cadoz 1995), who came up with cells as discrete-time models of small mass–spring–damper systems, with the possible introduction of nonlinearities. The main goal of the CORDIS-ANIMA project was to achieve high degrees of modularity and parallelism, and to provide a unified formalism for rigid and flexible bodies. The technique is very expensive for an accurate sequential simulation of wide vibrating objects, but is probably the only effective way in the case of a multiplicity of micro-objects (e.g. sand grains) or for very irregular media, since it allows an embedding of the material characteristics (viscosity, etc.). An example of CORDIS-ANIMA network discretising a membrane is shown in figure 3, where we have surrounded by triangles the equal cells which provide output variables depending on the internal state and on input variables from neighbouring cells. Even though the CORDIS-ANIMA system uses heterogeneous elements such as matter points or visco-elastic links, figure 3 shows how a network can be restated in terms of a cellular automaton showing functional homogeneity and isotropy.

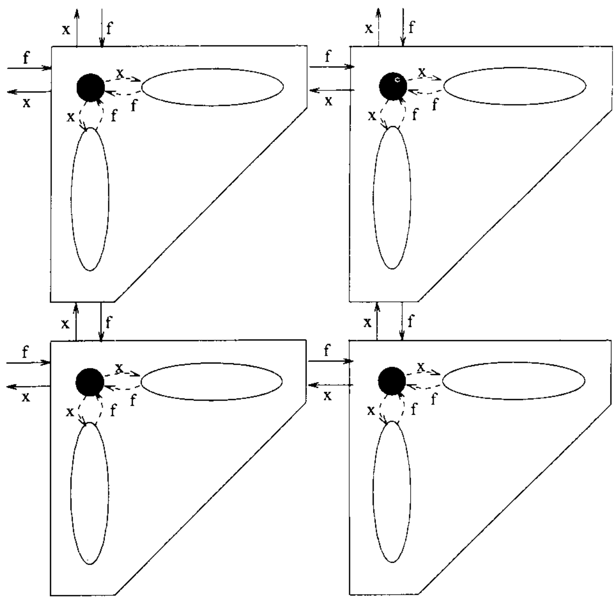

Figure 3. A CORDIS-ANIMA network (a piece of a rectangular mesh) restated as a cellular automaton. Black dots indicate mass points and white ovals indicate link elements, such as a visco-elastic connection. x represents a position variable and f represents a force variable.

https://doi.org/10.1017/S1355771898009182 Published online by Cambridge University Press

A cellular automaton is inherently parallel, and its implementation on a parallel computer shows excellent scalability (Spezzano, Talia, Gregorio, Rongo and Spataro 1996). Moreover, in the case of the multiplicity of micro-objects, it has shown good effectiveness for joint production of audio and video simulations (Cadoz, Luciani and Florens 1994).

In Chareyon (1990) it is shown how a one-dimensional cellular automaton can implement the model of a plucked string, which is equivalently obtained with the waveguide technique (Jaffe and Smith 1983, Karplus and Strong 1983) presented in section 3.4. Similarly, it might be possible to show that a two-dimensional cellular automaton can implement the model of a membrane as it is expressed by a waveguide mesh. However, as we will see in sections 3.3 and 3.4, when the system to be modelled is the medium where waves propagate, the natural approach is to start from the wave equation and to discretise it or its solutions. In the fields of finite-difference methods or waveguide modelling, theoretical tools do exist for assessing the correctness of these discretisations. On the other hand, only qualitative criteria seem to be applicable to cellular automata in their general formulation.

3.3. Finite-difference models

When modelling vibrations of real-world objects, it can be useful to consider them as rigid bodies connected by lumped, idealised elements (e.g. dashpots, springs, geometric constraints, etc.) or, alternatively, to treat them as flexible bodies where forces and matter are distributed over a continuous space (e.g. a string, a membrane, etc.). In the two cases the physical behaviour can be represented by ordinary or partial differential equations, whose form can be learned from physics textbooks and whose coefficient values can be obtained from physicists' investigations or from direct measurements. These differential equations often give only a crude approximation of reality, as the objects being modelled are just too complicated. Moreover, as we try to solve the equations by numerical means, a further amount of approximation is added to the simulated behaviour, so that the final result can be quite far from the real behaviour.

One of the most popular ways of solving differential equations is finite differencing, where a grid is constructed in the spatial and time variables, and derivatives are replaced by linear combinations of the values on this grid. The two main problems to be faced when designing a finite-difference scheme for a partial differential equation are: numerical losses and numerical dispersion. There is a standard technique (Press, Flannery, Teukolsky and Vetterling 1988) for evaluating the performance of a finite-difference scheme in contrasting these problems: the von Neumann analysis. It can be quickly explained for the simple case of the ideal string (or the ideal acoustic tube), whose wave equation is (Morse 1991)

$$\frac{\partial^2 p(x,t)}{\partial t^2} = c^2 \frac{\partial^2 p(x,t)}{\partial x^2}, \tag{1}$$

where c is the wave velocity of propagation, t and x are the time and space variables, and p is the string displacement (or acoustic pressure). By replacing the second derivatives by central second-order differences, the explicit updating scheme for the ith spatial sample of displacement (or pressure) is

$$\begin{aligned} p(i, n+1) = {} & 2\left(1 - \frac{c^2 \Delta t^2}{\Delta x^2}\right) p(i,n) - p(i, n-1) \\ & + \frac{c^2 \Delta t^2}{\Delta x^2}[p(i+1,n) + p(i-1,n)], \end{aligned} \tag{2}$$

where Δt and Δx are the time and space grid steps. The von Neumann analysis assumes that the equation parameters are locally constant and checks the time evolution of a spatial Fourier transform of equation (2). In this way a spectral amplification factor is found whose deviations from unit magnitude and linear phase give respectively the numerical loss (or amplification) and dispersion errors. For the scheme (2) it can be shown that a unit-modulus amplification factor is ensured as long as the Courant–Friedrichs–Lewy condition (Press *et al.* 1988)

$$\frac{c \Delta t}{\Delta x} \leqslant 1 \tag{3}$$

is satisfied, and that no numerical dispersion is found if equality applies in expression (3). A first consequence of expression (3) is that only strings having a length which is an integer number of $c\Delta t$ are exactly simulated. Moreover, when the string deviates from ideality and higher spatial derivatives appear (physical dispersion), the simulation becomes always approximate. In these cases, the resort to implicit schemes can allow the tuning of the discrete algorithm to the amount of physical dispersion, in such a way that as many partials as possible are reproduced in the band of interest (Chaigne 1992).

It is worth noting that if c in equation (1) is a function of time and space, the finite-difference method retains its validity because it is based on a local (in time and space) discretisation of the wave equation. Another advantage of finite differencing over other modelling techniques is that the medium is accessible at all the points of the time–space grid, thus maximising the possibilities of interaction with other objects.

