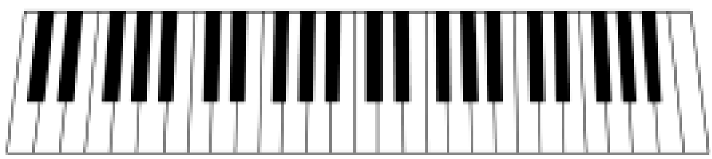
OBJECT-BASED SOUND SYNTHESIS for Virtual Environments

Using Musical Acoustics

Federico Pedersini, Augusto Sarti, and Stefano Tubaro

The physical modeling of complex sound generators can only be approached by individually synthesizing and discretizing the objects that contribute to the generation of sounds. This raises the problem of how to correctly implement the interaction between these objects. In this article we show how to construct an object-based environment for sound generation, whose objects can be individually synthesized and which can interact with each other through the modeling of a potential interaction topology. We will also show how this interaction topology can be made dynamic and time varying. We will further discuss how we envision an object-based environment that integrates geometric, radiometric, and intrinsic/extrinsic acoustic properties. We will finally illustrate our first results toward the modeling of complex sound generation systems.



Background

Recent achievements in the areas of computer vision, digital image processing, and computer graphics have made it possible to model realistic visual three-dimensional (3-D) environments [17], [27], [31], [33], [72], [2] and 3-D objects [4], [68], [39], [52]-[55] with efforts and costs that tend to decrease day by day. At the same time, a great deal of effort has also been spent in the area of audio analysis, synthesis, and processing, which has brought significant results in the perceptual encoding/compression of natural audio [10], in the realistic production of synthetic sounds [48], [20], [6], [8] and in their realistic spatialization [29], [36], [46], [47], [70]. Behind this tidal wave of results is an exponential increase of computational power at low cost, combined with an ever more capillary diffusion of domestic Internet connections and a consequent explosion of new forms of networked services. An attempt to prevent research from producing just a disorganized collection of heterogeneous results has been made by the MPEG standardization board, which in the past few years has spent tremendous efforts toward a standardization for multimedia encoding/decoding and document retrieval and is still working toward this goal. This effort has encouraged the research community to channel the new results within the standards and has motivated technology providers to focus on the production of low-cost multimedia set-top boxes that could enable even computer-illiterate users to access shared networked virtual environments.

In the past decade, the market has been preparing itself for the maturing of the technology related to the creation of interactive multimedia environments. In fact, electronic commerce, with virtual shopping centers and market places, is already a booming reality. All this is creating a tremendous demand for authoring tools that could speed up the process of creation and personalization of these environments, which are expected to become more and more realistic in the near future. The concept of virtual university is also stirring a great deal of interest as more and more academic institutions are getting ready to become didactic providers for remote learners.

In spite of the abundance of results in the many areas of interest for multimedia applications, little has been done for synergically exploiting those that concern synthetic and natural modeling/rendering of visual 3-D environments and acoustic 3-D environments. From a certain viewpoint, this is rather surprising as the modeling/rendering of 3-D objects and environments has a great deal in common with the modeling/spatialization of sounds. For example, the photorealistic rendering of a 3-D environment requires the specification of a geometric model and a radiometric model for the surfaces of the environment, plus a model for the illumination. Similarly,

1053-5888/00/$10.00©2000IEEE

realistic sound rendering requires a description of the surfaces of the environment (as far as both geometry and sound reflection/absorption are concerned), and a model of the sound source. Consequently, a number of similarities can be found between classical visual rendering techniques and advanced sound spatialization models.

A strong parallel can also be recognized between advanced visual 3-D modeling techniques and sound synthesis techniques. For example, 3-D visual modeling started out with completely synthetic (CAD) techniques and progressed with measurement-driven solutions (e.g., modeling based on images, range cameras, or laser scanners). More recently, some 3-D modeling techniques have appeared, which go beyond the modeling of surfaces, and try to also describe the mechanism that animates them (e.g., physical modeling of human faces with description of dermis and muscles [66], [67]). The evolution of sound synthesis techniques [8] is similar. Initially, the most widespread solutions for sound generation were fully synthetic (e.g., oscillator-based synthesis, nonlinear distortion, frequency modulation synthesis, etc.) [20], [8], and then a number of methods appeared which used natural sounds to model synthetic ones. Some popular examples are granular synthesis [20], [8] (mosaicing of natural sound particles) and wavetable synthesis [48] (based on the postprocessing of natural sound samples). More recently, a number of solutions have appeared in the literature, which are based on the modeling of the mechanism of production of sound rather than on the modeling of the sound itself [6], [7], [50], [51], [60], [61], [64]. This physical modeling approach is gaining more and more popularity for a number of reasons.

Object-Based Audio Environments

A common way to model a basic visual 3-D environment is to describe it as a set of objects which interact with each other and with the users in 3-D space. Objects can be surfaces that generate, reflect, refract, or diffuse light or points or curves that act as light sources. Passive surfaces are assigned a position in space, a shape, and some radiometric properties (albedo or texture and reflectivity or transparency model). Light sources (active objects) are assigned a radiation model. Objects can be time varying and can be made "sensitive" to events, which could be within the 3-D environment (e.g., the contact with other surfaces) or external (user's action through some input device). When one such event occurs, an appropriate reaction of the object and/or of the environment is triggered. The whole 3-D scene can thus be thought of as a set of individually modeled objects, which interact with each other depending on their mutual positioning within the 3-D environment and on external events. A simple approach to the rendering of this scene is done through an analysis of the mutual location of active/passive objects and viewpoints, based on some ray-tracing strategy [31], or some more sophisticated approach based on "radiosity" [33], with the help of some post-processing tools aimed at improving the final rendering quality (e.g., radiometric surface smoothing through an interpolation of the normals of the triangles of the surface mesh).

An object-based environment for the generation and the spatialization of sounds could be envisioned in a way that is quite similar to what is described above. In fact, a generic object of this environment could be a functional element of a sound-generating device, or it could be an element of the environment that contributes to the sound spatialization/reverberation. As an object could play both roles at once, sound attributes that should be attached to it are:

- *Intrinsic attributes,* which describe the internal vibrational properties, the mechanical/fluidodynamical properties of the object;
- *Extrinsic attributes,* which describe the way the object irradiates soundwaves in the environment and/or influences soundwaves propagated in the environment, i.e., how the object irradiates sounds and/or reflects/absorbs/diffuses incident soundwaves.

Intrinsic attributes are invoked when the object is involved in the generation of sounds, while extrinsic attributes are mainly invoked in the sound rendering phase. In fact, the spatialization process requires the specification of both shape and sound reflectivity properties of the surfaces of the environment. Quite clearly, an object may have both intrinsic and extrinsic sound attributes, as it can itself act as a sound generator or it can act as a sound scatterer for other generators. Its extrinsic properties may be both active and passive: the active attributes, in fact, correspond to the description of how the resonator irradiates soundwaves in the environment, while passive attributes describe how the object reflects, diffuses, and absorbs incident soundwaves. For example, the surface of a gong has both intrinsic (vibrational) and extrinsic (radiational) properties. The active radiational properties describe how the vibrations of its surface are irradiated into the environment, while the passive radiational properties characterize the reflection and the scattering of pressure waves of external origin due to the object's surface.

The audio rendering of the scene should be done through an analysis of the mutual location of surfaces, sound sources, and position of the auditory points (virtual microphones) by adopting a strategy that plays the role that ray tracing and/or radiosity play in the photorealistic rendering process [17], [33].

Sound objects should be time varying, and their behavior should be made sensitive to events, which could take place within the 3-D environment (e.g., contact condition) or with the user (user's action through some input device such as a mouse or a MIDI actuator). When one such event occurs, an appropriate reaction of the objects and/or the environment should be triggered.

The above discussion on the attributes that characterize the generation of sounds and the influence of the environment on the rendering and the spatialization raises questions on how, in fact, these attributes can be used for the creation of an object-based environment that involves multisensory forms of content. We believe that the ambitious goal of an object-based interactive hybrid audio/3-D environment can be achieved through the development of three object-based subenvironments (see Fig. 1) that interact with each other:

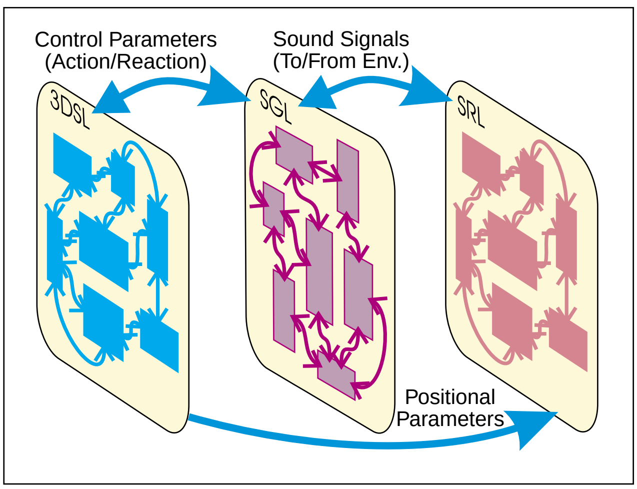
- An object-based 3-D scene layer (3-DSL);
- An object-based sound generation layer (SGL);
- An object-based sound rendering layer (SRL).

The object subdivision in the 3-D layer, in the sound synthesis layer, and in the spatialization layer may not be the same. In fact, 3-D objects are subdivided depending on the 3-D modeling approach and on the surface shape, audio objects are subdivided according to their functional role in the sound generation structure, while sound sources and scatterers are subdivided according to both geometrical and functional properties of such objects. The motivations behind a three-layered structure are the differences between such object subdivisions and, even more, the fact that keeping the problems of image rendering, sound generation, and sound rendering separate makes the synthesis far more flexible than otherwise believed. Anyway, no complications should arise from this layered approach if the mapping of signals between these three structures is done correctly.

