TUTORIAL ARTICLE

Virtual musical instruments – natural sound using physical models*

VESA VÄLIMÄKI[1] and TAPIO TAKALA[2]

[1]Helsinki University of Technology, Laboratory of Acoustics and Audio Signal Processing, Otakaari 5A, FIN-02150 Espoo, Finland
[2]Helsinki University of Technology, Department of Computer Science, Otakaari 1, FIN-02150 Espoo, Finland

Physical modelling of musical instruments is an exciting new paradigm in digital sound synthesis. The basic idea is to imitate the sound production mechanism of an acoustic musical instrument using a computer program. The sound produced by such a model will automatically resemble that of the real instrument, if the model has been devised in a proper way. In this article we review the history and present techniques of physical modelling. It appears that the many seemingly very different modelling methods try to achieve the same result: to simulate the solutions of the wave equation in a simplified manner. We concentrate on the digital waveguide modelling technique which has gained much popularity among both researchers and engineers in the music technology industry. The benefits and drawbacks of the new technology are considered, and concurrent research topics are discussed. The physical modelling approach offers many new applications, especially in the fields of multimedia and virtual reality.

1. INTRODUCTION

Physical modelling of musical instruments is a new approach to sound synthesis using computers. It is based on the idea that when the vibrating structure is simulated in exactly the right way, the sound produced by that model is identical with the sound of the corresponding physical object. Thus, physical modelling of musical instruments simply means that the physical structure of a musical instrument is being modelled with mathematical and physical formulae and these formulae are realised with a computer.

While being run, this computer program generates numbers (samples) that can be processed in a way familiar to the computer music community: the sequence of numbers (a digital signal) is fed into a digital-to-analogue (D/A) converter, which changes every number to a corresponding electrical voltage level. This time-varying voltage can be amplified and listened to through loudspeakers or headphones. All the phases of this synthesis process, except for the initial creation of the number sequence, are identical to what has been done since the early days of computer music. The only difference between this new digital synthesis technique and all the earlier ones is thus the method that is used for producing the digital sound signal.

A fundamental difference between the physical modelling approach and other synthesis techniques is that the former tries to imitate the properties of the sound source (type of excitation and resonator, resonances of a soundboard, etc.), while the latter focus on the properties of the sound signal heard by the listener (waveform, spectrum, etc.). It is astonishing that none of the earlier music synthesis techniques really considered the sound source, whereas in the field of speech synthesis, the simulation of the sound source (the vocal tract resonances and the vibration of the vocal cords) has been a popular technique from the beginning of the 1960s.

The main advantages of physical modelling synthesis are that the parameters of the technique are physically meaningful, such as the blowing pressure in wind instruments, and important parts of the evolution of single tones (e.g. the attack and decay) are generated automatically in a correct way.

The physical modelling approach is reaching a mature phase. The first commercial products based on a physical model have been released recently, including the Yamaha VL-1, Korg Wavedrum, and Roland VG-8 guitar processor. Of these, only the VL-1 is a purely physical model. The others use sound processing algorithms that are motivated by the principles of physical modelling.

2. OVERVIEW OF DIGITAL SYNTHESIS TECHNIQUES

Before discussing the details of physical modelling, we now review other sound synthesis techniques. This will be beneficial in understanding some of the recent modelling algorithms. Later on we will see that many of the principles used in traditional sound synthesis also play an essential role in physical models.

2.1. Traditional methods

2.1.1. Linear and nonlinear techniques

In earlier days, sound synthesis techniques were usually divided into two categories according to lin-

* The authors are grateful to Professor Matti Karjalainen for his comments on the manuscript and many fruitful discussions.

earity. An acoustic system is said to be linear if (i) the sum of outputs produced by two different input signals is the same as the output signal produced when the sum of these two signals has been used as an input, and (ii) the amplification of the input signal by some factor causes the output to be scaled by the same factor. These are known as the principles of superposition and homogeneity, respectively. All the techniques that do not fulfil these principles are called nonlinear. Some basic features of nonlinear systems are that the output signal may contain other frequencies than those present in the input and that the spectral content of the output signal depends on the amplitude of the input signal.

The class of linear techniques includes additive and subtractive synthesis. *Additive synthesis* is one of the oldest techniques for sound synthesis, yet only recently has it emerged as a tool that can be easily used for practical work. Additive synthesis basically means adding a number of sine waves to construct a desired spectrum. The same principle has been used in electric organs. An advantageous feature of additive synthesis is the availability of a corresponding analysis technique, the *Fourier transform*, which takes as an input a signal waveform and fits a number of sine and cosine waves into the data. As a result the Fourier transform produces a set of numbers that tell us how much and in what phase each of the sine and cosine components contribute to the input signal. These data can be used as parameters for additive synthesis. The parameters to be controlled are the frequency, amplitude and phase of each sine wave.

The Fourier transform can analyse both harmonic and nonharmonic signals. What is even better is that there exists an efficient way to compute the Fourier transform using a computer – the fast Fourier transform (FFT) algorithm. This makes the analysis stage much faster, but, even more importantly, additive synthesis can be realised very efficiently by first defining the spectrum of the signal and then converting this spectrum into a time waveform using the inverse FFT algorithm. Rodet and Depalle (1992) have recently shown that it is possible to radically reduce the amount of data in the spectral representation and yet produce high-quality musical tones via the inverse FFT.

Subtractive synthesis refers to linear filtering of an input signal so that some frequency components are attenuated (subtracted) and some emphasised. A better term is *source–filter model* since the basic idea of subtractive synthesis is that sound production is divided into two parts: an excitation signal (source) and a resonator (filter). The excitation is typically a spectrally rich waveform, such as noise or some periodic pulseform whose spectrum has energy at several frequencies. This is helpful when we realise that the filter is a linear system and the output signal cannot include any other frequencies than those included in the input signal. The filter not only attenuates some of the frequency components of the input signal but may also boost some of them. This can be used to imitate formants which are resonances that characterise the timbre of an instrument. In analog synthesizers this principle was extremely popular.

The nonlinear synthesis techniques are commonly called modulation techniques since they usually employ multiplication of two signals. The *FM synthesis* technique (Chowning 1973) is perhaps the best known representative of nonlinear techniques. It is based on the principle of frequency modulation where one oscillator (the modulator) controls the frequency of another one (the carrier). The Yamaha DX-7 released in 1983 was the first product to use the FM principle and also the first fully digital synthesizer. It represented a breakthrough for digital sound synthesis.

