Sound models for synthesis: a structural viewpoint

Giovanni De Poli

Centro di Sonologia Computazionale - University of Padova

1. Introduction

At the beginning of the 20th century, some musicians had already started to turn their attention to the search for new forms of sonority. They were of the opinion that the new technologies being developed would not only enhance the evolution of existing instruments, but that these technologies were also a potential source of alternative sounds that were unlike traditional sounds. They sought to identify these, and therefore, stimulate new organizational criteria in composing music.

In the scientific field, the development of new methods connected to information technology offered a growing number of instruments that, even though they were designed for other applications, could also be used to produce sounds. The combination of these two factors and the enthusiastic collaboration between musicians and researchers led to an intensive research activity and experimentation on new sounds. After an initial period during which only a few pioneers went ahead in almost complete isolation, in the Seventies a small community strongly felt the need to meet and join together. Over the years, the study of sound, and above all, producing sound by new methods, has become focal points of attention for researchers and musicians. This considerable interest is reflected in the names of the computer music centers that have arisen in that period; for example, the Institute of Sonology in Utrecht, the Center for Computer Research in Music and Acoustics (CCRMA) at Stanford University the Institute for Research and Coordination of Acoustics and Music (IRCAM) in Paris, and the Centro di Sonologia Computazionale (CSC) in Padova.

The underlying hope was that, using digital technology, it would be possible to generate any sound that the human ear can hear. But soon it became clear that while any sound, once recorded, can be reproduced, a new sound can be generated only when a computing procedure (i.e. a synthesis algorithm) can be described for its generation.

This idea gave great impulse to the search for algorithms and models for sound synthesis and their successive utilization in creating music. In a certain sense, there

Musica/Tecnologia, 16 (2022), ISSN online 1974-0050, ISSN print 1974-0042, pp. 7-75
DOI: 10.36253/music_tec-14417 – http://www.fupress.com/mt

was a tendency to identify the synthesis technique with the concept of instrument, not only as a method for generating sounds, but also as something that describes a class of sonority. In fact, the same synthesis model can often produce many different sounds, all of which have a common method of production and therefore share some acoustical properties.

A sound model is implemented by means of sound synthesis and processing techniques. A wide variety of sound synthesis algorithms is currently available either commercially or in the literature. Each one of them exhibits some peculiar characteristics that could make it preferable to others, depending on goals and needs. Technological progress has made enormous steps forward as far as the computational power that can be made available at low cost is concerned. At the same time, sound synthesis methods have become more and more computationally efficient, and the user interface has become friendlier and friendlier. Therefore, musicians can nowadays access a wide collection of synthesis techniques (all available at low cost in their full functionality) and concentrate on their timbre properties.

Sounds within a class differ according to the parameters provided to the synthesis model. Using the basic set of parameters produces the basic sounds of that particular class. However, when looking for richer and more interesting musical sounds, different and well-calibrated parameter sets should be used. The problem becomes one of knowing how to describe the desired sound in terms of the parameters of the chosen model. This so-called *synthesis control* problem requires special attention in being dealt with. If a synthesis model is compared to an instrument, then much experimentation is required to explore the class of sounds that can be produced and to understand how to obtain them. Furthermore, a great deal of time is necessary to learn how to play the instrument and the process of experimental creativity takes even longer.

Aim of this paper is to review some of the main computational models that are being used in musical sound production. In this work, the focus is on sound generation and not on sound processing, i.e. systems for sound transformation. This article is organized as follow. Section 2 discusses how the different approaches to modeling can be conceptually organized. Section 3 presents the main strategies to model the sound as it reaches our ear, regardless of the physical mechanisms underlying the sound production. The focus is on the listener. Section 4 deals with modeling the source of a sound and obtaining its synthesis through a simulation of the physical phenomena that produce sound. The focus is on the source. Section 5 discusses how synthesis models can be re-interpreted as control signal models. Section 6 briefly summarizes the main scientific and musical research on sound synthesis at Centro di Sonologia Computazionale during the Seventies and the Eighties. Finally, section 7 draws some conclusions and perspectives.