Interaction between 3-DSL and SGL: The interaction between these layers is in terms of control (positional) signals and is two directional:

- Physical modeling of sounds enables the use of physical inputs, therefore the positional parameters of some 3-D objects can be mapped onto inputs of some sound objects;
- Physical modeling of sounds enables responsive input/output, therefore positional feedback can be easily provided. Such parameters can be mapped onto some 3-D objects.

Interaction between SGL and SRL: The interaction between these layers is in terms of vibrational signals and is generally two directional:



1. A schematic representation of the interactions between the 3-DSL, the SGL, and the SRL.

- The vibrations that take place in SGL's resonators need to be transferred onto the SRL and irradiated into the environment, using the extrinsic attributes of the corresponding objects;
- The sound irradiated into the environment may produce sympathetic vibrational phenomena in other resonators, therefore there may be signal transferral from SRL to SGL as well.

Interaction between 3-DSL and SRL: The interaction between these layers is in terms of control (positional) signals and is basically feed forward (the interaction between 3-DSL and SRL becomes two directional when the geometry of the environment is modified to achieve a desired spatialization quality). In fact, the positional parameters of generators, scatterers, and viewpoints (which correspond to auditory points as well) are directly passed by the 3-D object layer to the spatialization layer to change the parameters of the spatialization model. Indeed, sound scattering depends also on the 3-D shape of the reflectors, and that information must be kept current.

In the next section we will focus on the SGL, and we will illustrate our object-based approach to sound synthesis through automatic physical modeling.

Sound Generation Layer

In principle, the SGL could be defined and implemented using traditional sound synthesis algorithms [20], [6], [8] based on *direct generation* (sound sampling, additive synthesis, granular synthesis, etc.) or on *signal modification* (subtractive synthesis, nonlinear distortion synthesis, frequency modulation, etc.). These methods, however, are characterized by a certain timbral rigidity because they model sounds rather than the sound generation mechanism. Such sound synthesis methods, in fact, are not suitable for interpreting positional parameters from the 3-D scene layer to modify the timbre of the generated sounds in a meaningful, predictable, and plausible fashion. Furthermore, they do not allow the specification of a physically plausible feedback signal to be mapped back onto the 3-D scene layer. To make interaction between SGL and 3-DSL possible and realistic, one reasonable solution is to implement sound synthesis with the physical modeling of the sound generation mechanism [5]-[7]. This choice has a number of advantages:

- The action on the model is specified through control signals with physical meaning;
- A precise relationship exists between the reaction of the reference physical instrument to a certain action and the reaction of its model;
- The model can easily be made responsive in the sense that it can return a physical positional feedback;
- Timbral richness is determined by the model structure rather than by the complexity of its parametric control (the model has its own timbral dynamics);
- A physical model simplifies the specification of the sound radiating surface in the spatialization layer.

The physical synthesis of sounds consists of modeling the vibrational phenomena that occur in a complex resonating structure, which can be made of a number of simpler resonators connected together. The vibrational phenomena are normally caused and, possibly, sustained by the interaction with other structures. Examples of such interactions are a gong hammered by a mallet or a bowed violin string.

The Issue of Local Discretization

Indeed, a sound generating system made of two or more interacting objects could be modeled and discretized as a whole, as normally done in the literature. This choice, however, would dramatically increase the complexity of the synthesis problem and reduce its flexibility. In fact, to account for all possible interactions between various objects, the sound environment would end up being modeled as a wide collection of complex and autonomous systems. As a consequence, to be able to construct an object-based sound environment with a reasonable effort, we need to develop a strategy that allows us to manage all possible interactions between individually synthesized objects, by planning and implementing the interaction topology and solving all possible computability and stability problems beforehand.

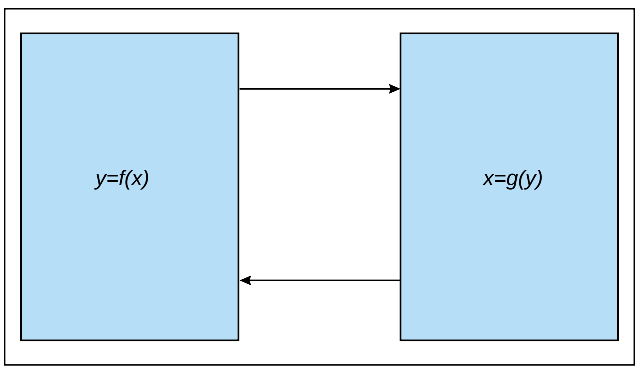
The problem of computability arises when we need to connect together two discrete-time models, each of which exhibits an instantaneous connection between input and output (see Fig. 2). In fact, the direct interconnection of the two systems would give rise to a delay-free loop (an implicit equation) in their implementation algorithm. This type of problems typically occurs when we try to connect together two individually discretized systems without taking into account any global interconnection constraint. To overcome this difficulty, the simplest solution, which is often adopted in the literature, consists of inserting a delay element in the noncomputable loops [71] (which correspond to deciding an artificial ordering in the involved operations). A more sophisticated approach is adopting some iterative numerical approach for solving the implicit equation that describes the noncomputable loop [21]. Whatever the solution may be, it involves a certain cost or risk in the final digital implementation, especially when discontinuous nonlinearities are present in the model. In fact, too simple a solution will tend to modify the system's behavior and, often time, to cause severe instability. Conversely, a more sophisticated iterative solution will dramatically increase the computational cost, as an implicit equation will have to be solved at each time instance.

This last discussion brings us to the problem of stability of the global implementation. It would be highly desirable for a block-based synthesis strategy to be able to preserve the stability properties of the analog reference system. This would allow us to select a sampling frequency that is only related to the involved signal bandwidths, rather than to the adopted discretization strategy. In other words, we would like to keep the oversampling factor (of the temporal discretization) as low as possible, without giving up the physicality or the behavioral plausibility of the system. Unlike what it may seem, this problem is quite critical when highly nonlinear elements are involved in the model implementation, which is our case not just because systems may be intrinsically nonlinear, but because contact conditions are modeled by step functions.

The sound objects that we are interested in are *resonators* [6]-[8], which are the sites where the vibratory phenomena take place. Such elements are modeled as linear dynamic systems that may incorporate an instantaneous nonlinear element to model the contact condition with other blocks and, in some cases, some sort of contact deformation. Examples of resonators are the string-soundboard structure of a piano or a violin, the acoustic tube of woodwinds or brass instruments, and the whole metal structure of a gong or a bell.

There are also other types of dynamic systems that need to be modeled, which play more the role of *exciters* [6]-[8] than that of resonators. These are elements whose only role is to cause and, possibly, support the vibratory phenomenon in the resonator and are usually modeled as nonlinear dynamic systems. Examples of exciters are the drumstick of percussions, the bow of a string instrument, the reed of a clarinet, and the human lips for brass instruments. Vibratory phenomena, however, can also be mutually caused by a collision between two resonators, in which case a nonlinearity must be included to model the contact condition. One difference that often discriminates between an exciter and a resonator is the fact that the former is usually modeled with lumped parameters (i.e., with a set of differential or even algebraic equations), while the latter is usually modeled with distributed parameters (i.e., with a set of partial differential equations).

The subdivision into blocks is induced by their functional role within the structure, and it would be desirable to preserve it during the synthesis phase. It is the goal of this section to show how it is possible to adopt a local approach to synthesis, which allows us to individually syn-



▲ *2. The problem of computability created when two digital systems that exhibit an instantaneous I/O connection are connected together. The absence of a delay element in the loop generates an implicit equation between inputs and outputs.*

thesize the building blocks and take care of their interaction later on.