When the objects being simulated are rigid bodies, they can be described by ordinary differential equations, and a plethora of techniques are available for their integration (Press *et al*. 1988). However, attention must be paid to stability issues and to the correct reproduction of important physical attributes. These issues are strongly dependent on the numerical integration technique and on the sampling rate which are to be used. In most of the cases there is no better method than trying several techniques and comparing the results, but the task is often facilitated by the fact that the strong nonlinearities are lumped. For example, in Gazengel, Gilbert and Amir (1995), the dynamics of a clarinet reed is discretised by using a fourth-order Runge–Kutta method, Euler differencing, and bilinear transformation (Oppenheim and Schafer 1989). The Runge–Kutta method turns out to be unstable for low sampling rates, while Euler differencing shows a poor reproduction of the characteristic resonance of the reed, due to numerical losses. For that specific case, the best choice seems to be the bilinear transform, which corresponds to a trapezoidal integration of the differential equations, possibly with some warping of the frequency axis (Moorer 1983) for adjusting the resonance central frequency. The discretisation by impulse invariance (Oppenheim and Schafer 1989) is also a reliable tool when aliasing can be neglected, and its performance is often preferable to bilinear transforming in acoustic modelling (van Walstijn and Dubbelboer 1997) because it is free of frequency warping and artificial damping.

Further studies are needed for establishing the most suitable discretisation techniques for the many kinds of lumped dynamics, with special attention to be paid to the looped connection of lumped nonlinear elements with memoryless nonlinearities and distributed resonators. Several techniques from signal processing and numerical analysis are yet to be evaluated, while some general methodologies are just being proposed. In this respect, section 3.4 will show how, switching to a wave-variable representation of the physical quantities, it is possible to apply the paradigms of wave digital filters and waveguide networks to the lumped and distributed elements, respectively.

3.4. Wave models

When discretising physical systems a key role is played by the efficiency and accuracy of the discretisation technique. Namely, we would like to be able to simulate simple vibrating structures and exciters with no artefacts (e.g. aliasing or noncomputable dependencies) and with low computational complexity. Due to its good properties with respect to these two criteria, the most popular way of approaching physical modelling of acoustic systems makes use of wave variables instead of absolute physical quantities. When wave variables are adopted in the digital domain for representing lumped components, this approach is called wave digital filtering (Fettweis 1986) and proceeds roughly as follows:

- Given the dual physical variables p and u (let us call them pressure and velocity), define the pressure waves

$$\begin{cases} p^+ = (p + Z_0 u)/2 \\ p^- = (p - Z_0 u)/2, \end{cases} \tag{4}$$

  where Z_0 is an arbitrary reference impedance.
- It is possible to show that a lumped component having impedance $Z(s)$ in the complex variable s can be represented in pressure waves by

$$p^- = \frac{Z(s) - Z_0}{Z(s) + Z_0} p^+. \tag{5}$$

- Apply the bilinear transform $z = (1 + s)/(1 - s)$ to equation (5) and choose Z_0 in such a way that there is at least one delay element in any signal path connecting p^+ with p^- in the digital version of equation (5).
- Apply the Kirchhoff principles (Belevitch 1968) to junctions of components derived by the previous steps (wave-scattering formulation of the network).

On the other hand, when the components to be modelled are distributed wave-propagating media, digital waveguide networks (Smith 1992) can be used for simulating them. In these models the physical variables are decomposed into their respective wave variables and their propagation is simulated by means of delay lines. Low-pass and all-pass filters are added to simulate dissipative and dispersive effects in the medium.

As opposed to finite differencing, which discretises the wave equation (see equations (1) and (2)), waveguide models come from discretisation of the solution of the wave equation. The solution to the one-dimensional wave equation (1) was found by D'Alembert in 1747 in terms of travelling waves:

$$p(x, t) = p^+(t - x/c) + p^-(t + x/c). \tag{6}$$

Equation (6) shows that the physical quantity p (e.g. string displacement or acoustic pressure) can be expressed as the sum of two wave quantities travelling in opposite directions. In waveguide models, waves are sampled in space and time in such a way that equality holds in expression (3). If propagation is ideal, i.e. linear, nondissipative and nondispersive, wave propagation is represented in the discrete-time domain by a couple of digital delay lines.

As an example, let us consider a rough model of a clarinet. The nonlinear block representing the reed can simply be an instantaneous nonlinear map relating the particle velocity u to the pressure difference,

https://doi.org/10.1017/S1355771898009182 Published online by Cambridge University Press

as

$$\begin{cases} \Delta p = P_0 - p \\ u = F(\Delta p), \end{cases} \tag{7}$$

where P_0 is the player's mouth pressure and p is the pressure inside the bore at the excitation point (McIntyre, Schumacher and Woodhouse 1983). Let us assume that there is, right after the exciter, a tract of constant-section tube having the characteristic impedance Z_0, and terminated by the radiation impedance of the bell. In order to simplify the analysis, we can see the tube as a lossless transmission line for the dual quantities flow and pressure, so that the D'Alembert decomposition (6) does hold. Introducing the analogy between electrical and mechanical variables, we can depict the system as an electric circuit, and recognise the three sections of generator, line and load. This permits us to introduce qualitative reasoning about the conditions for establishing oscillations and the periodic, multiperiodic or chaotic nature of the oscillations themselves (Rodet 1993, 1994).

In order to achieve a satisfactory audio quality for physics-based models, it is often necessary to use structures which are far more complex than that of the simplified clarinet. While more sophisticated exciters are considered in section 3.5, let us consider deviations from ideal propagation due to losses and dispersion in the resonator. Usually, these linear effects are lumped and simulated with a few filters which are cascaded with the delay lines. Losses due to terminations, internal frictions, etc., give rise to gentle low-pass filters, whose parameters can be identified from measurements (Valimaki *et al*. 1996). Wave dispersion, which is often due to medium stiffness, is simulated by means of all-pass filters whose effect is to produce a frequency-dependent propagation velocity (Rocchesso and Scalcon 1996).

In order to increase the computational efficiency, delay lines and filters should be lumped into as few processing blocks as possible. However, when considering the interaction with an exciter or signal pick-up from certain points of the resonator, the process of commuting and lumping linear blocks must be done with care. If the excitation is a velocity signal injected into a string, it will produce two velocity waves outgoing from the excitation point, and therefore at least two delay lines will be needed to represent propagation. The process of commuting and lumping must maintain the semantics at the observation points, while at the other points of the structure it is not necessary to have a strict correspondence with the physical reality.

Another aspect that we like to mention is that of simulating fractional delays. This is necessary when modelling musical instruments, since the proper tuning usually requires a space discretisation much finer than dictated by the sample rate. More generally, fractional delay lengths are needed whenever time-varying acoustic objects (such as a string which is varying its length or tension) are being modelled by digital waveguide networks. For this purpose, all-pass filters or Lagrange interpolators of various orders can be used (Laakso, Valimaki, Karjalainen and Laine 1996), the former suffering from phase distortion at high frequency, the latter suffering from both phase and amplitude distortion. However, low-order filters of both families can be used satisfactorily in most practical cases (Laakso *et al*. 1996). In other cases, the problem of designing a tuning filter is superseded by the more general problem of modelling wave dispersion (Rocchesso and Scalcon 1996).