Another nonlinear synthesis technique is *waveshaping* (Arfib 1979, Le Brun 1979). In this method the idea is to feed a signal (typically a sine wave) into a function that maps the amplitude values in a nonlinear manner. An example of this principle is the clipping caused by overdriving a guitar amplifier. Here the large amplitude values are compressed, while the small values are not much affected. Beauchamp (1979) applied this method to produce brass instrument sounds from pure sine waves. Interestingly enough, waveshaping is nowadays part of some physical models.

2.1.2. Wavetable synthesis

Another very popular synthesis technique is *sampling*, also called *wavetable synthesis*. It simply means recording, processing and playback of sounds. While it can be argued that sampling is not a synthesis technique at all (since it does not create sound from scratch) it must be said that this technique offers an infinite variety of possibilities: any sound (acoustic or synthetic) can be recorded digitally, filtered or edited or combined with other signals, and finally the processed version can be listened to. The processing can be so violent that the original sampled sound may not be recognisable, although it still affects the result. The reader should not be too surprised to learn that this principle of processing recorded sounds is also a useful tool for physical modelling.

In practice samplers are distinguished from other wavetable synthesizers in the way the original samples are generated and played back. While in samplers the playable signal is recorded in its whole length, in wavetable synthesizers the stored waveform is repeated in order to have a periodic tone. The original samples are synthesized with some other technique, such as additive or subtractive synthesis, or even by drawing the waveform by hand using a graphical interface. Unfortunately, the visual appearance of the waveform has very little to do with the audible timbre. The essential feature in periodic signals is their spectral content rather than the variation of the signal waveform over time. It is also known that the ear is not sensitive to the phase of a signal. This implies that we can have signals with different waveforms but with the same frequency spectrum and yet have similar timbre.

2.2. Recent developments

The old categorisation of synthesis techniques according to linearity is no longer particularly useful. There are several new approaches that are not essentially linear or nonlinear but which have some other property or principle that is of interest. Furthermore, in some advanced algorithms linear and nonlinear techniques appear as components.

Smith (1991) suggested another classification for sound synthesis techniques into four categories: (i) abstract algorithms, (ii) processing of recorded samples, (iii) spectral models, and (iv) physical models. The first class includes techniques that produce sound based on some mathematical formula which has no direct relation to real-world sounds or acoustic principles of sound generation. Thus it is difficult to predict what kind of sound is produced by a particular algorithm, or to design an algorithm that results in a particular sound. The traditional nonlinear techniques such as FM synthesis and waveshaping are well-known representatives of this class. The second class, processing of recorded samples, includes, for example, wavetable synthesis.

Spectral and physical models have recently gained more interest than the other approaches. For this reason we discuss them in more detail.

2.2.1. Spectral modelling

Spectral models concentrate on the frequency-domain properties of sound waves at the ear drum of the listener, taking into account how sounds are psychologically perceived. This class includes traditional linear techniques, such as basic additive and subtractive synthesis, but recently several new approaches have been developed. In particular, these methods utilise the fact that interesting timbres are not stationary but are characterized by the evolution of their spectral components.

The major problem with the implementation of additive synthesis has been the huge numbers of control parameters (frequency, amplitude, and phase of each sine wave as a function of time). The first attempts at reducing the quantity of data were to neglect the phase and to determine linear trajectories for the amplitude and frequency of every sine wave. However, only recent advances by McAuley and Quatieri (1986), Rodet and Depalle (1992), and Serra and Smith (1991) have expanded this technique into a useful one. As is often the case in the history of computer music, an important technique was first applied to processing speech rather than musical signals: McAuley and Quatieri (1986) showed that a speech signal can be represented with a number of sine waves whilst still retaining a high quality. The amplitudes and frequencies must vary with time according to trajectories measured from real signals. The MQ algorithm utilises the concepts of death and birth in order to know when to drop one sine component and to replace it with another one of a different frequency.

Serra and Smith (1991) expanded additive synthesis into a *sines-plus-noise model*, where the harmonics and other prominent frequency peaks of a sound are imitated by sine waves, and the rest of the sound (the residual) is modelled as noise. This model helps considerably in music synthesis where sounds can be very noisy. Resynthesis of a noisy musical tone was one of the weakest points of traditional additive synthesis, since a huge number of sine waves was required to reproduce broadband noise. In the sines-plus-noise model, the noise component is created by filtering white noise (a random signal with a flat spectrum) with a time-varying filter.

3. PRINCIPLES OF PHYSICAL MODELLING

The roots of physical modelling lie in the history of mathematics and physics. Many of the principles used in physical models today were first discovered in the eighteenth century by mathematicians such as D'Alembert, Euler, Bernoulli, and Lagrange. The single most important development was the *wave equation*, also known as Helmholtz' equation. It describes how waves in general, including mechanical vibrations, propagate in a homogeneous medium. In this section we show how the equation is derived from a simple physical system, and survey methods for solving the equation. As will be seen, the discrete formulation used in the derivation appears to be an efficient solution algorithm, too.

3.1. Wave transmission in a physical system

To understand how vibrations behave, let us observe a simple mechanical system consisting of a chain of equal point masses connected with massless springs (figure 1). At rest all the masses are equidistant. Assume that the first mass is forced to move a small amount to the right. The first spring is now compressed and tries to return to its rest length by exerting a force at both ends proportinal to the con-

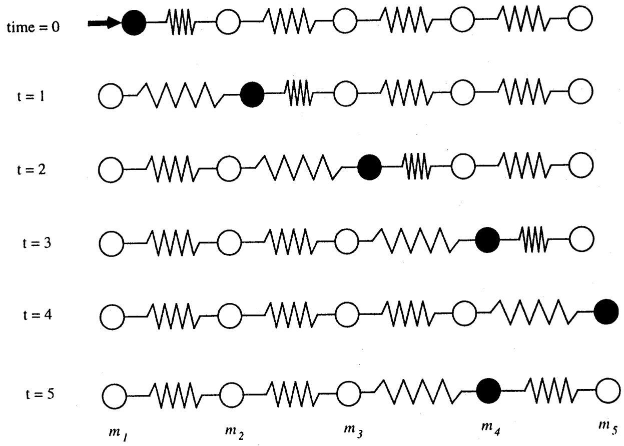

Figure 1. Propagation of displacement in a chain of masses and springs.

traction of the spring. By Newton's second law, this force causes the second mass to accelerate to the right, and similarly the first mass to accelerate to the left. After some time the first mass has returned to its rest position, but the second mass is now displaced. Meanwhile, the second spring has been compressed and produces a force on and acceleration of the third mass. In this way the initial contraction impulse gradually propagates as a wave along the chain. The propagation speed is determined by the material properties: stiffer springs make the wave move faster, whereas larger masses moderate its speed due to inertia.