2. Models and sound synthesis

Sound synthesis is a procedure to produce a sound, without the help of acoustic instruments, where no real-time acoustic input is used (see Fig. 1, left). In digital

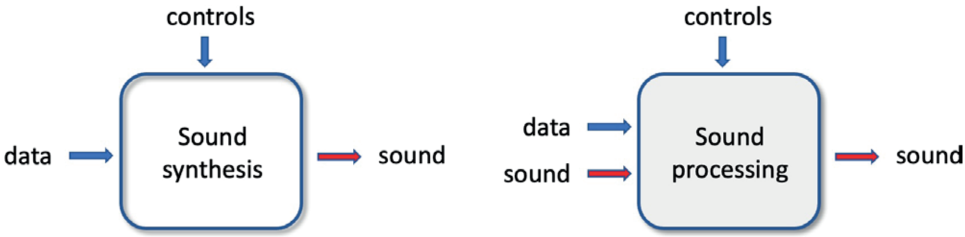

Figure 1. Sound synthesis vs. sound processing.

synthesis, a sound is represented by a sequence of numbers (samples). Hence, a digital synthesis technique consists of a computing procedure or a mathematical expression, which computes each sample value. Normally, the synthesis formula depends on some values, that is, *control parameters*. Frequency and amplitude are examples of such parameters. Parameters can be constant or slowly time variant during the sound. Time-variant parameters are also called *control functions*.

In early times, the performer's gesture was not taken into account: the composition of the sound and the composition with sounds replaced the traditional performance with instruments. In the 1980s, real-time music systems started to be realized, which allowed an increasingly effective interaction between the performer, the machine and the listener and fostered new performance practices. Thus, gestures started being used for synthesis control.

Audio can be a direct input to the computer system in the case of live electronics or *sound processing* systems (see Fig. 1, right). *Live electronics* is a musical practice where sounds are processed live with an electroacoustic system; acoustic and electroacoustic sounds are present simultaneously; the electroacoustic system becomes an extension of the voice and/or the instrument, i.e., the interaction between the acoustic and electronic performer. Only sound synthesis will be treated in this paper. A good overview on sound processing techniques can be found in (Zölzer, 2011).

2.1 Models

A useful approach for dealing with complexity is to use a *model* to evidence and abstract some relations that can be hypothesized, discarding details that are felt to be irrelevant for what is being observed and described. In this way models allows selective examination with the essential aspects. Models can be used to predict behavior in certain conditions and compare these results with observations. In this sense, they serve to generalize the findings and have both a descriptive and predictive value. Yet, an abstraction is task-dependent and it is used for a particular purpose, which in turn determines what is important and what can be left out. Thus, there are several ways to describe a phenomenon.

One level of abstraction allows us to derive *mathematical models*, which describe relations in models is by using mathematical expressions composed of observable (and often measurable) facts called variables or parameters. Developing and then validating mathematical models is the typical way to proceed in science and engineering. Often the variables are divided into input variables, supposedly known, and output variables, which are deduced by the model: the mathematical model describes the relations between input and output variables. In this case, inputs can be considered as the causes and output the effect of the phenomenon. For example, in a sound synthesis model (Fig. 1, left), the input variables are data and controls, while the output is the sound to be generated.

Once the equations are discretized, a *computational model* can compute the values of output variables corresponding to the provided values of inputs. This process is called simulation and it is widely used to predict the behavior of the phenomenon in different circumstances. Computational models inherit the abstractions of mathematical models and add one more level of abstraction by imposing a *synthesis algorithm* for solving them. Among many possible choices, digital signal processing (DSP) provides an advanced theory and tools that emphasize computational issues, particularly maximal efficiency, for sound synthesis.

2.1.1 Computational models as sound abstraction

In order to generate, manipulate, and think about sounds, it is useful to organize our intuitive sound abstractions into objects, in the same way as abstract categories are needed for defining visual objects. The first extensive investigation and systematization of sound objects from a perceptual viewpoint was done by Pierre Schaeffer in the Fifties (Schaeffer, 1966).

For effective generation and manipulation of sound objects it is necessary to define models for sound synthesis, processing, and composition. Identifying models, either visual or acoustic, is equivalent to making high-level constructive interpretations, built up from the zero level (i.e. pixels or sound samples). It is important for the model to be associated with a *semantic interpretation*, in such a way that an intuitive action on model parameters becomes possible.

Each sound synthesis algorithm can be thought of as a computational model for the sound itself. Though this observation may seem quite obvious, its meaning for sound synthesis is not so straightforward. As a matter of fact, modeling sounds is much more than just generating them, as a computational model can be used for representing and generating a whole class of sounds, depending on the choice of control parameters. The idea of associating a class of sounds to a digital sound model is in complete accordance with the way we tend to classify natural musical instruments according to their sound generation mechanism. For example, strings and woodwinds are normally seen as timbre classes of acoustic instruments characterized by their sound generation mechanism.