So far, we have talked about one-dimensional resonators, but many musical instruments (e.g. percussion) and most of the real-world objects are subject to deformation along several dimensions. The algorithms presented so far can be adapted to the case of multidimensional propagation of waves, even though new problems of efficiency and accuracy arise. All the models grow in computational complexity with the increase of dimensionality, and for any of them, the choice of the right discretisation grid is critical. For example, a rectangular waveguide mesh can be effective for simulating vibrating surfaces, but in this case the wave propagation is exact only along the diagonals of the mesh, while elsewhere it is affected by a dispersive phenomenon due to the fact that we are simulating circular waves by portions of plane waves. Waveguide meshes are shown to be equivalent to special kinds of finite-difference schemes (Van Duyne and Smith 1993b), so that the von Neumann analysis can be used for evaluating the numerical properties of the algorithms. Special attention has to be paid to the dependence of the dispersion factor on frequency, direction and mesh topology, because this influences the distribution of resonances, and therefore affects the tone colour and intonation. For membrane simulation, one of the most accurate yet efficient meshes is the triangular mesh (Fontana and Rocchesso 1998). Interpolated waveguide meshes have recently been introduced for improving the accuracy while using simple topologies (Savioja and Valimaki 1997). For three-dimensional wave propagation, the tetrahedral mesh is very attractive because its junctions can be implemented without any multiplication (Van Duyne and Smith 1995).

3.5. Models of nonlinearities

Nonlinearities assume a great importance in acoustic systems, especially where a wave-propagation medium is excited. A good model for these nonlinear mechanisms is essential for timbral quality, and is the

https://doi.org/10.1017/S1355771898009182 Published online by Cambridge University Press

real kernel of a physical model, which could otherwise be reduced to linear postprocessing of an excitation signal. Since the area where the excitation takes place is usually small, it makes sense to use lumped models for the excitation nonlinearity. In some cases, physical measurements provide a representation of the relation among some physical variables involved in excitation, and this relation can be directly implemented in the simulation. For example, for a simplified bowed string the transversal velocity as a function of force can be found in the literature (McIntyre *et al.* 1983) for different values of bow pressure and velocity (which are control parameters). The instantaneous nonlinear function can be approximated analytically or sampled and stored in a lookup table, which is in general multidimensional (Rocchesso and Turra 1993).

A first improvement over instantaneous nonlinear excitation is considering the dynamics of the exciter: this implies the introduction of a state (i.e. memory) inside the nonlinear block. When physicists study the behaviour of musical instruments, they often use dynamic models of the exciter. A nontrivial task is the translation of these models into efficient computational schemes for realtime sound synthesis. A general structure has been found for good simulations of wide instrumental families (Borin *et al.* 1992b, Rocchesso 1993). In figure 4, this structure is schematically depicted. The block NL is a nonlinear instantaneous function of several variables, while the block L is a linear dynamic system enclosing the exciter memory. Studies and simulations have shown that reeds, air jets, bows and percussions, can all be represented by this scheme to a certain extent.

If a lumped modelling of dynamics is used, it has to be represented by discretisation of some differential equations. To this end, one approach is to apply a finite-difference scheme to these equations and introduce the instantaneous nonlinearity in the

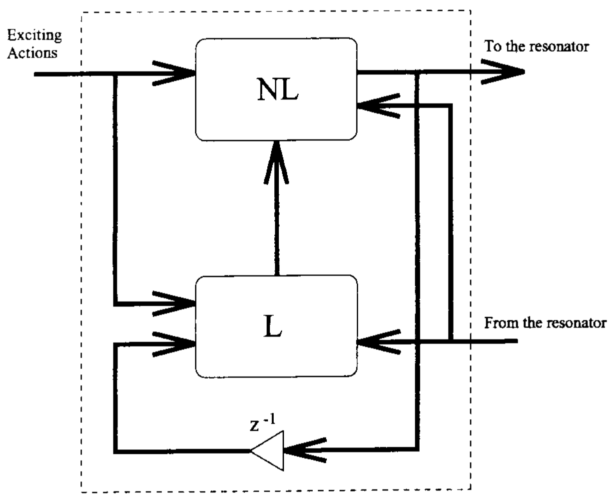

Figure 4. Scheme for a dynamic exciter.

resulting signal flowgraph (Chaigne and Askenfelt 1994, Van Duyne, Pierce and Smith 1994). This procedure must be tackled carefully because it is easy to come up with delay-free (noncomputable) loops: in some cases these inconsistencies can be overcome by introducing fictitious delay elements, but at the sampling rates that we can afford it is likely that these delays are not acceptable, so that other techniques are needed. A notable example is found in the piano hammer-string interaction (Borin and De Poli 1996), whose model can be applied to several percussive sound sources. The computable discretisation scheme of the nonlinear hammer is obtained from the straightforward finite difference approximation of the dynamics, from which a term dependent on previous and known terms is separated from an instantaneous and unknown term; finally, the instantaneous nonlinearity is recast into new variables in such a way that the delay-free loop is cancelled (Borin, De Poli and Rocchesso 1997).

Another approach to the representation of the exciter dynamics is to resort to a model based on lumped linear or nonlinear circuit components. The circuit components can be translated into the discrete-time domain using wave digital filter modelling (Fettweis 1986), which has recently been extended to cover nonlinear elements without (Meerkotter and Scholtz 1989) or with (Sarti and De Poli 1996) memory, i.e. resistances or reactances.

When the need is not for a special-case model, but for a model which is adaptable to several sound sources, or when it is desirable to tune the model from sampled sounds, other representations have to be chosen for the nonlinear exciter. It has been proposed to express one-dimensional memoryless nonlinearities as a polynomial function whose coefficients can be identified by Kalman filtering (Cook 1991), or adaptively by the LMS algorithm (Scavone 1994).

Alternatively, some general-purpose approximation network might be used as a generalised nonlinearity. For example, a radial-basis-function network (Poggio and Girosi 1990) built with a small number of Gaussian kernels has shown its effectiveness in representing the severe nonlinearities found in wind and string instruments (Drioli and Rocchesso 1997). The parameters of these networks have to be identified by some global optimisation technique, such as the genetic algorithm, applied to solve a spectral or waveform matching problem.

Nonlinearities might occur in the resonator as well, especially when it is driven into large vibrations. These nonlinearities are usually mild, so that simple saturation characteristics introduced somewhere in the resonator work just fine. However, for some sound sources the resonator nonlinearity is more critical, and special techniques must be devised in order to ensure that energy is not introduced or lost

https://doi.org/10.1017/S1355771898009182 Published online by Cambridge University Press

improperly. For instance, a time-variant one-pole allpass filter has been proposed for reproducing the kind of nonlinearity found in the bridge of the sitar (Van Duyne *et al*. 1994). Similar nonlinearities are found distributed in two-dimensional resonators, such as plates, and it is not yet clear how to simulate them by means of a small number of filters, properly placed in the computational structure.