What happens as the wave comes to the end of the chain depends on how the last mass is constrained. If it is free to move, it will stretch the last spring (as in figure 1), which will cause a strain wave to propagate to the left along the chain. Thus the original contraction wave has been reflected at the end, and its phase inverted. If the last mass were fixed, it would not release the contraction of the last spring, which would then reflect the compression wave back and retain its original phase. A similar reflection occurs later at the other end and a wave, once started, will move back and forth along the chain.

If we move a mass in the middle of the chain, two waves will emanate in opposite directions, each reflecting at the end and returning towards the centre. Due to the superposition principle of linear systems, they will not disturb each other while meeting, but simply sum temporarily and continue their motion. Instead of a single impulse, an arbitrary waveform can be generated by forcing a mass to move as a function of time. A sinusoidal force with period equal to the two-way travelling time of the wave in the chain will produce a *standing wave*: the two opposite travelling waves sum such that they alternately reinforce and cancel each other – seemingly there is no wavefront moving along the chain, but only strains and compressions alternating in time. The same phenomenon also happens at any other frequency which is an integral multiple of the basic resonance.

3.2. The wave equation

Exact physical analysis of our example system requires writing down its equations of motion. This derivation is rather mathematical, so a non-interested reader may want to skip to the solutions decribed in the next section.

Referring to figure 2, let us use p_i to denote the displacement of the mass m_i from its rest position along the line. The spring to the left is compressed if its left end is more displaced than its right end. It acts on m_i with a force $F_{\text{left}} = k\Delta p_i = k(p_{i-1} - p_i)$ to the right where k is the spring stiffness constant which relates the force exerted to the degree of compression of the spring. Similarly, the spring to the right produces a force $F_{\text{right}} = k\Delta p_{i+1} = k(p_i - p_{i+1})$, but in the opposite direction. Thus the total force on m_i is $F_i = k(\Delta p_i - \Delta p_{i+1}) = k(p_{i-1} + p_{i+1} - 2p_i)$. This can be written in abbreviated form as $F_i = k\Delta^2 p_i$, indicating that we have differentiated twice (taking differences of differences in p).

By Newton's second law, $F = ma$, and so the acceleration caused by the force on the mass is $a_i = F_i/m = (k/m)\Delta^2 p_i$ (as all masses are equal, we simply use m for m_i). During a short time interval Δt, this updates the velocity of the mass as $v_i(t + \Delta t) = v_i(t) + a_i\Delta t$, which updates its position as $p_i(t + \Delta t) = p_i(t) + v_i\Delta t$. Although these approximate formulae can be used for simulation, a method first put forward by Euler, the discrete time step causes errors which have to be specially compensated for. Smaller steps will increase accuracy, to a limit which results in infinitesimal intervals and mathematically exact integral formulae.

Now if we wish to approximate a continuous material we can scale down the system and use smaller masses with denser spacing and shorter springs in between. In this case it is better to use relative measures, i.e. displacements per unit length, and to indicate position by a continuous variable s instead

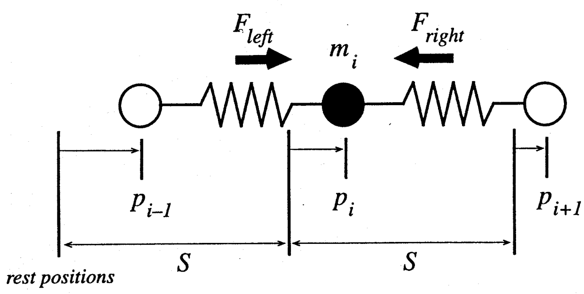

Figure 2. Forces due to the springs acting on the ith point mass in the chain.

of the integer index i. This means we have to scale each difference by the spacing S of masses along the line, yielding $F_{\text{left}} = kS\Delta p_i/S$, and similarly for F_{right}. In total, the force acting on the mass is $F_i = kS^2\Delta^2 p_i/S^2$, resulting in an acceleration of $a_i = F_i/m = (kS^2/m)\Delta^2 p_i/S^2$. When the spacing S becomes infinitesimally small, we replace finite differences in p by differentials: $\Delta^2 p_i/S^2 \rightarrow \partial s^2$. Also we use the mass density $\rho = m/s$ instead of infinitesimal masses, and the stiffness of the material $\sigma = kS$ instead of infinitesimal springs. These change the first term as $kS^2/m = kS/(m/S) \rightarrow \sigma/\rho$. Further analysis reveals that this constant factor equals the speed of motion of the propagation squared, or $\sigma/\rho = c^2$. Recalling that acceleration is the second derivative of displacement, we arrive at the equation

$$a(s) = \frac{\partial^2 p(s)}{\partial t^2} = \frac{\sigma}{\rho}\frac{\partial^2 p(s)}{\partial s^2} = c^2\frac{\partial^2 p(s)}{\partial s^2}.$$

Although we started the discussion with longitudinal displacement along a one-dimensional (1D) string, the same result holds for transverse displacement as well, except that the speed constant is based on different stiffness measures. Alternatively, we can replace masses with velocity potential and springs with pressure, arriving at a similar equation for the motion of air in a tube. Our 1D equation can also be generalised to a higher-dimensional material by simply taking the spatial derivative with respect to each coordinate axis and summing together. For three dimensions with coordinates x, y and z, the wave equation is

$$\frac{\partial^2 p}{\partial t^2} = c^2\left(\frac{\partial^2 p}{\partial x^2} + \frac{\partial^2 p}{\partial y^2} + \frac{\partial^2 p}{\partial z^2}\right),$$

or more concisely,

$$p_{tt} = c^2(p_{xx} + p_{yy} + p_{zz}),$$

where each subscript means differentiation with respect to the corresponding variable.

3.3. Solutions of the wave equation

If we wish to calculate the vibration of an object, we have to solve the quantity p as a function of time and position from this differential equation. For this we need to know the initial state of the system (for example, displacement and velocity at each point), boundary conditions (i.e. what happens at the end points of the material – usually if they are free or fixed) and possibly the acting external forces (excitation) as a function of time. We also need an algorithm to compute the solution.