It should be clear that the degree of compactness of a class of sounds is determined, on one hand, by the sensitivity of the digital model to parameter variations and, on the other hand, the amount of control that is necessary to obtain a certain desired sound. As an extreme example we may think of a situation in which a musician is required to generate sounds sample by sample, while the task of the computing equipment is just that of playing the samples. In this case the control signal is represented by the sound itself, therefore the class of sounds that can be produced is unlimited, but the instrument is impossible for a musician to control and play. An opposite extreme situation is that in which the synthesis technique is actually the model of an acoustic musical instrument. In this case the class of sounds that can be produced is much more limited (it is characteristic of the mechanism that is being modeled by the algorithm), but the degree of difficulty involved in generating the control parameters is quite modest, as it corresponds to physical parameters that have an intuitive counterpart in the experience of the musician.

2.1.2 Musical objectives in computational models

Technological limitations make us value a series of compromises we must take into account when designing or using a specific synthesis technique.

- *Sound quality.* By sound quality we mean the internal richness of sound. A sound with a great quality would be a natural sound while at the other extreme we could have a simple sound, electronically synthesized, with no microvariation during its duration.
- *Flexibility.* This term describes the ability of a specific synthesis technique to modify sound from a series of control parameters. With this criterion, a sampler would not be a very flexible instrument, and frequency modulation synthesis would be very flexible.
- *Generality.* By generality we understand the possibility of one synthesis technique to generate a great many timbres. Additive synthesis would be a very general technique and the recording of a sound would be very specific.
- *Robustness* concerns with how well the model succeeds in generating a family of perceptual related sounds, retaining the sound identity in the context of parameter variations. Physical modeling is very robust, while sampling is the opposite.
- *Playability* refers to the musician needs of an intuitive and easy access to the control parameters during both the sound design process and the performance. A physics-based model is very playable from a performance point of view, while it is the opposite from a design point of view.
- *Efficiency* refers to the number of computer instructions needed to generate each of the sound samples synthesized and to memory requirements. In this sense, frequency modulation synthesis is a very economical technique and additive synthesis requires much more computing time and memory.

An interesting conclusion that could be already drawn in the light of what we stated above is that the generality of the class of sounds associated to a sound synthesis algorithm is somehow in contrast with the playability of the algorithm itself. One should remember that the playability is of crucial importance for the success of a specific sound synthesis algorithm as, in order for a sound synthesis algorithm to be suitable for musical purposes, the musician needs an intuitive and easy access to its control parameters during both the sound design process and the performance. Such requirements often represent the reason why a certain synthesis technique is preferred to others.

From a mathematical viewpoint, the musical use of sound models opens some interesting issues: description of a class of models that are suitable for the representation of musically relevant acoustic phenomena; description of efficient and versatile algorithms that realize the models; mapping between meaningful acoustic and musical parameters and numerical parameters of the models; analysis of sound signals that produces estimates of model parameters and control signals; approximation and simplification of the models based on the perceptual relevance of their features; generalization of computational structures and models in order to enhance versatility.

2.2 Classification of sound models

Sound models can be classified from different points of view. The reason is that each classification has been introduced to best meet the needs of a specific audience; it then relies on a series of features and can be useful in different contexts.

2.2.1 Classification based on model structure

In the case of computers and digital instruments the generation mechanism is represented by the synthesis algorithm. We can classify the main classes of algorithms from the internal *model structure* point of view, pointing out the ways in which simple elements are used to obtain complex dynamic behavior.

- *Direct generation* which includes all the techniques that are based on one or more independently operating blocks and on the sum of the results. Examples are additive and granular synthesis and sampling.
- *Feed-Forward Structure* that is given by the feed-forward multi-block structures in which some blocks generate the signal and supply it to the other blocks for post-processing. This class includes all the linear and nonlinear transformation models such as subtractive synthesis, ring and frequency modulation synthesis.
- *Interacting Structure* that is characterized by an interacting multi-block structure. The simplest example of this scheme consists of a pair of blocks, which result in a feed-back connection. In this model, the dynamics of the sound are mainly due to the interaction between the blocks. If the system is completely linear the overall behavior is reduced to a linear filter. However, often a feature of this approach is the presence of nonlinear elements that give rise to complex behavior. Most physics-based models belong to this class.

This approach is in line with the classic Hornbostel-Sachs music instruments categorization system, which is based on how an instrument vibrates to produce sound (Sachs, 1940).