4. MODELS FOR SOUND PROCESSING

The mechanical or fluid-dynamical behaviour of actual objects can transfer some energy to the air, which propagates it by means of pressure and particle-velocity waves. If generative models account for the production of these pressure waves in the proximity of the sounding object, processing models account for the effects introduced on pressure waves by propagation in air, interaction with surrounding objects, and enclosure surfaces. In multimedia systems there is also the need of special sound modifications, such as flanging or phasing, which are not necessarily related with actual physical phenomena, and which are out of the scope of this paper. The reader interested in these 'special effects' is referred to Bloom (1985) and Orfanidis (1996).

4.1. Sound spatialisation

Any human being is capable of detecting the direction a certain sound is coming from with a precision of a few degrees (Blauert 1983). The three necessary (although not always sufficient) parameters needed for detecting the direction in which a sound source is located are: the interaural time difference (ITD), interaural intensity difference (IID) and head-related transfer functions (HRTF) (Kendall 1995a). The ITD is a time shift between the two signals arriving at the ears, and is due to the different lengths of the direct paths going from the source to the left and right ears. The IID is due to the different attenuation the waves are subject to along these paths. These two quantities alone do not solve the localisation problem in an unambiguous manner; they can only discriminate among source positions in a hemiplane. For a three-dimensional localisation we have to deal with HRTFs, which are different filtering patterns offered by the head to signals coming from different directions (Begault 1994). These three parameters have to be simulated with good accuracy if a realistic three-dimensional acoustic field has to be provided through headphones. For simulations of immersive environments, the ITD, IID and HRTF should be updated in realtime in order to follow the movements of the source and of the user's head. A great deal of effort has been devoted to the efficient yet accurate implementation of time-varying filters for sound localisation (Wightman and Kistler 1989a, b, Burgess 1992, Kistler and Wightman 1992, Wenzel 1992, Begault 1994, Begault and Erbe 1994, Gardner and Martin 1994, Jot and Warusfel 1995). The introduction of physical attributes of the enclosure can add other cues for having a correct spatial impression of the source. For example, the ratio between the intensities of the direct and reverberated sound is often sufficient to control the perceived distance from the source (Chowning 1971). We will see in section 4.2 how a physically based model of room acoustics can give convincing reverberation.

While spatial audio display has clearly shown its effectiveness in reducing the user's reaction time to an alert signal (Begault 1994), its influence on the perceived sence of presence or realism is somewhat controversial and seems to depend on the kind of auditory interface. Hendrix and Barfield have shown that the addition of spatialised auditory events, even though uncorrelated with visual events, significantly increases the sense of presence while leaving unchanged the sense of realism of a virtual environment (Hendrix and Barfield 1996). Further studies are needed to assess the effects of a more refined spatialisation (e.g. with head tracking) of object-dependent sounds. We expect that these studies will show an increase in realism as well as in presence for navigation in virtual environments.

The techniques for sound spatialisation that we have described so far are based on psychological and physiological attributes of the listener, and on the position of the source with respect to the ears. Since these techniques do not rely on the physics of wave propagation in an enclosure, and rely heavily on the use of headphones, they have some drawbacks that we are going to list briefly:

- HRTF should be individualised (earprints). Any use of standard HRFTs introduces perceptual distortions such as front–back confusion or poor source externalisation (Wenzel 1992, Begault 1994).
- Head tracking is needed when the user is not constrained by the visual display to keep the head in a steady position.
- Listening through headphones is an obstacle to a true sensory equivalence, since the listener receives cutaneous information that an acoustic display is present (Hendrix and Barfield 1996).
- Listening through headphones is not practical when the audience is somewhat wide.

The adoption of loudspeakers for acoustic display forces us to change the model of sound spatialisation. In fact, a couple of loudspeakers can reproduce the same spatial image as a headphone only in an anechoic chamber (Schroeder 1973), which is a quite rare listening room. A model which seems to be

https://doi.org/10.1017/S1355771898009182 Published online by Cambridge University Press

suited to a set of loudspeakers is the two-room model by Moore (1990). In this model the actual listening space is embedded inside the larger virtual room that we intend to simulate. We assume that sound is coming into the inner room from the outer room by holes in the walls, and these holes correspond to the actual positions of the loudspeakers. The model consists of a prescribed number of delay lines, each of them corresponding to a path connecting the source with a loudspeaker. The number of paths depends on the number of wall reflections we are considering. For a convincing perception of space it is often sufficient to consider only the direct paths and the first reflections, leaving to a diffuse reverberation model the simulation of the sound tail produced by further reflections. Such a simplification allows realtime implementation on existing digital signal processors (Ballan, Mozzoni and Rocchesso 1994).

Spatialisation models should take into account a couple of important physical attributes of a sounding object, namely its size and its radiation properties. The perceived size of an object can be controlled by the cross-correlation of signals fed to different channels in a multiple-loudspeaker display. Special filters are devised which allow a gradual adjustment of object size (Kendall 1995b). For simulating the radiation properties of actual sounding objects, all we have to do is specify an intensity function dependent on direction. This radiation pattern is inserted in the localisation model, and rays departing from the source are weighted by this direction-dependent pattern. A more complex problem is constructing loudspeaker systems which can reproduce a specific radiation pattern as found in a given sounding object (Warusfel, Kahle and Jullien 1993, Causse, Derogis and Warusfel 1995).

4.2. Room modelling and reverberation

The environment participates in sound propagation by direct modification of sound waves when they hit objects and enclosure surfaces. A real surface typically shows a behaviour between a perfect mirror and a perfect (Lambertian) diffusor, and its reflecting properties are frequency dependent (Kuttruff 1991).

The simulation of the diffusive properties of all the objects involved in sound propagation in a given environment is a hard task. The problem is similar to that encountered in illumination of scenes (Cohen and Wallace 1993), with the additional problem that propagation time of sound in air cannot be neglected. Even recasting the problem in terms of interaction of objects by force fields, the computational time is far beyond a possible implementation in real time. Somehow luckily, for a satisfactory perception of ambience, the human ear requires only a rough simulation of the room reverberation properties. Classically, reverberation was obtained by recursive filters having little resemblance of the physical reality of the target room. On the contrary, they only reproduced some statistical qualities of the reverberated sound (Schroeder 1962, 1973, Moorer 1979). This implied a difficult parametrisation of reverberators and, almost always, the user relied on some sets of magic numbers, with an evident sacrifice in versatility.