In some simple cases an analytical solution can be found. In 1D systems like our example above, it can be shown that the solution is the sum of two wave functions, one moving to the left and the other to the right along the chain. The initial state determines the shape of these waves. At the boundaries the waves reflect, retaining their phase if the end is fixed and inverting if the end is free. They cause standing waves, if their wavelength and the length of the chain are in integer proportions. Following the superposition principle, all possible standing wave motions can be separated into components called *modes*, each of which has a specific frequency and specific locations of nodes. In typical musical instruments with 1D resonators such as strings and tubes, the wavelengths are in almost integer proportions, so that they form a harmonic series.

For 2 and 3D objects, analytic solutions can be found for some simple geometries like rectangular plates, bricks and rooms, circular membranes, or cylindrical rods. For rectangular objects the modal frequencies form separate harmonic series based on each of the dimensions. In other cases they are generally inharmonic, as can be heard in the timbre of bells and drums, for example.

Finite element methods (FEMs) can be used to analyse the vibration of objects or rooms with irregular shapes or inhomogeneous material. The basic idea is to subdivide the object into simple homogeneous elements (for example, instead of point masses and massless springs we would distribute the mass evenly along the springs). Within each element an analytic solution can be formulated, but the boundary conditions tie mating elements together. Thus a set of simultaneous equations is formed, which can be solved numerically for all elements at once. For irregularly shaped but homogeneous material, it is also possible to use *boundary element methods* (BEMs) with somewhat different integral equations that are solved only at the boundary, but determine the inside behaviour of motion also. Both FEMs and BEMs have been used in acoustics to find the modal frequencies of rooms and objects, but they are generally too slow to be used directly for sound signal generation. However, the analysed modes can be used as a physical basis for spectral synthesis.

Approximative methods that simulate the behaviour of a system consisting of discrete mechanical components similar to our spring example, have been used both in computer music to generate timbres of very complicated systems (Cadoz, Luciani and Florens 1984) and in computer animation to visualise soft materials (Terzopoulos, Platt, Barr and Fleischer 1987). The system may consist of any configuration of basic mass and spring components. The simulation starts from an initial state and moves on step-by-step in discrete time intervals (Euler's method). At each step, the forces acting on each mass are first calculated, the corresponding accelerations used to update velocities, and finally each position is updated by velocity multiplied by the time step. This method of finite differences is fairly simple to program, and with the evolution of faster computers it may gradually become feasible even as a real-time synthesis method.

One way to make computation of discrete simulations more efficient is to make the system completely regular with a rectangular array of evenly spaced masses connected by equal springs. In particular, if we choose the time step such that the signal propagates exactly one mesh unit in the time step, the equations for each node reduce to simple summation of the signal values at neighbouring nodes, and appear to be mathematically exact. Moreover, a 1D string becomes a *waveguide* where two signals travel undistorted in opposite directions corresponding to the analytic solution. This can therefore be implemented very efficiently as two separate transmission lines (see figure 4), rather than computing the forces and velocities at each point of a single line. The use of waveguides as a physical model for sound synthesis is discussed thoroughly later in this paper.

The wave equation as such is somewhat idealised because it does not take into account the viscosity of the material that causes dissipation of energy and thus damping of the motion. In our mechanical model we can easily add another term to represent these phenomena: for each mass there is a force resisting its motion relative to its neighbours, with magnitude proportional to the relative velocity and direction opposite to the velocity. The effect is that vibrations of higher frequency, due to their faster motion, tend to get damped and die out sooner. This fact can readily be applied when using results of modal analysis in additive synthesis: each sinusoidal component has a different decay rate that is proportional to the frequency. In a discrete simulation the damping can readily be applied as additional forces when calculating the motion of each point mass. In waveguides, however, we would lose efficiency if we computed the attenuation of the signal at each point. Instead, all losses are collected to one filter placed at the end of the waveguide.

4. A SHORT HISTORY OF PHYSICAL MODELLING

By now it is generally known that while digital synthesis can in principle create any sound, a large majority of these possibilities are musically uninteresting. Thus, a better strategy for the search for new interesting timbres must be to find a technique that is capable of recreating at least some useful sounds, and then generalising these timbres to the unknown timbre spaces with careful modifications. Chowning (1973) suggested in his FM synthesis article that an investigation into the possibilities of his technique should first focus on the simulation of natural timbres. His main argument was that this approach would give knowledge of the fine details that discriminate synthetic sounds from natural ones.

In the case of physical modelling, the initial concentration on natural sounds can be stressed even more strongly, and usually the research is focused on the sounds of acoustic instruments. A fascinating theme is, of course, synthesis of sounds produced by new imagined instruments. A physical modelling synthesizer provides the user with natural ways of creating sounds and thus many of the strange, out-of-this-world sounds generated with a physical model also have certain acoustical properties that make them sound natural.

One example of this is the brightness of a tone: many musical instrument sounds (e.g. those of string and wind instruments) as well as the human voice have a typical low-pass characteristic which is caused by the fact that losses (caused by friction, dissipation of heat, etc.) are stronger at high frequencies. A sound produced by an abstract digital technique, on the other hand, can in principle have a spectrum that gains more energy towards high frequencies, at least up to a certain point. These kinds of spectra are neither what we are used to hearing nor what our auditory system has been built for, and they are therefore regarded as unnatural or unpleasant. When the parameters controlling the amount of losses in a physical model are not changed radically from their ‘correct’ values, the resulting sounds will have a familiar low-pass characteristic. Other features of natural sounds can be retained in a similar fashion when modifying the model. Physical modelling can offer a huge range of possibilities in the world of nonexisting sounds that have a familiar ‘acoustic’ identity.

In physical modelling, a musical instrument is typically divided into parts with respect to functional properties. Commonly the model consists of three parts: (i) the excitation mechanism, (ii) the resonator, and (iii) the radiator. In many instruments the excitation and resonator are connected with a nonlinear feedback. Each of the three parts is modelled separately. This provides the flexibility to combine parts freely in a modular way and to build models of different instruments.

An important step in the development of a synthesizer based on a physical model is to reduce the model as much as possible in order to be able to compute the output signals in real time with an affordable computer system. In this reduction one has to take into account the target of the produced sound–the human ear. Although the basic idea of physical modelling is to simulate the details of some sound production mechanism, it is of no value to concentrate on details which are not relevant to the timbre. The rule of thumb is that anything that will not have an audible effect on the sound signal can be disregarded.