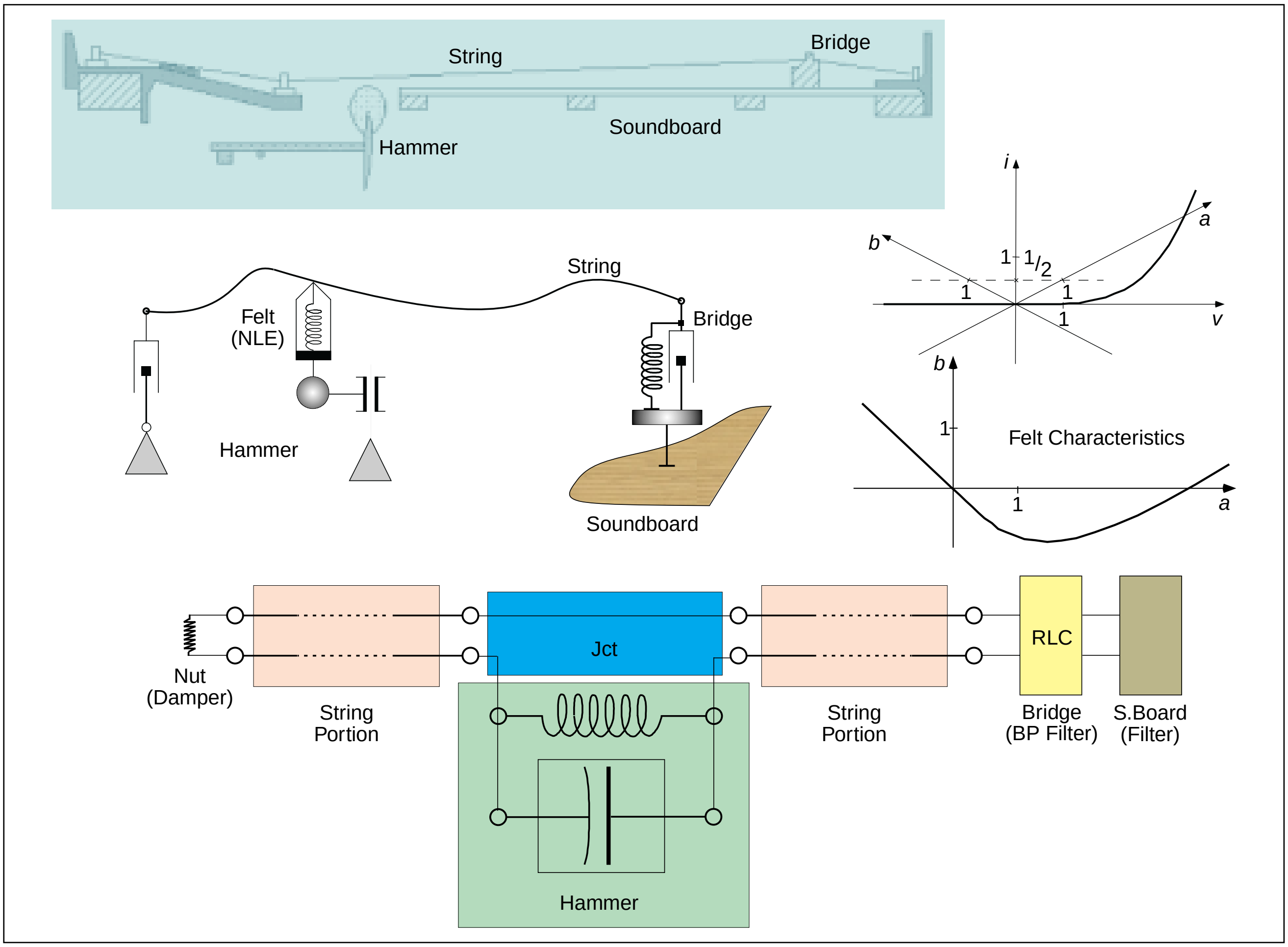
Wave Digital Structures

A physical structure (mechanical or fluidodynamical) can be described by an electrical equivalent circuit made of lumped or distributed elements. The equivalence can be made rather arbitrarily as a physical model is always characterized by a pair of extensive-intensive variables (e.g., voltage current, force velocity, pressure flow, etc.), and reciprocity principles can always be invoked. For example, if we wanted to model the hammer-string interaction in a piano we could first select a simplified model of the actual piano mechanism and then adopt an electrical equivalent of it, as shown in Fig. 3. In this case the equivalence is established by having forces and velocities correspond to voltages and currents, respectively.

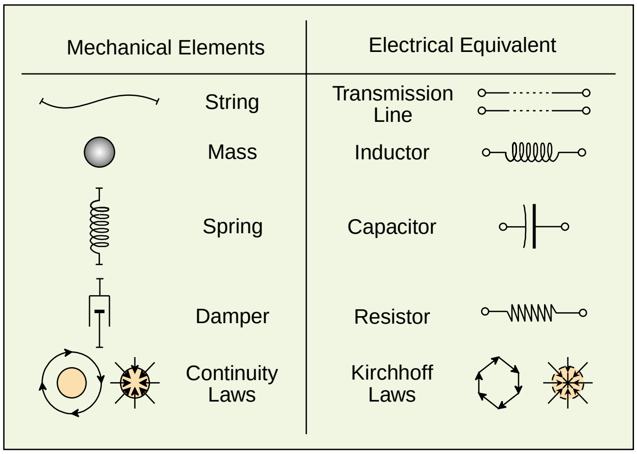
In general, we can recognize a number of equivalences between mechanical and electrical models, which can be used to automatize the construction of the electrical model. Some of these correspondences are shown in Fig. 4. Similar equivalences can be established between electrical and fluidodynamical variables/laws, for the modeling of interactions between acoustic tubes and specific exciters such as the jet flow in flutes and organ pipes or the reed in woodwinds.

Going through the electrical equivalent of the sound-production mechanism provides us with a standard representation of physical models. This representation cannot be digitally implemented using a local approach, however, as a direct interconnection of individually discretized elements would give rise to problems of computability. This is to be attributed to the fact that, when using extensive-intensive (voltage-current) pairs of variables, a direct interconnection of the blocks will not account for global constraints such as Kirchhoff laws.

One way to overcome this difficulty is to describe the system by means of scattering parameters [3]. This allows us to exploit the concept of adaptation to avoid computability problems. A well-known local method for designing filters after linear circuits, which is based on this approach, is that of wave digital filters (WDFs) [23], [24], [26]. The method consists of adopting a different pair of wave variables $a = v + Ri$ and $b = v - Ri$ for each element of the circuit, with R being a free parameter called reference resistance. This corresponds to a linear change of reference frame, from a (v,i) pair to an (a,b) pair, performed with a linear transformation with one degree of



▲ *3. Construction of the electrical equivalent of a piano model. When the hammer is in contact with the string, the velocities of hammer and string are the same at the contact point, therefore the contact junction is a series junction (current corresponds to velocity, voltage corresponds to force).*



▲ *4. Some examples of correspondences between physical elements/laws and electrical elements/laws.*

freedom (reference resistance R). The global constraints (Kirchhoff laws) are modeled in the interconnection phase, using multiport series and parallel adaptors, which also account for all the changes in the reference frames from point to point. The degree of freedom in the specification of the reference frame can be exploited to satisfy an adaptation condition on one of the ports of each adaptor. An adapted port, in fact, will not exhibit a local instantaneous wave reflection, thus guaranteeing that no computability problems will take place.

One key aspect of WDFs is that they preserve many properties of the analog filters that are used as a reference, such as passivity and losslessness [25]. Because of that, in the past few years we witnessed renewed interest in WDFs as the research in musical acoustics started to turn toward synthesis through *physical modeling*. This interest in WDFs is also due to the popularity gained in the past few years by *digital waveguides* (DWGs), which are close relatives of WDF's. Such structures, in fact, are suitable for modeling resonating structures in a rather versatile and simple fashion.

The DWG modeling approach consists of implementing the general solution of the equation that describes the propagation of perturbations in the structure [63], [64], together with its boundary conditions. For example, the general solution of the differential equation that describes the vibration of an infinitely long string (the ideal one-dimensional wave equation) is a pair of waves that propagate undistorted in the system, which can thus be modeled by a pair of delay lines. Such waves travel undistorted as long as the propagation structure is homogeneous (constant characteristic impedance). When a discontinuity occurs, wave scattering is modeled with a scattering junction structured like an adaptor of the WDF theory. A DWG structure will thus be made of an interconnection of delay lines, scattering junctions, and filters, which can be rather simply generalized to model lossy and dispersive propagation [64]. DWGs are suitable for simulating distributed resonating structures such as elastic strings, acoustic tubes, or even membranes and bells.

The similarity between DWGs and WDFs is not incidental, as the former represent the *distributed*-parameter counterpart of WDFs. In fact, they both use (incident and reflected) waves and scattering junctions. Thanks to such similarities, WDFs and DWGs turn out to be fully compatible with each other. However, while DWG waves are defined with reference to a *physical* choice of wave parameters such as propagation velocity and characteristic impedance, the reference parameters for WDF waves represents a degree of freedom to be used to avoid computability problems.

It is quite clear that hybrid WDF/DWG structures seem to offer a flexible solution to the problem of sound synthesis through physical modeling. One should keep in mind, however, that both the classical WDF theory and the DWG theory are inherently linear, which raises the problem of how to incorporate nonlinearities into a generic wave digital (WD) structure, as they are predominant in musical acoustics.

Nonlinear WD Structures

The need to incorporate nonlinear elements in WDF structures was first recognized by Meerkötter [49], who noticed that in any linear WDF structure there is always one degree of freedom left in the global combination of reference resistances, which can be exploited to adapt the port where the nonlinear element needs to be connected to. Indeed, since the wave variables are either voltage or current waves, the nonlinear elements that can be incorporated in WDF structures this way are nonlinear resistors. Quite clearly, the wave nonlinearity (a $b-a$ curve) that will be connected to the reflection-free (adapted) port of the WDF structure is obtained from the Kirchhoff characteristic of the nonlinear resistor (a $v-i$ curve) using the same transformation that defines pairs (a,b) of waves as a function of Kirchhoff pairs (v,i) of variables (voltage and current).

Nonlinear resistors, however, represent only a subset of the so-called "algebraic" nonlinearities [16] encountered in nonlinear circuit theory and in musical acoustics. Algebraic bipoles are described by an equation between the two port variables $v^{(j)}$ and $i^{(k)}$, where $j,k\in\{0,\pm1,\pm2,\ldots\}$ denote time differentiation (if positive) or integration (if negative) of v and i. Nonlinear devices that are not algebraic are called *dynamic* [16] elements. The simplest examples of nonlinear algebraic bipoles are nonlinear resistors, capacitors, and inductors, although many others can be found in the literature of nonlinear circuit theory (FNDRs, supercapacitors, superinductors, memristors, etc. [16]).

Modeling nonresistive algebraic nonlinearities with classical WDF principles is known to give rise to problems of computability, since closed loops without delays cannot be avoided in the resulting WD structure. An example of WD implementation of a circuit containing a nonlinear reactance can be found in [21], whose scheme exhibits a problem of computability that is numerically avoided by solving a nonlinear implicit equation at every time instance. Other authors [71], in similar situations, choose a more rudimental solution that consists of inserting a delay element where the noncomputable connection (delay-free loop) is found. This solution, however, could easily introduce unacceptable discretization errors or instability problems.

To overcome computability problems without having to solve implicit equations, a different solution for a wave implementation of circuits that contain reactive nonlinearities was proposed in [60] and [22]. In this solution, new waves are defined to be suitable for the direct modeling of algebraic nonlinearities such as capacitors and inductors. In fact, with respect to the new waves, the description of the nonlinear element becomes purely algebraic, so that the results already formulated for nonlinear resistors [49] can be applied. To adopt such new waves, a special two-port element that performs the change of variables is defined and implemented in a computable fashion. The reactive nonlinear element is thus modeled in a new WD domain, where its description becomes memoryless. Roughly speaking, with respect to the new wave variables, the behavior of the nonlinear bipole becomes resistor like, therefore the two-port junction that performs the change of wave variables plays the role of a device that transform the reactance into a resistor.