One of the authors has recently introduced a sound-processing model called 'the Ball within the Box' (BaBo) which is based on three different perspectives on wave propagation in an enclosure (Rocchesso 1995). In the proximity of the sound source a time-domain perspective best describes wave propagation, and ray tracing or the image method (Borish 1984) can be used to come up with a tapped delay which accounts for the direct signals and the early reflections. Whenever a sufficiently long distance has been travelled by waves from the sound source, we can adopt the simpler plane-wave description, which is easily interpreted in the frequency domain in terms of normal modes. This level is well described by a bunch of recirculating delay lines which can be exactly parametrised from the geometrical description of a rectangular box. Each of the delay lines corresponds to a plane-wave closed path. The third level of description takes into account fine-grain phenomena, such as diffusion, which are missed by the other perspectives. These effects are averaged out and lumped in a single scattering object (i.e the Ball) inserted in a perfectly reflecting enclosure (i.e. the Box). From a computational viewpoint the scattering object is represented by a feedback matrix which redistributes energy among the delay lines. Figure 5 shows the computational structure of the BaBo model, where additional filters T, G and A have been inserted for simulating losses and, possibly, dispersion. The coefficients b and c are used to control the location of zeros of the structure: while the poles are independent of source and listener positions, the zeros are not. The feedback matrix is chosen to be circulant because of the good properties of the resulting recursive structure. Such a structure is called

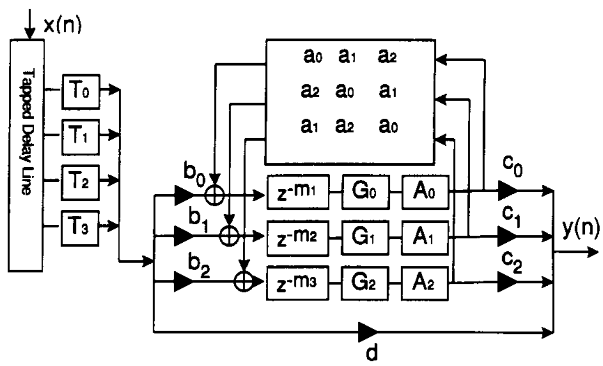

Figure 5. Computational structure of the BaBo model.

https://doi.org/10.1017/S1355771898009182 Published online by Cambridge University Press

a feedback delay network (Rocchesso 1997, Rocchesso and Smith 1997).

The physical representation offered by the BaBo model, even though simple, allows the determination of the delay lengths based on physical distances and not on number-theoretic considerations, as was done with classical structures. Moreover, the object properties have a close correspondence with the coefficients of the feedback matrix, thus allowing a tuning of the computational structure based on the diffusive properties we intend to obtain.

5. PHYSICALLY BASED MODELLING FOR JOINT PRESENTATION OF SOUND AND ANIMATED IMAGES

After an extensive presentation of physically based sound models, it is worth pointing to some peculiar applications which have already been developed, and which show the strength of the physics-based approach for integrating sounds into multimedia applications. One of the most natural applications of the techniques illustrated in section 3 is in the field of joint scientific visualisation and auralisation. Since sounds are produced as a side effect of system simulations, it is just as easy to produce images which can be of help for understanding complex body dynamics or wave propagation phenomena. For example, two-dimensional wave propagation and the interaction with an exciter have been clearly visualised by means of waveguide meshes (Van Duyne and Smith 1993a, b, 1996, Fontana and Rocchesso 1997). It has been commonly acknowledged that the effect of numerical artefacts is better appreciated when both visual and aural demonstrations are given. Three-dimensional waveguide or finite-difference meshes have been used for visualising wave propagation in rooms (Savioja, Rinne and Takala 1995), where effects such as diffraction introduced by walls or pillars are not always obvious to the architect. Other notable examples of three-dimensional wave visualisation are the simulation of wave propagation in a bent tube (see Van Duyne and Smith (1996) for snapshots of an animation by Stilson) or in arbitrary three-dimensional shapes (Rossiter, Horner and Baciu 1996). Researchers at the University of Illinois have experimented with the simulation of a chaotic nonlinear system (Bargar, Choi, Das and Goudeseune 1994) in their CAVE virtual environment (De Fanti, Cruz-Neira, Sandin, Kenyon and Hart 1992). They applied control and navigation techniques to chaos for generating meaningful auditory and visual cues, thus providing a multi-sensorial and immersive exploration of complexity.

Other examples of the integration of physically based techniques in an audiovisual context are found in the works of Cook (1995) and Hanninen *et al.* (Hanninen, Savioja and Takala 1996, Takala, Hanninen, Valimaki, Savioja, Huopaniemi, Huotilainen and Karjalainen 1996), where a virtual musical instrument is rendered in both the visual and aural domains. In particular, the second of these examples shows a full integration of physics-based animation, sound generation and sound processing: an animated musician plays a physically modelled flute in a fully synthetic acoustically modelled concert hall. Instrument fingerings are controllable in real time among a set of finger positions precomputed by inverse kinematics, so that gestures are synchronised with sound generation.

In the field of artistic animation, we would like to point out the work by Cadoz, Luciani and Florens (1994), who used the CORDIS-ANIMA system for joint production of sound and animated images. For example, a complex clockwork mechanism was simulated, and the visual motion was synchronised with a set of sounds obtained by the same computational structures.

Only a few experiences are reported in the physically based integration of environmental sounds into virtual environments or animation. In Munteanu, Guggiana, Darvishi, Schauer, Rauterberg and Motavalli (1995) the spectral properties of simple interactions, such as collisions between spheres and plates, are analysed by exploitation of physical properties of the objects, and the parameters extracted from analysis are used to drive signal-based sound synthesis techniques, such as additive or subtractive synthesis. A similar approach is also followed in van den Doel and Pai (1996), where a vibrating object, such as a plate, is divided into small areas, each of them with characteristic weights of the vibrational modes. In other words, while the same resonances are maintained all over the object, the system zeros vary with the excitation point. These approaches, which use physics in the analysis phase and nonphysical methods in the synthesis phase, are attractive for their efficiency but are not capable of rendering the sound of sustained excitations (e.g. friction) or multiple strokes in rapid sequence (like in a musical *ribattuto*).

A promising direction for the integration of sound and motion was indicated by Takala and Hahn in their seminal paper (Takala and Hahn 1992), where they defined *sound rendering* as 'the process of forming the composite soundtrack from component sound objects'. They outlined two parallel pipelines for image and sound rendering, and proposed to synthesise sound objects from physical principles in a way similar to motion synthesis. The Hahn's research group further investigated the relationship between sound and motion by introducing *timbre trees*, an acoustic analogy of shade trees, which are well known to the computer graphics community (Takala *et al.* 1993, Hahn, Geigel, Lee, Gritz, Takala and

Mishra 1995). These contributions provided good solutions to the problems of sound–motion synchronization and automatic generation of background music, although they relied heavily on heuristics due to difficulties in accurate physical modelling.