In general, physical models are computationally more expensive to implement than popular synthesis methods such as FM or sampling synthesis. Fortunately, computers and signal processors are becoming faster so that real-time implementation of more complicated models will become feasible.

4.1. Physical modelling techniques

The methods used for physical modelling can be divided into five categories: (i) source–filter modelling, (ii) numerical solution of partial differential equations, (iii) vibrating mass–spring networks, (iv) modal synthesis, and (v) waveguide synthesis. The source–filter method is included here because it can in some cases be interpreted as a physical modelling technique. An example is a vocal tract model that is used for synthesising speech or singing: the periodic pulses generated by the vocal cords are the source and the vocal tract is the filter (see, for example, Rodet 1984). In some musical instruments it is also easy to separate the source and the filter. In many percussion instruments, for example, the excitation is a short impulse that is not affected by the feedback from the resonator. However, most musical instruments are more complicated systems than just a combination of two subsystems. In wind instruments, for instance, feedback from the resonator to the excitor is required, and in addition the interaction of the excitation mechanism and the feedback signal is nonlinear.

The first attempts to use a physical model for generating musical sounds were made by Hiller and Ruiz (1971). They started with the differential equations that govern the vibrations of a string and approximated this equation with finite differences. This technique is computationally very intensive. Even today real-time sound synthesis with this approach using affordable hardware is out of the question. This line of physical modelling has been continued by Chaigne, Askenfelt and Jansson (1990) (see also Chaigne 1992).

Later in the 1970s another approach to physical modelling was taken in Grenoble, France. A system called CORDIS was developed which simulates a musical instrument as a collection of point masses that have certain elasticity and frictional characteristics (Cadoz *et al*. 1984, Florens and Cadoz 1991). This approach is closely related to the so-called finite element method (FEM) that is used in mechanical engineering to simulate vibration of structures. The object is divided into a large number of pieces in space. Each piece is connected to its neighbours with springs and microdampers. These elements form networks that imitate the vibrating object. At the beginning of the 1980s a special-purpose processor was built which enabled real-time sound synthesis based on physical modelling to be realised for the first time (Cadoz *et al*. 1984).

At IRCAM in Paris a third technique for physical modelling was developed (Adrien 1991). It is called *modal synthesis* and is based on a representation of a vibrating structure as a collection of frequencies and damping coefficients of resonance modes and coordinates that describe the mode shapes. When the instrument is excited at some point, this force excites some or all of the modes. An advantage of modal synthesis is that analysis tools exist which are not too laborious to use. Adrien (1991) points out that modal analysis of a new object, say a violin body, only takes a couple of days. For some simpler structures, such as a string, the model data can be computed in an analytical form.

In modal synthesis, one of problems to be solved is where to take the output of the model. At IRCAM a clever solution was found using the body of a real instrument as the ‘loudspeaker’: in the case of the violin, for example, this means that the simulated vibration of strings is fed to electrical shakers that are attached to a violin body, the strings of which have been carefully damped (Adrien 1991). This approach has the very nice advantage that the radiational properties of the instrument need not be simulated. In virtual reality applications, however, this practical trick is not amenable, since all the vibrating structures must create numerical data to be used by other parts of the virtual reality system.

4.2. Waveguide synthesis

The waveguide synthesis technique will be discussed in more detail than the other techniques because in the first half of this decade it has turned out to be the most important of all the physical modelling methods. This applies to both the academic and commercial communities: the majority of recent advances in the theory of physical modelling have incorporated digital waveguide techniques, and all the commercial products utilise digital waveguides.

The aim in digital waveguide modelling is to design computationally efficient models that behave like physical systems. This technique is especially well suited to the simulation of 1D resonators, such as a vibrating string, a narrow acoustic tube, or a thin bar. The method was, however, first applied to artificial reverberation using delay line networks (Smith 1985) and only after that to the synthesis of wind and string instruments (Smith 1986). The theory of digital waveguide modelling has been primarily developed by Smith (Smith 1987, 1992, 1993). Discrete-time systems which implement a waveguide model are called *waveguide filters* (WGFs). A particularly nice feature of the digital waveguide approach is that it simulates physical phenomena directly in a digital way, that is, there is no need to first develop a continuous-time model and then discretise this in time.

First we discuss a well-known forerunner of waveguide models, the Karplus–Strong algorithm.

4.2.1. *The Karplus–Strong model*

Kevin Karplus and Alex Strong developed a simple algorithm for the synthesis of plucked string sounds (Karplus and Strong 1983). Their method is based on the idea that a wavetable, i.e. a table containing a sampled waveform of an audio signal, is modified while it is read. The technique was found to be a special case of a string simulation studied by McIntyre, Woodhouse and Schumacher (1983) and it was instantly extended by Jaffe and Smith (1983). The Karplus–Strong algorithm is an important predecessor of current waveguide models and recently it has led to quite detailed models of string instruments (see, for example, Karjalainen and Välimäki 1993, Karjalainen, Välimäki and Jánosy 1993, Smith 1993, Välimäki, Huopaniemi, Karjalainen and Jánosy 1995).

The block diagram of the Karplus–Strong model is shown in figure 3. The system is a recursive comb filter. The delay line (or wavetable) is initialised with white noise. The output of the delay line is fed to a low-pass filter that is called the *loop filter*. The filtering result is the output of the system and it is also fed back to the delay line.

This technique does not produce a purely repetitive output signal as basic wavetable synthesis does. In particular, those frequency components of the random signal which coincide with the resonances of the comb filter will attenuate more slowly than the other components. Thus, the signal in the delay line progressively turns into a pseudo-periodic signal with a clearly perceivable fundamental frequency.

The original proposition of Karplus and Strong was that a two-point averaging filter should be used as the loop filter, but in practice a more sophisticated digital filter is employed. It is important to remember that the filter's magnitude response must not exceed unity at any frequency – otherwise the system becomes unstable. When the magnitude response of the filter is less than unity at all frequencies the resulting sound will gradually attenuate. The timbre will also vary over time if the magnitude response is not flat, since the harmonics of the signal will attenuate at different rates. A digital low-pass filter that causes the high frequency components to decay more rapidly than the lower frequencies corresponds to a physically meaningful model. Dispersion can be brought about using a filter with a nonlinear phase response.