The above idea of transforming reactances into nonlinear resistors is not new in the theory of circuit design. In fact, the literature on nonlinear circuits is rich with results that allow the designer to model arbitrary nonlinear networks by using just nonlinear resistors, operational amplifiers, and other linear circuit elements [11]-[15]. By doing so, it is possible to design arbitrary bipoles without ever using a nonlinear inductor or a nonlinear capacitor, which are more difficult to implement. This is possible by using special two-port analog devices called *mutators* [16], [12], [11], which are built using only operational amplifiers and linear passive resistors and capacitors. In general, mutators reduce the problem of realizing a wide class of nonlinear bipoles with memory to that of synthesizing a nonlinear resistor. The method proposed in [60] and [22] is the digital counterpart of this analog approach to the design of nonlinear circuits.

Generalized WDF structures

A further extension of the ideas introduced in [60] and [22] was recently proposed [61], which introduced a more general family of digital waves that allow us to model a wider class of algebraic and dynamic nonlinearities. The consequent generalization of the WDF principles include dynamic multiport junctions and adaptors, which synergically combine ideas of nonlinear circuit theory (mutators) and WDF theory (adaptors). This generalization provides us with a certain degree of freedom in the design of WD structures. In fact, not only can we design a dynamic adaptor in such a way to incorporate the whole dynamics of a nonlinear element into it, but we can also design a dynamic adaptor that will incorporate an arbitrarily large portion of a linear structure. It can be easily proven [61] that, under mild conditions on their parameters, such multiport adaptors are nonenergetic, therefore the global stability of the reference circuit is preserved by the wave digital implementation. For this reason, such multiport junctions can be referred to as dynamic *adaptors*.

The class of digital waves that we use for modeling a port in the WD domain is basically of the form

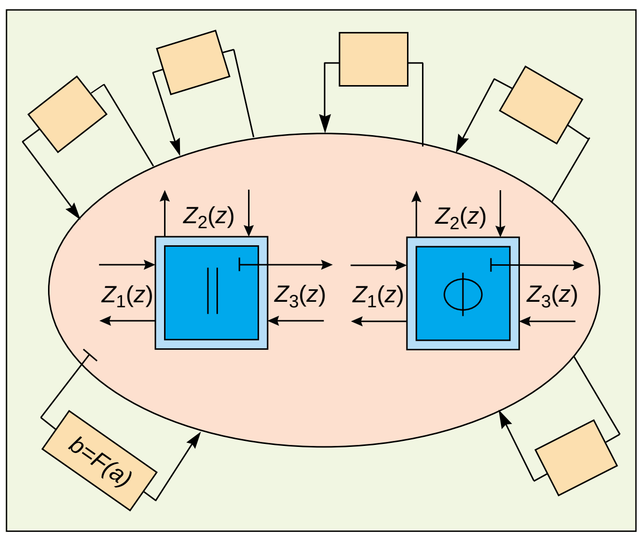
$$A(z) = V(z) + R(z)I(z)$$
$$B(z) = V(z) - R(z)I(z),$$

where $R(z)$ is a reference transfer function (RTF) [60], [61]. With this choice, the class of nonlinearities that can be modeled in the WD domain is, in fact, that of all algebraic bipoles of the form

$$p = g(q), \quad P(z) = H_v(z)V(z), \quad Q(z) = H_i(z)I(z),$$

where p and q are related to v and i, respectively, through a finite difference equation, while $R(z) = H_v(z)/H_i(z)$. The above choice of digital waves allows us to model a wide class of nonlinear dynamic elements, such as nonlinear reactances (e.g., nonlinear springs) or, more generally, linear circuits containing a *lumped* nonlinearity. The *memory* of the nonlinear element is, in fact, incorporated in the dynamic adaptor or in the mutator that the nonlinearity is connected to. As a consequence, our adaptors cannot be memoryless, as they are characterized by reflection filters instead of reflection coefficients.

With this more general definition of the digital waves, we can define the adaptation conditions for any linear bipole by selecting the RTF in such a way as to eliminate the instantaneous input/output connection in its WD implementation (instantaneous adaptation). An "adapted"



▲ *5. Macro-adaptors in extended WDF structures are obtained by arbitrarily interconnecting together a number of dynamic adaptors. Such macro-adaptors model the local topology of "instantaneously decoupled" subsystems.*

bipole will thus be modeled in the WD domain as $B(z)=z^{-1}K(z)A(z)$, where the delayed reflection filter $K(z)$ can also be identically zero.

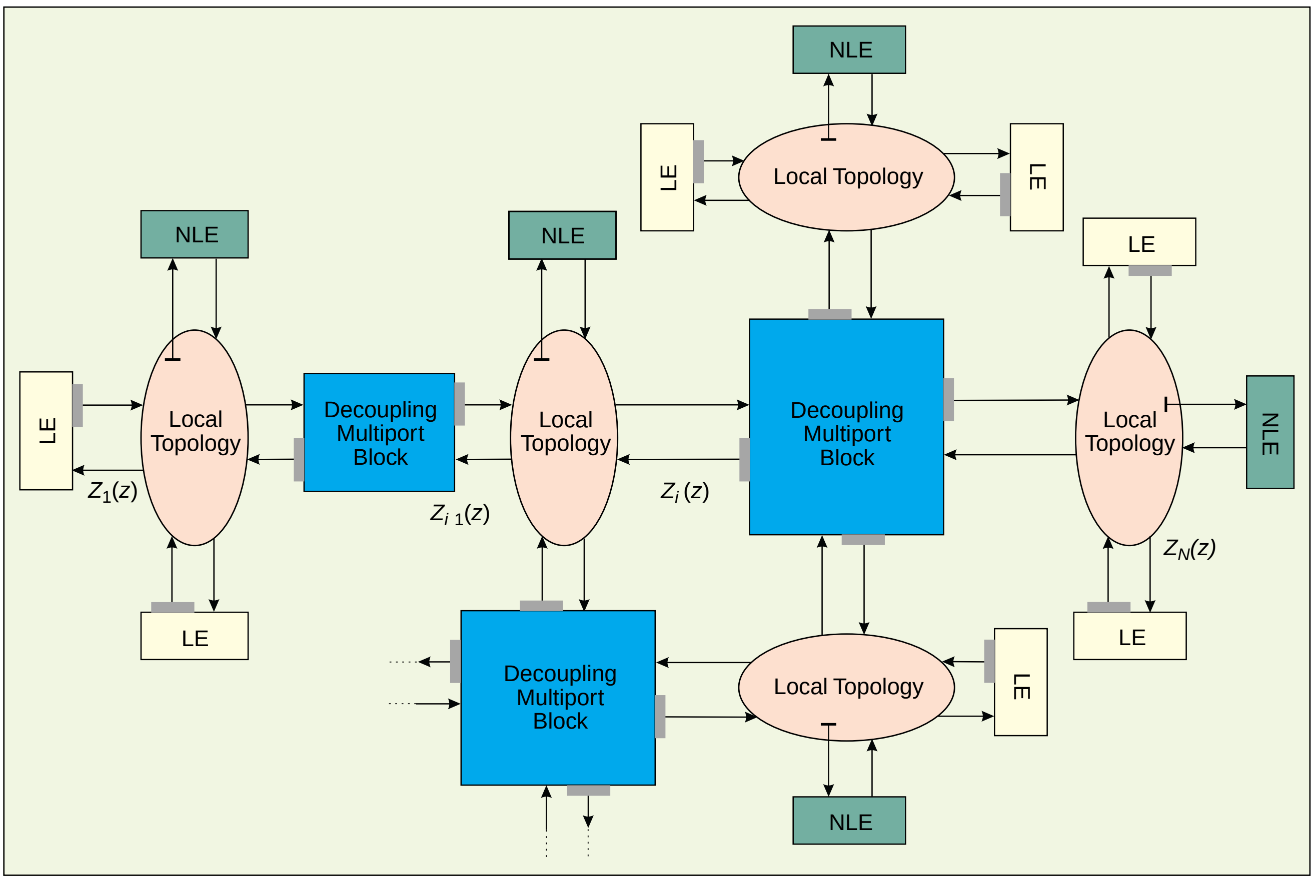
The interconnection between WD elements is implemented through a network of elementary (series or parallel) dynamic adaptors, as shown in Fig. 5. These adaptors take care of performing the necessary transformation (with memory) between variables, as each wave pair is referred to a different RTF. This network of elementary adaptors constitutes a dynamic macro-adaptor that can be proven to be nonenergetic [61]. This is an important feature of such elements as it allows us to guarantee that the passivity properties of the individual elements of the reference analog circuit be preserved by their WD counterpart. In fact, we have already verified that parallel and series multiport junctions are intrinsically nonenergetic provided that the port RTFs be stable. A computable interconnection through the nonenergetic junction of elements having the same passivity properties as the reference ones will preserve the stability properties of the reference analog circuit. We need to make sure, however, that the quantization of the filter coefficients will not affect the continuity constraints on the junctions and that the analog-to-digital mapping is always performed by means of the bilinear transformation.

Object Interaction

The sound synthesis approach summarized in the previous section, which we propose for the SGL, allows us to implement a fixed topology of interaction between sound objects, in the sense that the interconnection between objects needs to be specified in advance and cannot be changed on the fly. Furthermore, as complex resonating structures may become difficult to implement, initialize, and handle, it would be desirable to have some strategy for splitting the structure into smaller elements that are easier to deal with and may work in parallel. Finally, nothing has yet been said about how to implement and initialize macro-adaptors, which are the key elements of our structures. These problems are assessed in this section, together with the problem of how to make the synthesis structure time varying.

Planning the Topology

We will now show that there is a way to make the SGL interconnection topology dynamic, which exploits the nonlinear elements that implement the contact conditions. Let us consider an object that could potentially interact with a number of other objects in a sound environment. For example, we could think of a mallet that could collide, at different times, with a number of drum-like resonators.



6. Structure of a nonlinear block-based WD system with fixed interaction topology. The gray boxes at the ports of decoupling multiport block denote the presence of a delay element, which guarantees that neither instantaneous local reflections nor instantaneous reflections through outer loops will occur.

Indeed, this situation cannot be implemented with a fixed interaction topology such as the one of Fig. 6. To make this dynamic topology possible, we need to be able to connect or disconnect objects while the system is running. This can be achieved by exploiting the fact that a connection between systems is *irrelevant* when their contact condition is not satisfied.