6. CONCLUSION

With respect to static sampled sounds, parametric models give an evident advantage in all the applications where interactivity is a major concern, particularly with respect to static sounds. The new paradigm of physics-based models, in their physical or pseudo-physical interpretation, can provide new strength to audiovisual communication. We have indicated some of the relationships occurring between sound and image models, in both the synthesis and processing phases, in the hope that new efforts will be put into trying to build a tight coupling between images and sounds in multimedia environments.

REFERENCES

Ballan, O., Mozzoni, L., and Rocchesso, D. 1994. Sound spatialization in real time by first-reflection simulation. *Proc. Int. Computer Music Conf.*, pp. 475, 6. ICMA, Aarhus, Denmark.

Bargar, R., Choi, I., Das, S., and Goudeseune, C. 1994. Model-based interactive sound for an immersive virtual environment. *Proc. Int. Computer Music Conf.*, pp. 471–4. ICMA, Aarhus, Denmark.

Barzel, R. 1992. *Physically-based Modeling for Computer Graphics: a Structured Approach*. San Diego, CA: Academic Press.

Begault, D. R. 1994. *3-D Sound for Virtual Reality and Multimedia*. Boston, MA: Academic Press.

Begault, D. R., and Erbe, T. 1994. Multichannel spatial auditory display for speech communications. *Journal of the Audio Engineering Society* **42**(10): 819–26.

Belevitch, V. 1968. *Classical Network Theory*. San Francisco: Holden-Day.

Blauert, J. 1983. *Spatial Hearing: the Psychophysics of Human Sound Localization*. Cambridge, MA: MIT Press.

Blinn, J. F., and Newell, M. E. 1976. Texture and reflection in computer generated images. *Communications of the ACM* **19**(10): 542–7.

Bloom, P. J. 1985. High-quality digital audio in the entertainment industry: an overview of achievements and challenges. *IEEE Acoustics, Speech and Signal Processing Magazine* **2**: 2–25.

Borin, G., and De Poli, G., 1996. A hysteretic hammer–string interaction model for physical model synthesis. In *Proc. Nordic Acoustical Meeting (NAM–96)*, pp. 399–406. Helsinki, Finland.

Borin, G., De Poli, G., and Rocchesso, D. 1997. Elimination of delay-free loops in discrete-time models of nonlinear acoustic systems. *Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, pp. 12.1.1–4. IEEE, Mohonk, NY.

Borin, G., De Poli, G., and Sarti, A. 1992a. Algorithms and structures for synthesis using physical models. *Computer Music Journal* **16**(4): 30–42.

Borin, G., De Poli, G., and Sarti, A. 1992b. Sound synthesis by dynamic systems interaction. In D. Baggi (ed.) *Readings in Computer-Generated Music*, pp. 139–60. IEEE Computer Society Press.

Borish, J. 1984. Extension of the image model to arbitrary polyhedra. *Journal of the Acoustical Society of America* **75**: 1,827–36.

Burgess, D. A. 1992. Real-time audio spatialization with inexpensive hardware. *GVU Technical Report GIT-GVU-92-20*. Graphics and Usability Center, Georgia Institute of Technology, Atlanta, Georgia.

Cadoz, C., Luciani, A., and Florens, J.-L. 1993. CORDIS-ANIMA: a modeling and simulation system for sound synthesis – the general formalism. *Computer Music Journal* **17**(1): 19–29.

Cadoz, C., Luciani, A., and Florens, J.-L. 1994. Physical models for music and animated image. The use of CORDIS-ANIMA in "ESQUISSES" a music film by ACROE. *Proc. Int. Computer Music Conf.*, pp. 11–18. ICMA, Aarhus, Denmark.

Causse, R., Derogis, P., and Warusfel, O. 1995. Radiation of musical instruments and improvement of the sound diffusion techniques for synthesized, recorded or amplified sounds (revisited). *Proc. Int. Computer Music Conf.*, pp. 359–60. ICMA, Banff, Canada.

Chafe, C. D. 1990. Pulsed noise in self-sustained oscillations of musical instruments. *Proc. Int. Conf. on Acoustics, Speech, and Signal Processing*. IEEE, Albuquerque, NM.

Chaigne, A. 1992. On the use of finite differences for musical synthesis. Application to plucked stringed instruments. *J. Acoustique* **5**(2): 181–211.

Chaigne, A., and Askenfelt, A. 1994. Numerical simulations of piano strings. I. A physical model for a struck string using finite difference methods. *Journal of the Acoustical Society of America* **95**(2): 1,112–18.

Chareyon, J., 1990. Digital synthesis of self-modifying waveforms by means of linear automata. *Computer Music Journal* **14**(4): 25–41.

Chowning, J. M. 1971. The simulation of moving sound sources. *Journal of the Audio Engineering Society* **19**(1): 2–6. Reprinted in *Computer Music Journal*, June 1977.

Cohen, M. F., and Wallace, J. R. 1993. *Radiosity and Realistic Image Synthesis*. Cambridge, MA: Academic Press.

Cook, P. R. 1991. Non-linear periodic prediction for on-line identification of oscillator characteristics in woodwind instruments. *Proc. Int. Computer Music Conf.*, pp. 157–60. ICMA, Montreal, Canada.

Cook, P. R. 1995. Integration of physical modeling for synthesis and animation. *Proc. Int. Computer Music Conf.*, pp. 525–8. ICMA, Banff, Canada.

De Fanti, T., Cruz-Neira, C., Sandin, D., Kenyon, R., and Hart, J. 1992. The CAVE. *Communications of the ACM* **35**(6): 65–72.

De Poli, G. 1991. A tutorial on digital sound synthesis techniques. *Computer Music Journal* **7**(4): 8–26. Reprinted in C. Roads (ed.) *The Music Machine*, pp. 429–47. Cambridge, MA: MIT Press.

https://doi.org/10.1017/S1355771898009182 Published online by Cambridge University Press

De Poli, G., Piccialli, A., and Roads, C. (eds.) 1991. *Representations of Musical Signals*. Cambridge, MA: MIT Press.

Drioli, C., and Rocchesso, D. 1997. A generalized musical-tone generator with application to sound compression and synthesis. *Proc. Int. Conf. on Acoustics, Speech, and Signal Processing*. IEEE, Munchen, Germany.

Fettweis, A. 1986. Wave digital filters: theory and practice. *Proc. of the IEEE* **74**(2): 270–327.

Florens, J.-L., and Cadoz, C. 1991. The physical model: modeling and simulating the instrumental universe. In G. De Poli, A. Piccialli and C. Roads (eds.) *Representations of Musical Signals*, pp. 227–68. Cambridge, MA: MIT Press.

Fontana, F., and Rocchesso, D. 1998. Physical modeling of membranes for percussion instruments. *Acustica* **84**(3): 529–41.

Gardner, B., and Martin, K. 1994. HRTF measurements of a KEMAR dummy-head microphone. *Technical Report #280*. MIT Media Lab, Boston, MA.