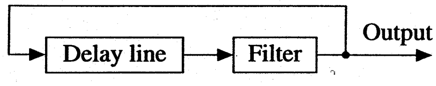

Figure 3. The Karplus–Strong model.

One of the major problems of the basic Karplus–Strong model is that the fundamental frequency of the tone cannot be accurately controlled. The fundamental frequency f_0 of the synthetic signal is determined by the length of the delay line, i.e.

$$f_0 = \frac{f_s}{M + 1/2},$$

where f_s is the sampling rate (in Hz), M is the length of the delay line in samples, and 1/2 refers to the delay caused by the two-point averaging filter. It is seen that this equation is a function of the integer-valued variable M. Thus the fundamental frequency is quantised and an arbitrary pitch cannot be produced.

The length of the delay line loop, and consequently the fundamental frequency f_0 of the synthetic signal, can be accurately controlled by adding a *fractional delay filter* in the feedback loop of the Karplus–Strong model. Jaffe and Smith (1983) introduced a first-order allpass filter for producing the desired fractional delay. Later, a linear interpolator (Sullivan 1990) and a third-order Lagrange interpolator (Karjalainen and Laine 1991) were used for this task.

4.2.2. *Advances in waveguide synthesis*

One of the first applications of the waveguide modelling technique was model-based sound synthesis of the clarinet (Smith 1986, Välimäki, Laakso, Karjalainen and Laine 1992, Rocchesso and Turra 1993). A waveguide flute model was introduced in Karjalainen, Laine, Laakso and Välimäki (1991) and Välimäki, Karjalainen, Jánosy and Laine (1992). In a wind instrument model, the waveguide consists of two delay lines which represent the propagation of wave components in opposite directions in an acoustic tube (figure 4). The length L of the delay lines is related to the effective length l_{eff} of the tube by

$$L = \frac{f_s l_{eff}}{c},$$

where f_s is the sampling rate (e.g. 22.05 kHz) and c is the speed of sound ($c \approx 340$ m s^{-1}). Note that in general

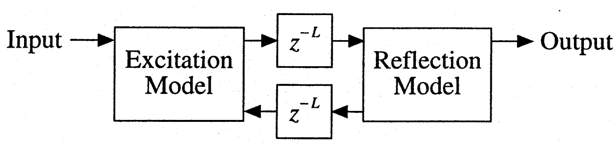

Figure 4. Waveguide wind instrument model.

L is not an integer and again a fractional delay filter must be used to implement the real-valued delay line.

The reflection model consists of a digital filter that brings about the frequency-dependent reflection and radiation of the sound wave at the end of the bore. Note that the reflection model has two outputs, one for the outgoing signal and the other for the reflected, ingoing signal.

The excitation model includes a *nonlinearity* which is an essential part of the sound production mechanism in wind instruments. In the case of the clarinet, this system models the operation of a reed that controls the air flow through the mouthpiece. The same nonlinearity is also suitable for other reed wind instruments, such as the saxophone (Cook 1988). In the flute, the nonlinearity models the airflow into the bore and is similar to a sigmoid function (Karjalainen *et al*. 1991, Välimäki, Karjalainen, Jánosy and Laine 1992).

Cook (1991, 1992) introduced a waveguide model that is capable of synthesising brass instrument tones. His system is also based on the principle shown in figure 4. However, the nonlinearity was obtained by modelling the lips of the player as a mass–spring oscillator. For discrete-time simulation the differential equations governing this oscillator were replaced with difference equations.

Välimäki, Karjalainen and Laakso (1993) have shown how finger holes may be incorporated in a waveguide woodwind instrument model. Recently, the digital waveguide technique has been generalised to two dimensions (Van Duyne and Smith 1993). This has allowed modelling of vibrating plates and, for example, drum heads. Further generalisation to three and more dimensions is straightforward. Savioja, Rinne and Takala (1994) have demonstrated that a 3D waveguide mesh can be used to simulate room acoustics.

Nowadays research waveguide modelling mainly concentrates on the nonlinear interaction of excitation and the vibrating objects. Other concurrent research topics include new techniques for modelling resonators, such as waveguide modelling of conical acoustic tubes (Välimäki and Karjalainen 1994), modelling the radiation characteristics of musical instruments (Huopaniemi, Karjalainen, Välimäki and Huotilainen 1994, Karjalainen, Huopaniemi and Välimäki 1995), and control of physical models (Jánosy, Karjalainen and Välimäki 1994).

5. USAGE OF VIRTUAL MUSICAL INSTRUMENTS

In the future, there will be more commercial synthesizers based on physical modelling. They will be more advanced synthesizers than those in use today. These will be welcomed by computer music professionals and also by composers and arrangers of acoustic music. For them physical models will offer the possibility of listening to a musical piece in progress before musicians start rehearsing it. Even now many composers use a MIDI synthesizer to listen to their pieces before introducing them to musicians or to the public. With the more realistic synthetic sounds of physical models, this electronic premiere can sound almost as natural as if it were played using acoustic instruments.

We may imagine that in the future when physical modelling has reached high accuracy in terms of simulating structures, it will be possible to simulate virtual musical instruments which are physically feasible, but which do not exist in physical reality. This could be useful for instrument builders when trying to improve or modify current acoustic musical instruments. Presently, instrument builders typically have to rebuild a prototype several times, since there is no other way of knowing whether a modification will yield a positive result. Even today, computers can aid the design of certain parts of instruments, e.g. the finger holes of wind instruments (Keefe 1982), but complete simulation and sound synthesis is yet to come.

Another exciting possibility is the restoration of historical musical instruments. When the analysis and parameter estimation tools for physical models have advanced to the point that a physical model can be made to simulate a recorded tone, it will be possible to imitate any instrument whose sound has been recorded. For example, the sound of antique violins that do not exist anymore could be recreated synthetically and be used for playing more beautiful music. Furthermore, old recordings that are partly destroyed could be restored by replacing the missing parts using a physical model of the old instrument. Of course, this might raise a discussion on the ethical suitability of such a procedure.

Researchers in computer music and physical modelling have begun to picture a new way of recording music. Instead of recording the music with a microphone and storing the sound samples on a CD record we could proceed as follows. Record the musical performance digitally and analyse it with a computer program that would perform two tasks:

(1) Estimate the parameter values for physical models of those instruments that are present in the performance. Since separation of sound signals from a recording is an extremely difficult task, it might be necessary to record some isolated tones from each instrument for the purpose of the analysis.
(2) Follow the music signal through time and estimate the time instants of changes in the control parameters. These changes would correspond to

change of notes, change of playing style, or some other phenomenon that would be caused by the musician.