As a simple example, let us consider the case of hammer-string interaction in the piano mechanism. The WD structure that corresponds to the equivalent circuit of Fig. 3 is shown in Fig. 7, where the macro-block M corresponds to the contact point between hammer and string. The nonlinear element (NLE) connected to the $R-C$ mutator [60], [51], [61] (the double-boxed two-port junction of Fig. 3, whose aim is to transform the nonlinear capacitor into a nonlinear resistor) corresponds to the nonlinear spring that models the felt deformation and, at the same time, the contact condition. It can be easily shown that when the contact condition is not satisfied, the series adaptor that connects the hammer to the two portions of the string becomes transparent for the two portions of waveguides that model the string. This fact suggests that removing the whole connection by replacing that series adaptor with a direct connection between the two waveguide portions would not modify the behavior of the resonator.

The above reasoning can be extended to more complex resonators and has a significant impact onto our implementation scheme. In fact, there are two important direct consequences that are worth mentioning:

- Systems that are not close to contact can be disconnected and may evolve independently;
- If the topology of the DWG network that implements the resonator is fixed, then a measure of proximity can be used for deciding whether and where to insert a transparent junction on the delay lines to preset the contact point.

An example of this situation is the physical model of a hammer dulcimer, where only two hammers are available for a set of many strings. In this case, in fact, the interaction topology between hammers and strings needs to be made dynamic and implemented as explained above.

Decoupling Elements

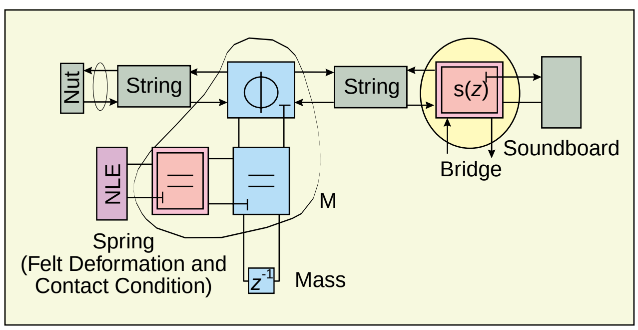
In general, while for a bipole the condition of adaptation corresponds to the possibility of extracting a delay element from it, for a multiport element this is no longer true. In fact, the port adaptation only implies that no *local* instantaneous reflections can occur, while nothing can be said about instantaneous reflections through outer paths. If it is true that a delay can actually be extracted from a port, then we talk about *instantaneous decoupling*, which is a stronger condition than adaptation. The concept of *instantaneous decoupling* is important as it allows us to split the synthesis and the initialization of large WD structures into that of smaller substructures [51], [50]. If N portions of a WD structure that are connected together through a decoupling N-port block ($N \geq 2$), which is a multiport element that exhibits at least $N-1$ decoupling ports, then such portions are said to be instantaneously decoupled, as they do not instantaneously interact with each other. One other reason why this decoupling condition is important is that it allows us to model WD structures that contain more than one nonlinearity. We know, in fact, that only one of all the ports of a macro-adaptor (the oval block of Fig. 5) can be adapted, therefore only one nonlinearity can be connected to it. Through a decoupling N-port block, however, we can connect together N macro-adaptors, each of which is allowed one nonlinear element.

Decoupling multiport blocks are quite frequently encountered in musical acoustics, especially when using DWG to implement reverberating structures. An example of a block-based sound synthesis structure where the decoupling condition allows us to model a large number of nonlinear elements is the acoustic piano. In this case, a number of wave digital hammers are connected, each through a DWG model of a string, to the same (decoupling) resonating structure (soundboard).

In conclusion, the global structure of a WD implementation of a physical model can be seen as a number of decoupled interconnection blocks such as those of Fig. 6, whose aim is to connect together either linear macro-blocks or instantaneous nonlinear blocks. The presence of decoupling ports allows us to approach the synthesis/initialization problem in a block-wise fashion. For example, if an interconnection block is connected to a set of adapted macro-blocks of the form $B(z) = z^{-1}K(z)A(z)$, then we can separate the synthesis/initialization of the macro-blocks of the form $K(z)$ from that of the interconnection block [51], [50]. A similar reasoning holds for two decoupled portions of the global WD structure.

Constructing Macro-Adaptors

The contact conditions allow us to *unplug* and isolate subsystems, while decoupling blocks allow us to approach the synthesis and the initialization of WD structures in a block-wise fashion. As a consequence, all that



7. WD structure for the modeling of piano sounds with fixed interaction topology. The contact condition is incorporated in the nonlinear element that is connected to the macro-adaptor M. When the contact condition is not satisfied, the macro-adaptor M becomes irrelevant and the string keeps evolving as if the macro-adaptor was not there.

is left to discuss is the construction and the initialization of a macro-adaptor.

An N-port macro-adaptor can be automatically built through a tableau-based approach, specifically designed for WD structures [51], [50]. Its description, in fact, can be given in the form

$$\mathbf{S}(z)\mathbf{C}(z) = \mathbf{0}^T,$$

where $\mathbf{S}(z)$ is a $2N \times N$ tableau matrix [3], $\mathbf{0}$ is a vector with N zero elements, and $\mathbf{C}(z) = [A_1 \cdots A_N \, B_1 \cdots B_N]^T$ is the vector of digital waves. A generic macro-adaptor can be thought of as a network of elementary (parallel or series) three-port adaptors with memory that belong to a predefined collection. This allows us to construct $\mathbf{S}(z)$ by "pasting" a number of predefined 6×3 matrices into a larger sparse matrix. This matrix equation can be quite easily rearranged and inverted to obtain a state-update equation, or else it can be solved iteratively using some efficient numerical method for sparse matrix equations.

As our macro-adaptors are not memoryless, they need to be properly initialized, which is a critical operation for WD models of mechanical systems as it usually affects the mutual position and contact conditions of mechanical elements. The determination of the state update equation can be seen as a direct form of the synthesis problem, as output signals are computed from input signals and memory content. Initialization, on the other hand, can be seen as an inverse problem, as memory content must be derived from output and input signals. As the nonlinearity is "lumped," this operation can be quite easily performed through nonlinearity inversion and matrix inversion.

Making the Structure Time Varying

Changing any model parameters in a WD structure usually affects all the other parameters as they are bound to satisfy global adaptation conditions. Temporal variations of the nonlinearities are easily implemented by employing special WD two-port elements that are able to perform a variety of transformations on the nonlinear characteristics (nonhomogeneous scaling, rotation, etc.). Temporal variations of RTFs, on the other hand, are implemented through a global recomputation of all model parameters on the behalf of a process that works in parallel with the simulator [51], [50]. This operation requires the remapping of the nonlinearities as well. This parameter update, however, is not computationally intensive as it is performed at a rate that is normally only a fraction of the signal rate (e.g., 100 times slower). It is important to remember, however, that abrupt parameter changes must be carefully dealt with not to affect the global energy in an uncontrollable fashion.

Automatic Implementation

Some methods are already available for synthesizing linear macro-blocks [63]; therefore the automatic synthesis procedure is based on the assumption that such elements are already available in the form of a collection of presynthesized structures. In its current state, the system that we developed is able to automatically compile the source code that implements a WD structure based on standard WDF adaptors and new dynamic adaptors chosen from a reasonably wide collection [51], [50]. The information that the system starts from is a semantic description of the network of interactions between all such elements.

Currently, the family of blocks includes WD mutators [60] and other types of adaptors developed for modeling typical nonlinear elements of the classical nonlinear circuit theory (both resistive and reactive). The available linear macro-blocks belong to the family of the DWGs [63], while the nonlinear maps are currently point wise described in the Kirchhoff domain and then automatically converted in a piecewise-linear WD map. Typical lumped WDF blocks are masses, springs, dampers, nonlinearities, ideal generators, and filters (especially allpass filters, for the fine tuning of strings or acoustic tubes or to account for the dispersive propagation in some enharmonic elastic structures such as bells, low piano strings, etc.). Typical distributed-parameter blocks are simple DWG implementation of strings and acoustic tubes, generalized DWG that accounts for rigidity and losses in a distributed fashion, reverberators based on Toeplitz matrices, green functions, and DWG models of 2-D and 3-D structures such as membranes and bells.

The parameters can be modified "on the fly" to make the structure time varying. A parallel process deals with the problem of recomputation of all WD parameters, depending on their changes expressed in the Kirchhoff domain.

An Example of Application

Our approach to the construction of the SGL has been tested on a variety of applications of musical acoustics. Starting from an appropriate semantic description of the building blocks and their topology of interconnection, we used our authoring tool to automatically generate C++ source code for the implementation of a number of typical acoustic musical instruments. The timbral classes implemented with this method are hammered strings (piano, electric piano), plucked strings (guitar), bowed strings (violin), reed instruments (clarinet, oboe), jet-flow acoustic tubes (flute, organ pipes), percussions, etc.

One of these examples, namely the grand piano, has been developed with a two-fold goal in mind: to test our solution to the problem of the mechanical modeling of a nontrivial acoustic instrument and to test our approach to the construction of a dynamic topology of interconnection.

The basic mechanism of hammer-string interaction is shown in Fig. 3, which corresponds to the block-based WD model of Fig. 7. As we can see in Fig. 8, the trajectories of the hammer and of the string at contact point and the temporal evolution of the force that the hammer exerts on the string are very "physical" and realistic. In fact, the hammer tends to bounce back a bit more every time a wave is reflected by the nut or the bridge and returns at the contact point, causing the ripples in the force's profile. This behavior turns out to have a very realistic impact on the resulting sound. The plotted output corresponds to the acoustic signal at the bridge.

The global implementation of the piano model has been entirely built using a rather extended network of WDF and DWG elements. The DWG model of each string includes stiffness [59] and losses [64]. The bridge is modeled as a bandpass filter (the WD equivalent of an RLC filter) and is connected to a rather complex soundboard model based on a DWG network. The string's fine tuning is performed using all-pass filters. A limited number of hammers are used dynamically to hit a full-scale resonator such as the one described above, with a dynamic management of the contact conditions. As for the spatialization layer, in its current state, a limited number of virtual pick-ups are scattered on the soundboard model and the vibrational signals are sent to a pseudophysical reverberator based on a circulant feedback delay network [57].