2.2.2 Classification based on cognitive representation

Gaver (1993) introduced the distinction between musical listening and everyday listening. *Musical listening* focuses on perceptual attributes of the sound itself (e.g., pitch, loudness), whereas *everyday listening* focuses on events to gather relevant information about our environment, that is, not about the sound itself but rather about sound sources and actions producing sound.

It is possible to carry out a classification of the synthesis algorithms on the basis on what the model aims to represent. In this case, we can distinguish:

- *Signal based models* : these models represent the shape of the sound wave that reaches our ear, without any reference to the physical mechanism underlying the production of sound (Fig. 2, right). They therefore refer to the properties of acoustic perception. The evaluation of a signal model should be done mainly according to perceptual cues. The focus is on the sound receiver, i.e. the human ear.
- *Physics based models* : these models obtain an acoustic signal as a by-product of a model simulating the physical mechanism of production of sound (Fig. 2, left). Physics-based models are better evaluated according to the physical behaviors involved in the sound production process. The focus is on the sound source or emitter.

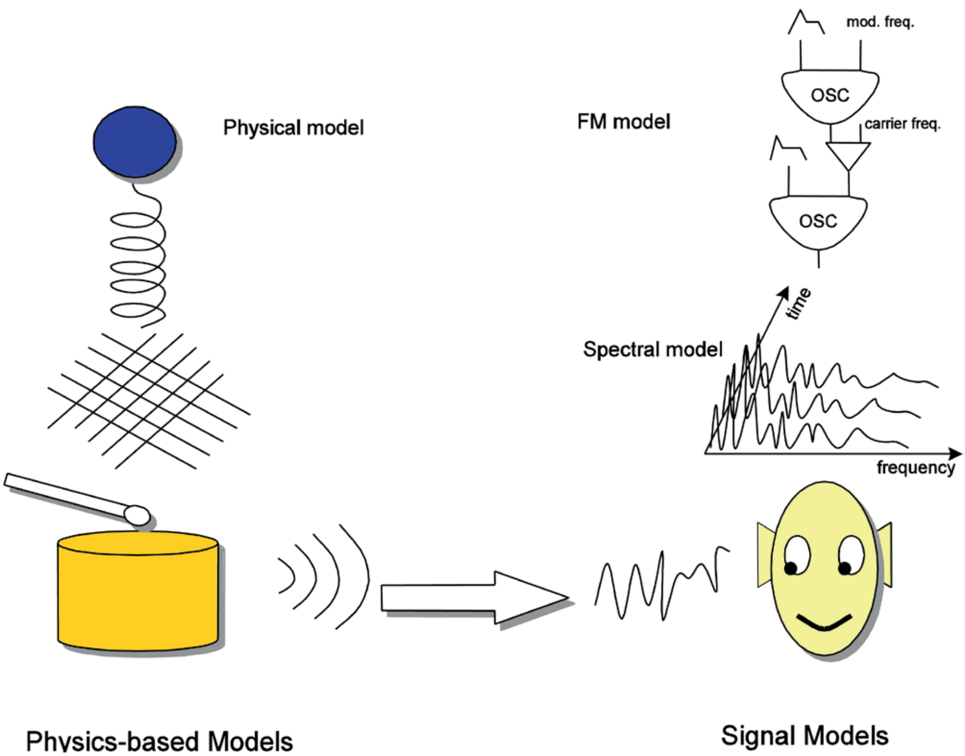

Figure 2. Signal vs. physics-based models.

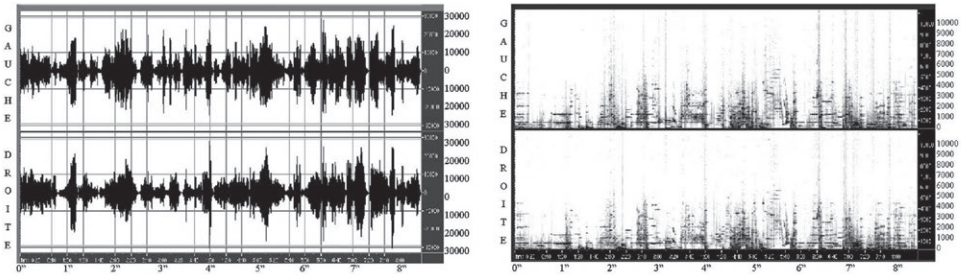

Figure 3. Time domain (left) and frequency domain (right) representation of an excerpt of *Winter Leaves* by Mauro Graziani (Zattra, 2004).