The collection of parameters of the physical model and the control parameters would include all the information necessary to reconstruct the musical performance. This technique could be called *model-based coding of music*. There is no doubt that the amount of information to be stored would be far smaller than that used today for high-quality sound recording, even when sophisticated psychoacoustic coding techniques are used. The sound quality of a physical model can be characterised with a small set of numbers (just a few dozen) and control events occur very seldom (a few events per second) when compared to sampling events in digital recording (44,100 samples per second). The technique may sound unrealistic today, but once again it is worthwhile considering speech processing technology: in digital mobile phones the voice of the speaker is analysed, coded and decoded (resynthesised) with a technique that has its roots in physical modelling of the speech processing mechanism. Thus, model-based coding of music – perhaps first applied to performances by a single musical instrument – may not be as far in the future as one might at first think.

Virtual reality is also an exciting field where structure and behaviour of real or imaginary worlds are modelled and animated. With special devices like head-mounted displays, a human observer can be immersed in a virtual environment so deeply that he/she feels like really being inside the simulation. Most of the research so far has been in developing 3D interaction devices and highly believable visual effects. Recently, however, sounds in virtual worlds have also gained interest (Begault 1994). As the animation of virtual objects is often physically based in order to achieve the most lifelike motion, it is natural to use similar principles for sound generation as well. Striking similarities can be seen in the methods used: Pentland and Williams (1989) utilised modal analysis to represent visible deformations and vibrations of colliding objects, and Terzopoulos *et al.* (1987) applied the masses-and-springs model for the same purpose. Gaver (1994) produced sound effects for user interfaces (so-called auditory icons) based on simulated collisions when an object breaks into pieces, for example. Concurrently, the synchronisation of animation and sound effects is an area of active research. Takala and Hahn (1992) have outlined a pipelined process architecture for rendering sounds in an analogous fashion to the way photorealistic images are rendered from 3D geometric models. Although high-quality rendering of sound and images cannot be performed in real time, it is still useful in the production of animated films. With substantial simplifications an interactive system can be built, where the user can hear the effects of his/her actions on the virtual objects, feeling also the direction and distance of the sound event (Astheimer 1993). Not so far in the future we may expect to experience a virtual concert hall where virtual musicians play virtual instruments, either by themselves or led by a human conductor.

REFERENCES

Adrien, J. M. 1991. The missing link: modal synthesis. In G. De Poli, A. Piccialli and C. Roads (eds.) *Representations of Musical Signals*, pp. 267–97. Cambridge, MA: MIT Press.

Arfib, D. 1979. Digital synthesis of complex spectra by means of multiplication of nonlinear distorted sine waves. *Journal of the Audio Engineering Society* **27**(10): 757–68.

Astheimer, P. 1993. What you see is what you hear – acoustics applied in virtual worlds. *Proc. IEEE Symp. on Research Frontiers in Virtual Reality*, pp. 100–107. San Jose, CA.

Beauchamp, J. 1979. Brass-tone synthesis by spectrum evolution matching with nonlinear functions. *Computer Music Journal* **3**(2): 35–43. Reprinted in Roads, C., and Strawn, J. (eds.). 1985. *Foundations of Computer Music*, pp. 95–113. Cambridge, MA: MIT Press.

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

Cadoz, C., Luciani, A., and Florens, J. 1984. Responsive input devices and sound synthesis by simulation of instrumental mechanisms: the CORDIS system. *Computer Music Journal* **8**(3): 60–73. Reprinted in Roads, C. (ed.). 1989. *The Music Machine*, pp. 495–508. Cambridge, MA: MIT Press.

Chaigne, A. 1992. On the use of finite difference for musical synthesis – Application to plucked string instruments. *Journal d'Acoustique* **5:** 181–211.

Chaigne, A., Askenfelt, A., and Jansson, E. V. 1990. Temporal synthesis of string instrument tones. *Speech Transmission Laboratory – Quarterly Progress and Status Report* (*STL-QPSR*) **4:** 81–100. Stockholm, Sweden: Royal Institute of Technology (KTH).

Chowning, J. M. 1973. The synthesis of complex audio spectra by means of frequency modulation. *Journal of the Audio Engineering Society* **21**(7): 526–34. Reprinted in Roads, C., and Strawn, J. (eds.). 1985. *Foundations of Computer Music*, pp. 6–29. Cambridge, MA: MIT Press.

Cook, P. R. 1988. Implementation of single reed instruments with arbitrary bore shapes using digital waveguide filters. *Technical Report No. STAN-M-50.* Dept of Music, CCRMA, Stanford University, Stanford, CA.

Cook, P. R. 1991. TBone: an interactive waveguide brass instruments synthesis workbench for the NeXT machine. *Proc. 1991 Int. Computer Music Conf.* (*ICMC'91*), pp. 297–9. Montreal, Canada.

Cook, P. R. 1992. A meta-wind-instrument physical model, and a meta-controller for real-time performance control. *Proc. 1992 Int. Computer Music Conf.* (*ICMC'92*), pp. 273–6. San Jose, California.

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.

Gaver, W. W. 1994. Using and creating auditory icons. In G. Kramer (ed.) *Auditory Displays*. SFI Studies in Sciences of Complexity, Proc. Vol. XVIII, pp. 417–46. Reading, MA: Addison-Wesley.

Hiller, L., and Ruiz, P. 1971. Synthesizing musical sounds by solving the wave equation for vibrating objects: parts I and II. *Journal of the Audio Engineering Society* **19**(6): 462–70 & **19**(7): 542–50.

Huopaniemi, J., Karjalainen, M., Välimäki, V., and Huotilainen, T. 1994. Virtual instruments in virtual rooms – a real-time binaural room simulation environment for physical models of musical instruments. *Proc. 1994 Int. Computer Music Conf.,* pp. 455–62. Aarhus, Denmark.

Jaffe, D., and Smith, J. O. 1983. Extensions of the Karplus–Strong plucked string algorithm. *Computer Music Journal* **7**(2): 56–69. Reprinted in Roads, C. (ed.). 1989. *The Music Machine*, pp. 481–94. Cambridge, MA: MIT Press.

Jánosy, Z., Karjalainen, M., and Välimäki, V. 1994. Intelligent synthesis control with applications to a physical model of the acoustic guitar. *Proc. 1994 Int. Computer Music Conf.*, pp. 402–6. Aarhus, Denmark.