Indeed, the computational complexity of the resulting algorithm in this case coincides with the complexity of the resonating structure, whose role in the characterization of timbres is predominant. Some simpler implementations, however, already run real time on low-cost PC platforms. For example, the WD model of an electromechanical piano (e.g., Wurlitzer or Fender-Rhodes) can easily run with full polyphony (61 or 73 keys) on a Pentium III (350 MHz).

Sound Rendering Layer: An Overview

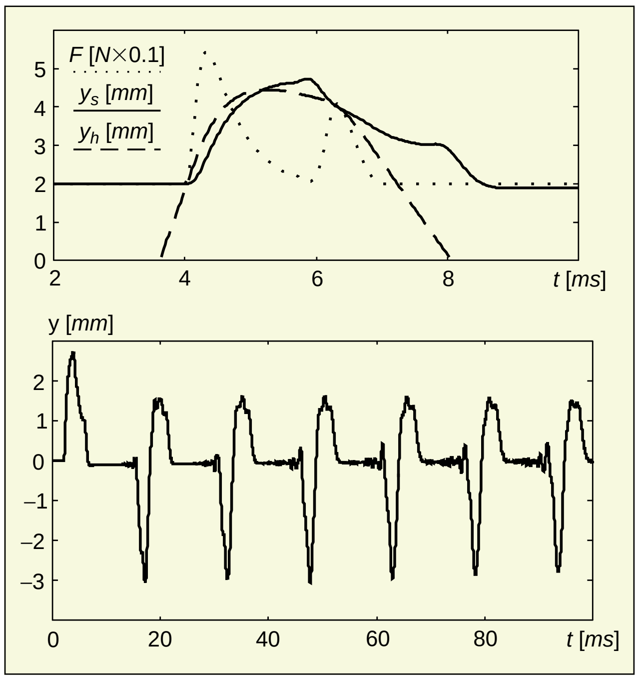
The literature is rich with sound reverberation techniques of practical usability [30], which range from simple comb-and-allpass filters that model early reverberation, to more complex resonating structures based on feedback delay networks [57], [58], DWG structures [64], [63] and multidimensional WDF [40]. Solutions based on DWG structures have the advantage of exhibiting a close similarity with the resonating structures used for sound generation through physical modeling. In fact, they are often implemented as a network of delay lines connected through multiport junctions.

When the goal is not just to obtain plausible reverberations, but to achieve audio-realistic rendering, more complex solutions need to be considered. Advanced rendering techniques, in fact, are aimed at creating a sense of presence by enabling a certain auditory comprehension of the proportions and the geometry of the surrounding space.

The sound rendering techniques that are available in the literature [29] can be roughly classified into five families: finite element methods, image source methods, boundary element methods, path tracing, and beam tracing. Aside from the first class of methods, all such techniques are inspired by visual rendering solutions. There are some strong differences between light and sound radiation, however, which complicate the situation in the acoustic case. First of all, sound wavelengths are much longer than light wavelengths. This does not allow us to ignore the phenomena of diffraction. Other consequences of having longer wavelengths are that specular reflections become dominant over diffuse reflections and that occlusions due to small objects have little impact on sound spatialization. A second significant difference is that sound propagation is much slower than light propagation. In fact, reverberations are the result of different propagation delays in different paths. Another crucial difference from visual rendering is that sound propagation is coherent. In fact, when modeling sound propagation we must carefully take the phase of acoustic waves into account. Because of these differences, successful visual rendering technique could easily translate into unfeasible sound rendering solutions.

Finite element methods consist of solving the wave propagation equation over a predefined volumetric grid. One interesting WDF-based solution within this category was proposed by Rabenstein [56], [40], [41]. These solutions produce very accurate results but they require the modeling of the propagation on all the points of a dense grid that samples the spatialization volume, which makes it quite demanding in terms of both memory and computations.

Image source methods [1], [9] are a first example of modeling of specular acoustic reflections in environments. They consist of computing specular reflection paths by "mirroring" the sound source over each surface of the environment. A line segment that connects a virtual (mirrored) source with the receiver is used to model a specular reflec-



▲ *8. Model simulation of an acoustic piano through a WD model. F is the force that the hammer applies to the string, y_h is the position of the hammer, and y_s is the position of the string at the contact point. Finally, y is the position of the string at the soundbridge.*

tion path. Higher order reflection paths can be obtained through a recursive generation of virtual sources through repeated mirroring. Indeed, this approach is robust, but it can only model specular reflections. Furthermore, its computational complexity increases exponentially with the number of reflections that we want to account for, which makes it feasible only for simple environments.

To approach the audio-realistic rendering problem in an object-based fashion, we need, once again, to invoke the parallelisms with classical visual rendering methods. Boundary element methods, path tracing, and beam tracing are all directly derived from visual rendering solutions.

Examples of boundary element methods that are extensively used in computer graphics are the so-called radiant exchange methods [17], [32], which model diffuse reflection of radiosity between surfaces. These techniques require the computation of form factors to measure the radiosity exchanged between pairs of surfaces patches. To obtain the radiosity of each surface patch of a given 3-D environment, it is necessary to simultaneously solve a set of transport equations.

Unfortunately, the success that this approach has had in visual rendering is difficult to match in the acoustic rendering case [46], [69]. This is because transport equations in acoustic modeling are expected to account for all the above-mentioned differences between light and sound radiation. In particular, they have to account for phase. Furthermore, besides specular reflections, they must consider paths of diffraction through extended form factor computations. Finally, to obtain good rendering results, the size of the surface patches must be much smaller than the acoustic wavelength and the solution must be computed for a number of frequencies, which makes the radiant exchange approach of little practical usability, especially for large 3-D environments.

Another class of image rendering techniques that has a counterpart in audio spatialization is ray tracing [31], [27], [73], [38]. Acoustic rays (paths) can be defined as small portions of spherical acoustic waves of negligible aperture, and geometrical acoustics can be applied to describe the reflections of such rays on objects [43]. A simple implementation of this approach for acoustic rendering can be obtained by ignoring the phenomenon of diffraction. This corresponds to assuming that acoustic wavelengths of interest are negligible compared with the size of the objects in the environment. In some applications where the frequencies that we care to spatialize are the higher ones, this might be not a restrictive assumption. Another assumption that needs to be made is the mutual incoherence (i.e., the absence of mutual interferences) between the waves associated to different acoustic rays. This last one is a difficult condition to guarantee in advance, but is generally verified with good approximation.

Along the direction of the acoustic ray, the pressure decreases with the square power of the distance. When the acoustic ray hits an object, a reflection occurs in a way that depends on the physical characteristics of the surface. If we focus just on specular reflections, we need to account for a filtering effect, due to the fact that the surface material interacts with the incident wave in a frequency-dependent fashion. This filtering can be modeled by a transfer function of the form

$$K = \frac{Z\cos\theta - 1}{Z\cos\theta + 1},$$

where θ is the angle of incidence of the acoustic ray and Z is the (frequency-dependent) characteristic impedance of the reflecting surface. When the surface is rigid and smooth, $Z \to \infty$ and $K \to 1$, providing a perfect reflection regardless of the incidence angle.

Path tracing techniques [42] consist of determining all the acoustic rays of interest between a sound source and a listening point (receiver). Rays are generated from the source point and followed through the environment until an appropriate set of significant rays is found to reach the listening point [38].

Path tracing methods have the advantage of being simple to implement, as only the interaction between rays and surfaces need to be computed. As a consequence, unlike finite element methods, the complexity depends less than linearly on the number of surfaces of the modeled environment. Furthermore, they can be devised and implemented in such as way as to model surface-ray interactions that are more complex than specular reflection. In fact, paths of diffuse reflection, diffraction, and refraction can be sampled as well. By doing so, it is possible to model any type of indirect reverberation and to accommodate arbitrarily shaped surfaces [18]. Conversely, path tracing techniques require a strong sampling of the fifth-dimensional (5-D) parameter space that describes all possible rays, which gives rise to approximations [45] and aliasing in the acoustic response of the environment. Approximations are due to the limited number of paths that can be modeled, while alias is due to the fact that there is no guarantee that some important acoustic ray will not be missed in the sampling process. Furthermore, such methods are difficult to implement when the auditory point is not stationary, as changes in the location of the receiver would require path retracing.

A very promising technique that overcomes, in part, these limitations is represented by beam tracing [29], which is the acoustic counterpart of the visual rendering technique described in [34]. Beam tracing methods are, in a way, a generalization of ray tracing techniques, as they trace pyramidal bundles of rays throughout the environment. Roughly speaking, the approach starts with the surface-based segmentation of the whole set of rays that emanate from the sound source. This first step produces a number of beams, each of which illuminates a different surface. Such beams are then clipped to remove the shadow region, which is then replaced by a properly constructed transmission beam. A reflection beam can then be obtained through the creation of a virtual source by simply mirroring the transmission beam over the surface.

The process is repeated for the reflected beam which, in turn, intersects some other surfaces. The main advantage of beam tracing over path tracing is that it exploits spatial coherence since, with a single beam, we can model an infinite number of ray-surface intersections. This overcomes the above-mentioned problems of approximation and aliasing that arise from a sampling of the 5-D parameter space that describes the rays. On the other hand, beam tracing methods are more difficult to implement than ray tracing techniques. In fact, determining intersection and clipping beams can become rather complicated from a geometric standpoint. Even more complex can be the modeling of refractions or reflections off curved surfaces.

An interesting characteristic of this approach [29] is the possibility of computing the topology of interaction between sources and surfaces through off-line preanalysis called beam tracing. This topology is described by a graph called a beam tree. By doing so, the rendering process consists of performing a lookup search for the beams of the beam tree that contain the moving receiver.