Gazengel, B., Gilbert, J., and Amir, N. 1995. Time domain simulation of single reed wind instrument. From the measured input impedance to the synthesis signal. Where are the traps? *Acta Acustica* **3**: 445–72.

Hahn, J. K., Geigel, J., Lee, J. W., Gritz, L., Takala, T., and Mishra, S. 1995. An integrated approach to motion and sound. *Journal of Visualization and Computer Animation* **6**(2): 109–23.

Hanninen, R., Savioja, L., and Takala, T. 1996. Virtual concert performance – synthetic animated musicians playing in an acoustically simulated room. *Proc. Int. Computer Music Conf.*, pp. 402–4. ICMA, Hong Kong.

Haus, G. (ed.) 1993. *Music Processing*. Madison, WI: A-R Editions Inc.

Hendrix, C., and Barfield, W. 1996. The sense of presence within auditory virtual environments. *Presence: Teleoperators and Virtual Environments* **5**(3): 290–301.

Incerti, E., and Cadoz, C. 1995. Topology, geometry, matter of vibrating structures simulated with CORDIS-ANIMA. Sound synthesis methods. *Proc. Int. Computer Music Conf.*, pp. 96–103. ICMA, Banff, Canada.

Jaffe, D. A., and Smith, J. O. 1983. Extensions of the Karplus–Strong plucked string algorithm. *Computer Music Journal* **7**(2): 56–69.

Jot, J.-M., and Warusfel, O. 1995. A real-time spatial sound processor for music and virtual reality applications. *Proc. Int. Computer Music Conf.*, pp. 294–5. ICMA, Banff, Canada.

Karplus, K., and Strong, A. 1983. Digital synthesis of plucked string and drum timbres. *Computer Music Journal* **7**(2): 43–55.

Kendall, G. S. 1991. Visualization by ear: auditory imagery for scientific visualization and virtual reality. *Computer Music Journal* **15**(4): 70–3.

Kendall, G. S. 1995a. A 3-D sound primer: directional hearing and stereo reproduction. *Computer Music Journal* **19**(4): 23–46.

Kendall, G. S. 1995b. The decorrelation of audio signals and its impact on spatial imagery. *Computer Music Journal* **19**(4): 71–87.

Kistler, D. J., and Wightman, F. 1992. A model of head-related transfer functions based on principal components analysis and minimum-phase reconstruction. *Journal of the Acoustical Society of America* **91**: 1,637–47.

Kuttruff, H. 1991. *Room Acoustics*, 3rd edn (1st edn 1973). Essex, England: Elsevier Science.

Laakso, T. I., Valimaki, V., Karjalainen, M., and Laine, U. K. 1996. Splitting the unit delay – tools for fractional delay filter design. *IEEE Signal Processing Magazine* **13**(1): 30–60.

MacIntyre, B., and Feiner, S. 1996. Future multimedia user interfaces. *Multimedia Systems* **4**: 250–68.

Madhyastha, T. M., and Reed, D. R. 1995. Data sonification: do you see what I hear? *IEEE Software* **12**(2): 45–56.

McAdams, S. 1987. Music: a science of mind? *Contemporary Music Review* **2**(1): 1–61.

McIntyre, M. E., Schumacher, R. T., and Woodhouse, J. 1983. On the oscillations of musical instruments. *Journal of the Acoustical Society of America* **74**(5): 1,325–45.

Meerkotter, K., and Scholtz, R. 1989. Digital simulation of nonlinear circuits by wave digital filter principles. *Proc. Int. Symp. on Circuits and Systems*, pp. 720–3 (IEEE).

Metlay, M. 1996. Technics WSA1 acoustic modeling synthesizer. *Computer Music Journal* **20**(3): 88–94.

Moore, F. R. 1990. *Elements of Computer Music*. Englewood Cliffs, NJ: Prentice-Hall.

Moorer, J. A. 1972. Music and computer composition. *Communications of the ACM* **15**(2): 104–13.

Moorer, J. A. 1977. Signal processing aspects of computer music. *Proc. of the IEEE* **65**(8): 1,108–37.

Moorer, J. A. 1979. About this reverberation business. *Computer Music Journal* **3**(2): 13–18.

Moorer, J. A. 1983. The manifold joys of conformal mapping: applications to digital filtering in the studio. *Journal of the Audio Engineering Society* **31**(11): 826–40.

Morse, P. M. 1991. *Vibration and Sound* (1st edn 1936, 2nd edn 1948). New York: American Institute of Physics for the Acoustical Society of America.

Munteanu, E., Guggiana, V., Darvishi, A., Schauer, H., Rauterberg, M., and Motavalli, M. 1995. Physical modelling of environmental sounds. *Proc. Second Int. Conf. on Acoustics and Musical Research*, pp. 107–12. AIA-CNR, Ferrara, Italy.

Oppenheim, A. V., and Schafer, R. W. 1989. *Discrete-Time Signal Processing*. Englewood Cliffs, NJ: Prentice-Hall Inc.

Orfanidis, S. J. 1996. *Introduction to Signal Processing*. Englewood Cliffs, NJ: Prentice-Hall Inc.

Poggio, T., and Girosi, F. 1990. Networks for approximation and learning. *Proc. of the IEEE* **79**(9): 1,481–97.

Press, W. H., Flannery, B. P., Teukolsky, S. A., and Vetterling, W. T. 1988. *Numerical Recipes in C*. Cambridge: Cambridge University Press.

Rich, C., Waters, R. C., Strohecker, C., Schabes, Y., Freeman, W. T., Torrence, M. C., Goldin, A. R., and Roth, M. 1994a. An animated on-line community with artificial agents. *IEEE Multimedia* **1**(1): 32–42.

Rich, C., Waters, R. C., Strohecker, C., Schabes, Y., Freeman, W. T., Torrence, M. C., Goldin, A. R., and Roth, M. 1994b. Demonstration of an interactive multimedia environment. *IEEE Computer* **27**(12): 15–22.

Risset, J.-C. 1969. An introductory catalog of computer-synthesized sounds. *Technical Report*. Bell Laboratories, Murray Hill, NJ.

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

Roads, C., Pope, S., Piccialli, A., and De Poli, G. (eds.) 1997. *Musical Signal Processing*. Lisse, the Netherlands: Swets & Zeitlinger.

Rocchesso, D. 1993. Modelli generalizzati di strumenti musicali per la sintesi del suono. *Rivista Italiana di Acustica* **17**(4): 61–71.

Rocchesso, D. 1995. The Ball within the Box: a sound-processing metaphor. *Computer Music Journal* **19**(4): 47–57.

Rocchesso, D. 1997. Maximally-diffusive yet efficient feedback delay networks for artificial reverberation. *IEEE Signal Processing Letters* **4**: 252–5.

Rocchesso, D., and Scalcon, F. 1996. Accurate dispersion simulation for piano strings. *Proc. Nordic Acoustical Meeting (NAM-96)*, pp. 407–14. Hensinki, Finland.