Karjalainen, M., Huopaniemi, J., and Välimäki, V. 1995. Direction-dependent physical modeling of musical instruments. *Proc. 15th Int. Congr. on Acoustics* (*ICA'95*), Vol. 3, pp. 451–4. Trondheim, Norway.

Karjalainen, M., and Laine, U. K. 1991. A model for real-time sound synthesis of guitar on a floating-point signal processor. *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing* (*ICASSP'91*), pp. 3,653–6. Toronto, Canada.

Karjalainen, M., Laine, U. K., Laakso, T. I., and Välimäki, V. 1991. Transmission-line modelling and real-time synthesis of string and wind instruments. *Proc. 1991 Int. Computer Music Conf. (ICMC'91)*, pp. 293–6. Montreal, Canada.

Karjalainen, M., and Välimäki, V. 1993. Model-based analysis/synthesis of the acoustic guitar. *Proc. Stockholm Music Acoustics Conf.* (*SMAC93*), pp. 443–7. Stockholm, Sweden.

Karjalainen, M., Välimäki, V., and Jánosy, Z. 1993. Towards high-quality sound synthesis of the guitar and string instruments. *Proc. Int. Computer Music Conf.* (*ICMC'93*), pp. 56–63. Tokyo, Japan.

Karplus, K., and Strong, A. 1983. Digital synthesis of plucked string and drum timbres. *Computer Music Journal* **7**(2): 43–55. Reprinted in Roads, C. (ed.). 1989. *The Music Machine*. Cambridge, MA: MIT Press.

Keefe, D. H. 1982. Theory of the single woodwind tone hole. *Journal of the Acoustical Society of America* **72**(3): 676–87.

Le Brun, M. 1979. Digital waveshaping synthesis. *Journal of the Audio Engineering Society* **27**(4): 250–66.

McAuley, R. J., and Quatieri, T. F. 1986. Speech analysis/synthesis based on sinusoidal representation. *IEEE Trans. on Acoustics, Speech and Signal Processing* **34:** 744–54.

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.

Pentland, A., and Williams, J. 1989. Good vibrations: modal dynamics for graphics and animation. Proc. SIGGRAPH'89. *Computer Graphics* **23**(3): 215–22.

Rocchesso, D., and Turra, F. 1993. A real time clarinet model on Mars workstation. *Proc. X Colloquio di Informatica Musicale*, pp. 210–13. Milan, Italy.

Rodet, X. 1984. Time domain formant-wave-function synthesis. *Computer Music Journal* **8**(3): 15–31.

Rodet, X., and Depall, P. 1992. A new additive synthesis method using inverse Fourier transform and spectral envelopes. *Proc. 1992 Int. Computer Music Conf. (ICMC'92)*, pp. 410–11. San Jose, California.

Savioja, L., Rinne, T. J., and Takala, T. 1994. Simulation of room acoustics with a 3-D finite difference mesh. *Proc. 1994 Int. Computer Music Conf.* (*ICMC'94*), pp. 463–6. Aarhus, Denmark.

Serra, X. J., and Smith, J. O. 1991. Spectral modeling synthesis: a sound analysis/synthesis system based on a deterministic plus stochastic decomposition. *Computer Music Journal* **14**(4): 12–24.

Smith, J. O. 1985. A new approach to digital reverberation using closed waveguide networks. *Proc. 1985 Int. Computer Music Conf.* (*ICMC'85*), pp. 47–53. Vancouver, Canada.

Smith, J. O. 1986. Efficient simulation of the reed-bore and bow-string mechanism. *Proc. 1986 Int. Computer Music Conf.* (*ICMC'86*), pp. 275–80. The Hague, The Netherlands.

Smith, J. O. 1987. Music applications of digital waveguides. *Technical Report no. STAN-M-39.* Dept of Music, CCRMA, Stanford University, Stanford, CA.

Smith, J. O. 1991. Viewpoints on the history of digital synthesis. *Proc. 1991 Int. Computer Music Conf.,* pp. 1–10. Montreal, Canada.

Smith, J. O. 1992. Physical modeling using digital waveguides. *Computer Music Journal* **16**(4): 74–87.

Smith, J. O. 1993. Efficient synthesis of stringed musical instruments. *Proc. 1993 Int. Computer Music Conf.* (*ICMC'93*), pp. 64–71. Tokyo, Japan.

Sullivan, C. S. 1990. Extending the Karplus–Strong algorithm to synthesize electric guitar timbres with distortion and feedback. *Computer Music Journal* **14**(3): 26–37.

Takala, T., and Hahn, J. 1992. Sound Rendering. Proc. SIGGRAPH'92. *Computer Graphics* **26**(2): 211–20.

Terzopoulos, D., Platt, J., Barr, A., and Fleischer, K. 1987. Elastically deformable models. Proc. SIGGRAPH'87. *Computer Graphics* **21**(3): 205–14.

Van Duyne, S. A., and Smith, J. O. 1993. Physical modeling with the 2-D digital waveguide mesh. *Proc. 1993 Int. Computer Music Conf.* (*ICMC'93*), pp. 40–47. Tokyo, Japan.

Välimäki, V., and Karjalainen, M. 1994. Digital waveguide modeling of wind instruments bores constructed of truncated cones. *Proc. 1994 Int. Computer Music Conf.* (*ICMC'94*), pp. 423–30. Aarhus, Denmark.

Välimäki, V., Huopaniemi, J., Karjalainen, M., and Jánosy, Z. 1995. Physical modeling of plucked string instruments with application to real-time sound synthesis. *AES 98th Convention.* Paris, France. A revised version will be published in the *Journal of the Audio Engineering Society*.

Välimäki, V., Karjalainen, M., Jánosy, Z., and Laine, U. K. 1992. A real-time DSP implementation of a flute model. *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing* (*ICASSP'92*), Vol. 2, pp. 249–52. San Francisco, California.

Välimäki, V., Karjalainen, M., and Laakso, T. I. 1993. Modeling of woodwind bores with finger holes. *Proc. Int. Computer Music Conf.* (*ICMC'93*), pp. 32–9. Tokyo, Japan.

Välimäki, V., Laakso, T. I., Karjalainen, M., and Laine, U. K. 1992. A new computational model for the clarinet. *Proc. 1992 Int. Computer Music Conf.* (*ICMC'92*), Addendum. San Jose, Califonia.