All the sound rendering solutions that have been briefly discussed in this section exhibit a number of advantages and disadvantages. When the goal is to obtain plausible reverberations rather than audio-realistic sound rendering, DWG techniques can be an interesting solution. In fact, they operate in a fully object-based fashion and are very similar in spirit to the sound synthesis approach proposed in this article for the SGL. The sound rendering techniques based on finite elements represent a good solution when we are interested in the modeling of the spatialization of low-frequency sounds, but they are not object based. Boundary element methods are also best suited for a realistic spatialization of low-frequency sounds, but their complexity is very high. When dealing with simple environments (typically a rectangular room), image source methods are probably the best solution, although they are limited to the modeling of specular reflections only. Depending on our expectations on the spatialization accuracy, path tracing and beam tracing appear to be very promising sound-rendering techniques. Path tracing is an excellent approach for modeling higher order reflections, although it is difficult to employ with dynamically moving auditory points. Although beam tracing appears to be more complex to implement, it enables the precomputation of the topology of interaction between surfaces and sources (beam tree). Both tracing solutions are also very close in spirit to the object-based layered organization of audiovisual 3-D environments, as presented in this article. Furthermore, we can outline a number of similarities between the object-based sound synthesis technique that we proposed for the SGL and such spatialization techniques for the SRL. In fact, in both cases, the construction and the managing of the topology of interactions is the most challenging and complex tasks to perform.

Conclusions

In this article we illustrated our approach to the construction of an object-based environment for sound generation, whose objects are individually synthesized and interact with each other through the modeling of their potential interaction topology. In particular, we showed how this interaction topology can be implemented in such a way to avoid problems of computability and to preserve the physical properties of the reference acoustic structure such as passivity and losslessness. We then illustrated our strategy to make this interconnection dynamic and time varying. We also discussed how we envision an object-based environment that integrates geometric, radiometric, and intrinsic/extrinsic acoustic properties.

The proposed approach has proven effective for the automatic and modular synthesis of a wide class of physical structures encountered in musical acoustics. In fact, the wave tableau approach we implemented makes the construction and the implementation of the interaction topology simple and systematic. In its current state, the implementation of the described synthesis system is able to assemble the synthesis structure from a syntactic description of its objects and their interaction topology, providing the user with a first CAD approach to the construction of an interactive sound environment.

We finally gave a brief overview on the available sound rendering methods and their potential of integration within a layered representation of audiovisual 3-D environments. The resulting scenario is challenging and promising, as new forms of interaction between users and acoustically responsive virtual environments will be possible, in which sounds will be generated through the modeling of the physical interaction between objects and will be correctly spatialized and auralized.

It is important to mention that a first example of virtual audio reality system has already been developed within the EC-sponsored DIVA project [36], [62], [47]. This work already constitutes a significant step forward in the direction of audiovisual integration, although full-scale interactivity and object-based solutions are not yet a reality in the achieved results. We are currently working on a full-scale implementation of the layered structure for audiovisual environments as presented in this article. In particular, we are focusing on the SRL and its interactions with the SGL.

One concluding remark on future directions of research is that there is a strong need of understanding what kind of approximations can be made on the synthesis-rendering structure in both SGL and SRL without significantly impacting on human perception of sound quality. As we know, there is virtually no limit in the level of detail that can be used in modeling physical reality. This is true for both spatialization and sound synthesis. A better understanding of the link between perceptual redundancy and physical accuracy of the models would thus help us simplify the implementation of audiovisual environments.

Acknowledgments

We wish to thank Dr. Enrico Bianchi of Phoenix Tools Ltd. for the fruitful discussions on problems of photo-realistic and audiorealistic rendering.

Federico Pedersini was born in Brescia, Italy, in 1965. He received the "laurea" degree (summa cum laude) in electrical engineering in 1991 and the doctoral degree in electrical engineering and communications in 1994, both from the Politecnico di Milano, Milan, Italy. He spent one year at the Intitut fur Theoretische Nachrichtentechnik of the University of Hannover, Germany, working on camera calibration for 3-D reconstruction systems and accurate modeling of camera imaging process. He is currently with the Politecnico di Milano, where he is doing research on image processing for 3-D measurement, reconstruction, and image synthesis.

Augusto Sarti was born in Rovigo, Italy, in 1963. He received the "laurea" degree (summa cum laude) in electrical engineering in 1988 and a doctoral degree in electrical engineering and information sciences in 1993, both from the University of Padova, Padua, Italy. He worked for one year for the Italian National Research Council, doing research on digital radio systems. He then spent two years doing research on nonlinear system theory at the University of California, Berkeley. He is currently an Assistant Professor at the Politecnico di Milano, Milan, Italy, and his research interests are in digital signal processing and, in particular, in video coding, image analysis for 3-D scene reconstruction, and audio processing and synthesis.

Stefano Tubaro was born in Novara, Italy, in 1957. He completed his studies in electrical engineering in 1982. He joined the Politecnico di Milano, Dipartimento di Elettronica e Informazione in 1984, doing research on voice and image coding. In 1986 he joined the CSTS (Study Center for Space Telecommunications) of the CNR (National Research Council). Since 1991 he has been an Associate Professor of Electrical Communications at the Politecnico di Milano. His current research interests focus on digital image processing and, in particular, on video sequence coding at low bit-rate through motion estimation and image segmentation. He also works on stereovision for 3-D scene reconstruction applied to remote manipulation, autonomous vehicle guidance, and telepresence.

References

[1] J. Allen and D. Berkley, "Image method for efficiently simulating small-room acoustics," *J. Acoustical Soc. Amer.*, vol. 65, no. 4, pp. 912-915, Apr. 1979.

[2] P. Beardsley, P. Torr, and A. Zisserman, "3-D model acquisition from extended image sequences," in *Proc. ECCV '96*, Cambridge, UK, Apr. 14-18, 1996, vol. 2, pp. 683-695.

[3] V. Belevitch, *Classical Network Theory*. San Francisco, CA: Holden Day, 1968.

[4] A. Blake and A. Zisserman, *Visual Reconstruction*. Cambridge, MA: MIT Press, 1987.

[5] G. Borin, G. De Poli, and A. Sarti, "A modular approach to excitator-resonator interaction in physical models synthesis," in *Proc. Int. Computer Music Conf.*, Columbus, OH, Nov. 1990, pp. 46-50.

[6] G. Borin, G. De Poli, and A. Sarti, "Sound synthesis by dynamic systems interaction," in *Readings in Computer-Generated Music*, D. Baggi, Ed. Los Alamitos, CA: IEEE Computer Society Press, 1992, pp. 139-160.

[7] G. Borin, G. De Poli, and A. Sarti, "Algorithms and structures for synthesis using physical models," *Computer Music J.*, vol. 19, no. 4, pp. 30-42, 1992.

[8] G. Borin, G. De Poli, and A. Sarti, "Musical signal synthesis," in *Musical Signal Processing*, C. Roads, S. Pope, A. Piccialli, and G. DePoli, Eds. Lisse, Netherlands: Swets and Zeitlinger, 1996, pp. 5-30.

[9] J. Borish, "Extension of the image model to arbitrary polyhedra," *J. Acoust. Soc. Amer.*, vol. 75, no. 6, pp. 1827-1836, June 1984.

[10] K. Brandenburg, "Perceptual coding of high quality digital audio," in *Applications of Digital Signal Processing to Audio and Acoustics*, M. Kahrs and K. Brandenburg, Eds. Norwell, MA: Kluwer, 1998, pp. 85-131.

[11] L.O. Chua, "Synthesis of new nonlinear network elements," *Proc. IEEE*, vol. 56, pp. 1325-1340, Aug. 1968.

[12] L.O. Chua, *Introduction to Nonlinear Network Theory*. New York: McGraw-Hill, 1969.

[13] L.O. Chua, "Memristor—The missing circuit element," *IEEE Trans. Circuit Theory*, vol. CT-18, pp. 507-519, Sept. 1971.

[14] L.O. Chua, "Device modeling via basic nonlinear circuit elements," *IEEE Trans. Circuits Syst.*, vol. CAS-27, pp. 1014-1044, Nov. 1980.

[15] L.O. Chua, "Dynamic nonlinear networks: State-of-the-art," *IEEE Trans. Circuits Syst.*, vol. CAS-27, pp. 1059-1087, Nov. 1980.

[16] L.O. Chua, "Nonlinear circuits," *IEEE Trans. Circuits Syst.*, vol. CAS-31, pp. 69-87, Jan. 1984.

[17] M.F. Cohen and J.R. Wallace, *Radiosity and Realistic Image Synthesis*. Reading, MA: Academic, 1993.

[18] R.L. Cook, T. Porter, and L. Carpenter. "Distributed ray tracing," in *Proc. SIGGRAPH'84*, vol. 18, pp. 137-146.

[19] N. Dadoun, D.G. Kirkpatrick, and J.P. Walsh. "The geometry of beam tracing," in *Proc. Symp. Computational Geometry*, Baltimore, MD, June, 1985, pp. 55-71.

[20] G. De Poli, "A tutorial on digital sound synthesis techniques," in *The Music Machine*, C. Roads, Ed. Cambridge, MA: MIT Press, 1992, pp. 429-447.

[21] T. Felderhoff, "Simulation of nonlinear circuits with period doubling and chaotic behavior by wave digital principles," *IEEE Trans. Circuits Syst. I*, vol. CAS-41, pp. 485-491, July 1994.

[22] T. Felderhoff, "A new wave description for nonlinear elements," *IEEE Int. Symp. Circuits and Systems*, Atlanta, GA, May 12-15, 1996, pp. 221-224.

[23] A. Fettweis, "Wave digital filters: Theory and practice," *Proc. IEEE*, vol. 74, pp. 327-270, Feb. 1986.

[24] A. Fettweis, "Some principles of designing digital filters imitating classical filter structures," *IEEE Trans. Circuit Theory*, vol. CT-18, pp. 314-316, Mar. 1971.

[25] A. Fettweis, "Pseudopassivity, sensitivity, and stability of wave digital filters," *IEEE Trans. Circuit Theory*, vol. CT-19, pp. 668-673, Nov. 1972.