Rocchesso, D., and Smith, J. O. 1997. Circulant and elliptic feedback delay networks for artificial reverberation. *IEEE Trans. on Speech and Audio Processing* **5**(1): 51–63.

Rocchesso, D., and Turra, F. 1993. A generalized excitation for real-time sound synthesis by physical models. *Proc. of the Stockholm Music Acoustics Conf.*, pp. 584–8. Royal Swedish Academy of Music, Stockholm.

Rodet, X. 1993. Models of musical instruments from Chua's circuit with time delay. *IEEE Trans. on Circuits and Systems-II* **40**(10): 696–701.

Rodet, X. 1994. Stability/instability of periodic solutions and chaos in physical models of musical instruments. *Proc. Int. Computer Music Conf.*, pp. 386–93. Computer Music Association, Aarhus, Denmark.

Rossiter, D., Horner, A., and Baciu, G. 1996. Visualization and manipulation of 3D digital waveguide structures for sound experimentation. *Proc. Int. Computer Music Conf.*, pp. 43–6. ICMA, Hong Kong.

Sarti, A., and De Poli, G. 1996. Generalized adaptors with memory for nonlinear wave digital structures. *Proc. EUSIPCO*, pp. 1,941–4. Trieste, Italy.

Savioja, L., Rinne, T. J., and Takala, T. 1994. Simulation of room acoustics with a 3-D finite difference mesh. *Proc. Int. Computer Music Conf.*, pp.463–6. ICMA, Aarhus, Denmark.

Savioja, L., and Valimaki, V. 1997. Improved discrete-time modeling of multi-dimensional wave propagation using the interpolated digital waveguide mesh. *Proc. Int. Conf. on Acoustics, Speech, and Signal Processing*. IEEE, Munich, Germany.

Scavone, G. P. 1994. Combined linear and non-linear periodic prediction in calibrating models of musical instruments to recordings. *Proc. Int. Computer Music Conf.*, pp. 433–4. ICMA, Aarhus, Denmark.

Schaeffer, P. 1966. *Traite des objets musicaux*. Paris: Editions du Seuil.

Schafer, M. R. 1977. *The Tuning of the World*. Toronto: McClelland and Stewart Limited.

Schroeder, M. R. 1962. Natural-sounding artificial reverberation. *Journal of the Audio Engineering Society* **10**: 219–33.

Schroeder, M. R. 1973. Computer models for concert hall acoustics. *American Journal of Physics* **41**: 461–71.

Smith, J. O. 1991. Viewpoints on the history of digital synthesis. *Proc. Int. Computer Music Conf.*, pp. 1–100. ICMA, Montreal, Canada.

Smith, J. O. 1992. Physical modeling using digital waveguides. *Computer Music Journal* **16**(4): 74–91.

Smith, J. O. 1993. Efficient synthesis of stringed musical instruments. *Proc. Int. Computer Music Conf.*, pp. 64–71. ICMA, Tokyo, Japan.

Spezzano, G., Talia, D., Gregorio, S. D., Rongo, R., and Spataro, W. 1996. A parallel cellular tool for interactive modeling and simulation. *IEEE Computational Science and Engineering* **3**(3): 33–43.

Takala, T., and Hahn, J. 1992. Sound rendering. *Proc. SIGGRAPH* **26**(2): 220–1. ACM Computer Graphics.

Takala, T., Hahn, J., Gritz, L., Geigel, J., and Lee, J. W. 1993. Using physically-based models and genetic algorithms for functional composition of sound signals, synchronized to animated motion. *Proc. Int. Computer Music Conf.*, pp. 180–3. ICMA, Tokyo, Japan.

Takala, T., Hanninen, R., Valimaki, V., Savioja, L., Huopaniemi, J., Huotilainen, T., and Karjalainen, M. 1996. An integrated system for virtual audio reality. *Audio Engineering Society Convention*. AES, Copenhagen, Denmark.

Truax, B. (ed.) 1978. *Handbook for Acoustic Ecology*. Vancouver: ARC Publications.

Valimaki, V., Houpaniemi, J., Karjalainen, M., and Janosy, Z. 1996. Physical modeling of plucked string instruments with application to real-time sound synthesis. *Journal of the Audio Engineering Society* **44**(5): 331–53.

van den Doel, K., and Pai, D. K. 1996. Synthesis of shape dependent sounds with physical modeling. *Proc. Int. Conf. on Auditory Display (ICAD-96)*. Palo Alto, CA.

Van Duyne, S. A., Pierce, J. R., and Smith, J. O. 1994. Traveling wave implementation of a lossless mode-coupling filter and the wave digital hammer. *Proc. Int. Computer Music Conf.*, pp. 411–18. ICMA, Aarhus, Denmark.

Van Duyne, S. A., and Smith, J. O. 1993a. Physical modeling with the 2-D digital waveguide mesh. *Proc. Int. Computer Music Conf.*, pp. 40–7. ICMA, Tokyo, Japan.

Van Duyne, S. A., and Smith, J. O. 1993b. The 2-D digital waveguide mesh. *Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*. Mohonk, NY.

Van Duyne, S. A., and Smith, J. O. 1995. The tetrahedral waveguide mesh: multiply-free computation of wave propagation in free space. *Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*. Mohonk, NY.

Van Duyne, S. A., and Smith, J. O. 1996. The 3D tetrahedral digital waveguide mesh with musical applications. *Proc. Int. Computer Music Conf.*, pp. 9–16. ICMA, Hong Kong.

van Walstijn, M., and Dubbelboer, F. 1997. Discrete modeling of the harmonic oscillator. *Technical Report*. Unpublished manuscript.

Vercoe, B. L., Gardner, W. G., and Scheirer, E. D. 1998. Structured audio: creation, transmission, and rendering of parametric sound representations. *Proc. of the IEEE* **86**(5): 922–40.

von Neumann, J. 1966. *Theory of Self-Reproducing Automata*. Urbana: University of Illinois Press.

Warusfel, O., Kahle, E., and Jullien, J. P. 1993. Relationships between objective measurements and perceptual interpretation: the need for considering spatial emission of sound sources. *Journal of the Acoustical Society of America* **93**: 2,281–2.

Wenzel, E. M. 1992. Localization in virtual acoustic displays. *Presence: Teleoperators and Virtual Environments* **1**(1): 80–107.

Wightman, F. L., and Kistler, D. J. 1989a. Headphone simulation of free-field listening. I: Stimulus synthesis. *Journal of the Acoustical Society of America* **85**(2): 858–67.

Wightman, F. L., and Kistler, D. J. 1989b. Headphone simulation of free-field listening. II: Psychophysical validation. *Journal of the Acoustical Society of America* **85**: 868–78.

Wolfram, S. 1984. Computational theory of cellular automata. *Communications in Mathematical Physics* **96**: 15–57.

Yamaha Corp. 1994. Yamaha physical modeling synthesizers. *Computer Music Journal* **18**(3): 101.