[26] A. Fettweis and K.Meerkötter, "On adaptors for wave digital filters," *IEEE Trans. Acoustic, Speech, Signal Processing*, vol. ASSP-23, pp. 516-525, Dec. 1975.

[27] J.D. Foley, A. van Dam, S.K. Feiner, and J.F. Hughes, *Computer Graphics: Principles and Practice*. 2nd ed. in C, Reading, MA: Addison-Wesley, 1996.

[28] F. Fontana and D. Rocchesso, "Physical modeling of membranes for percussion instruments," *Acustica - Acta Acustica*, vol. 84, no. 3, pp. 529-542, May-June 1998.

[29] T. Funkhouser, I. Carlbom, G. Elko, G. Pingali, M. Sondhi, and J. West, "A beam tracing approach to acoustic modeling for interactive virtual environments," in *Proc. SIGGRAPH'98*, pp. 21-32.

[30] W. Gardner, "Reverberation algorithms," in *Applications of Digital Signal Processing to Audio and Acoustics*, M. Kahrs and K. Brandenburg, Eds. Norwell, MA: Kluwer, 1998, pp. 85-131.

[31] A.S. Glassner, *An Introduction to Ray Tracing*. New York: Academic, 1989.

[32] C.M. Goral, K.E. Torrance, D.P. Greenberg, and B. Battaile, "Modeling the interaction of light between diffuse surfaces." in *Proc. SIGGRAPH'84*, vol. 18, pp. 213-222.

[33] R. Hall, *Illumination and Color in Computer-Generated Imaging*. New York: Springer-Verlag, 1989.

[34] P. Heckbert and P. Hanrahan, "Beam tracing polygonal objects," in *Proc. SIGGRAPH'84,*, vol. 18, pp. 119-127.

[35] M. Hodgson, "Evidence of diffuse surface reflections in rooms," *J. Acoust. Soc. Amer.*, vol. 89, pp. 765-771, 1991.

[36] J. Huopaniemi, L. Savioja, and T. Takala, "DIVA virtual audio reality system," in *Proc. Int. Conf. Auditory Display, (ICAD'96)*, Palo Alto, CA, Nov. 4-6, 1996, pp. 111-116.

[37] C.B. Jones, "A new approach to the 'hidden line' problem," *Computer J.*, vol. 14, no. 3, pp. 232-237, Aug. 1971.

[38] J.T. Kajiya, "The rendering equation," in *Proc. SIGGRAPH'86*, vol. 20, pp. 143-150.

[39] T. Kanade, P. Rander, and P.J. Narayanan, "Virtualized reality: Constructing virtual worlds from real scenes." *IEEE Multimedia*, vol. 4, pp. 34-47, Jan.-Mar. 1997.

[40] H. Krauß and R. Rabenstein, "Application of multidimensional wave digital filters to boundary value problems," *IEEE Signal Processing Lett.*, vol. 2, pp. 138-140, July 1995.

[41] H. Krauß, R. Rabenstein, and M. Gerken, "Simulation of wave propagation by multidimensional digital filters," *J. Simulation Practice Theory*, vol. 4, pp. 361-382, 1996.

[42] U.R. Krockstadt, "Calculating the acoustical room response by the use of a ray tracing technique," *J. Sound Vibr.*, no. 8, vol. 18, 1968.

[43] H. Kuttruff, *Room Acoustics*, 3rd ed. New York: Elsevier Science, 1991.

[44] H. Kuttruff, "A simple iteration scheme for the computation of decay constants in enclosures with diffusely reflecting boundaries," *J. Acoustic Soc. Amer.*, vol. 98, no. 1, pp. 288-293, July 1995.

[45] H. Lehnert, "Systematic errors of the ray-tracing algorithm," *J. Applied Acoustics*, vol. 38, nos. 2-4, pp. 207-221, 1993.

[46] T. Lewers, "A combined beam tracing and radiant exchange computer model of room acoustics," *J. Applied Acoustics*, vol. 38, nos. 2-4, pp. 161-178, 1993.

[47] T. Lokki, L. Savioja, and J. Hiipakka, "Immersive 3-D sound reproduction in virtual room," in *Proc. AES 16th Int. Conf. Spatial Sound Reproduction*, Rovaniemi, Finland, Apr. 10-12, 1999, pp. 172-177.

[48] D.C. Massie, "Wavetable sampling synthesis," in *Applications of Digital Signal Processing to Audio and Acoustics*, M. Kahrs and K. Brandenburg, Eds. Norwell, MA: Kluwer, 1998, pp. 85-131.

[49] K. Meerkötter and R. Scholtz. "Digital simulation of nonlinear circuits by wave digital filter principles," in *Proc. IEEE Int. Symp. Circuits and Systems*, Portland, OR, May 8-11, 1989, vol. 1, pp. 720-723.

[50] F. Pedersini, A. Sarti, and S. Tubaro, "Block-wise physical model synthesis for musical acoustics," *IEE Electron. Lett.*, vol. 35, no. 17, pp. 1418-1419, Aug. 1999.

[51] F. Pedersini, A. Sarti, S. Tubaro, and R. Zattoni, "Toward the automatic synthesis of nonlinear wave digital models for musical acoustics," in *Proc. IX Euro. Signal Processing Conf.*, Sept. 8-11, 1998, Rhodes, Greece, vol. IV, pp. 2361-2364.

[52] F. Pedersini, P. Pigazzini, A. Sarti, and S. Tubaro, "3-D area matching with arbitrary multiview geometry," in *Proc. EURASIP Signal Processing: Image Communications*, vol. 14, nos. 1-2, pp. 71-94, 1998.

[53] F. Pedersini, A. Sarti, and S. Tubaro, "multicamera systems: calibration and applications," *IEEE Signal Processing Mag.*, vol. 16, pp. 55-65, May 1999.

[54] F. Pedersini, A. Sarti, and S. Tubaro, "Visible surface reconstruction with accurate localization of object boundaries," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 10, pp. 278-291, Mar. 2000.

[55] F. Pedersini, P. Pigazzini, A. Sarti, and S. Tubaro, "Multicamera motion estimation for high-accuracy 3-D reconstruction." *EURASIP Signal Processing*, vol. 80, no. 1, pp. 1-21, Jan. 2000.

[56] R. Rabenstein and H. Krauß: "Discrete simulation of uniform transmission lines by multidimensional digital filters." *Int. J. Numer. Model.*, vol. 9, pp. 271-294, 1996.

[57] D. Rocchesso J and O. Smith: "Circulant and elliptic feedback delay networks for artificial reverberation," *IEEE Trans. Speech Audio Processing*, vol. 5, no. 1, pp. 51-63, Jan. 1997.

[58] D. Rocchesso, "Maximally diffusive yet efficient feedback delay networks for artificial reverberation," *IEEE Signal Processing Lett.*, vol. 4, no. 9, pp. 252-255, Sept. 1997.

[59] D. Rocchesso and F. Scalcon, "Bandwidth of perceived inharmonicity for physical modeling of dispersive strings," *IEEE Trans. Speech Audio Processing*, vol. 7, pp. 597-601, Sept. 1999.

[60] A. Sarti and G. De Poli, "Generalized adaptors with memory for nonlinear wave digital structures." in *Proc. VIII European Signal Processing Conf.*, Trieste, Italy, 1996, vol. 3, pp. 1773-1776.

[61] A. Sarti and G. De Poli, "Toward nonlinear wave digital filters," *IEEE Trans. Signal Processing*, vol. 47, pp. 1654-1668, June 1999.

[62] L. Savioja, J. Huopaniemi, T. Lokki, and R. Väänänen, "Virtual environment simulation advances in the DIVA project," in *Proc. Int. Conf. Auditory Display (ICAD'97)*, Palo Alto, CA, Nov. 3-5, 1997, pp. 43-46.

[63] J.O. Smith, "Acoustic modeling using digital waveguides," in *Musical Signal Processing*, C. Roads, S.T. Pope, A. Piccialli, and G. DePoli, Eds. Lisse, The Netherlands: Swets and Zeitlinger, 1997, pp. 221-263.

[64] J.O. Smith, "Principles of digital waveguide models of musical instruments," in *Applications of Digital Signal Processing to Audio and Acoustics*, M. Kahrs and K. Brandenburg, Eds. Norwell, MA: Kluwer, 1998, pp. 417-466.

[65] S.J. Teller and C.H. Séquin, "Visibility preprocessing for interactive walkthroughs," in *Proc. SIGGRAPH'91*, vol. 25, pp. 61-69.

[66] D. Terzopoulos and A. Witkin, "Physically based models with rigid and deformable components," *IEEE Computer Graphics and Applications*, vol. 8, no. 6, pp. 41-51, 1988.

[67] D. Terzopoulos and D. Thalmann, *Computer Animation and Simulation*. New York: Springer-Verlag, 1995.

[68] http://www.3-Dscanners.com/

[69] N. Tsingos and J. Gascuel, "A general model for simulation of room acoustics based on hierarchical radiosity, technical sketches," in *SIGGRAPH 97 Visual Proc.*, 1997.

[70] Väänänen, V. Välimäki, J. Huopaniemi, and M. Karjalainen, "Efficient and parametric reverberator for room acoustics modeling," in *Proc. Int. Computer Music Conf. (ICMC'97),* Thessaloniki, Greece, Sept. 25-30, 1997, pp. 200-203.

[71] S.A. Van Duyne, J.R. Pierce, and J.O. Smith, "Traveling-wave implementation of a lossless mode-coupling filter and the wave digital hammer," in *Proc. 1994 Intl. Computer Music Conf.*, 1994, pp. 411-418.

[72] L. Van Gool and A. Zisserman, "Automatic 3-D model building from video sequences," *Euro. Trans. Telecommun.*, vol. 8, no. 4, pp. 369-378, July-Aug. 1997.

[73] T. Whitted, "An improved illumination model for shaded display," *Commun. ACM*, vol. 23, no. 6, pp. 343-349, June 1980.