2.2.3 Classification based on underlying techniques

Classification based on processing domain Another classification of synthesis algorithms is based on the domain where the signal processing is applied (namely time or frequency) and on the perceptual interpretation of the user, since our ear is approximately a spectral analyzer.

- *Time domain models*, which are best interpreted in term of their time characteristics.
- *Frequency domain models*, which are best interpreted in term of their spectral characteristics.

Figure 3 shows the time domain (left) and frequency domain (right) representation of an excerpt of *Winter Leaves* by Mauro Graziani. Being both algorithmically and semantically significant, we will use this distinction in the presentation of signal-based models.

Linearity In earlier days, sound synthesis techniques were usually divided into two categories with respect to the *linearity* of generative operations, thus saying that filters are linear transformation while modulations and distortions are non-linear techniques. 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. This feature was very stimulating for composers.

Sample vs block processing Signal processing algorithms usually process signals by either block processing or sample-by-sample processing. For block processing, samples are first collected in a memory buffer and then processed each time the buffer is completely filled with new data. Examples of such algorithms are fast Fourier transforms (FFTs) for spectra computations and fast convolution. In sample processing algorithms, each input sample is processed on a sample-by-sample basis, as for example happens in frequency modulation synthesis.

Hardware vs. software implementation A final distinction can be made regarding the technology used in implementing the computing models. At the beginning, synthesis was done *offline* with *software* developed specifically for the specific computational model and then with modular languages, such as MusicV and derivatives, that allowed a general environment for synthesis (Mathews et al., 1969). In the Eighties the advancement of technologies permitted the development of *real-time hardware* processors, which allowed for the sound synthesis and transformation in real-time. Finally, today most synthesis is performed in real time on *general purpose computers*.

3. Sound modeling: signal-based approaches

With the words *signal-based models*, we want to indicate the whole family of synthesis algorithms which are aimed to modeling sound as it reaches the ear, regardless of the physical mechanisms underlying the sound production. Sound perception is a complex phenomenon, involving a signal analysis in both frequency and time domains. Signal models can be divided into two classes, according to the natural interpretation given by the user in terms of time or spectral characteristics. Along with this dichotomy, we can say that sampling and granular synthesis are *time-domain models;* while additive and subtractive synthesis, frequency modulation, and non-linear distortion can be described as *frequency-domain models*.

Let us briefly illustrate the algorithms which are, in our opinion, the most significant in the field of sound synthesis. The description will be given at a generic high level, paying special attention to music implications, and leaving to the vast literature the task of describing mathematical and implementation details.

3.1 Time domain models

To analyze the various methods of audio signals it is best to group them according to structural and functional parameters. A fundamental characteristic of a musically interesting sound is its complex dynamic behavior. It depends on external control and on internal structure of the sound generating process. Internal structure sets limits and suggests usage. Thus, it is a good starting point for classification. Therefore, an initial distinction shall be made by analyzing the way in which simple elements are combined to obtain complex dynamic behaviors.

3.1.1 Basic generators

A synthesis technique which directly produce a simple signal from given data, is called *Generator*. In this class, which we shall call direct, we include all the techniques that are based on one or more blocks which operate independently among themselves (Fig. 4).

Techniques of digital synthesis inherited the knowledge developed for synthesis by analog means and introduced the concept of the unit generator as a digital version of

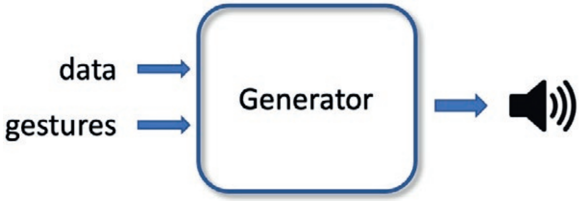

Figure 4. Generator module.

analog devices or the modules of analog synthesizers. A unit generator is a fundamental building block, or module, used to build a variety of sound-generating algorithms. Each unit generator accepts numeric parameters and/or audio signal(s) as input and produces an output signal. Basic signal processing and synthesis modules could be interconnected to create interesting synthetic sounds. We will see the oscillator for periodic signals and the noise generator for random signals.

The Wavetable oscillator or Table lookup oscillator In many musical sounds, pitch is a characteristic to which we are quite sensitive. In examining the temporal shape of pitched sounds, we see a periodic repetition of the waveform without great variations. The simplest synthesis method attempts to reproduce this characteristic, generating a periodic signal through continuous repetition of the waveform. The technique is carried out by a module called an oscillator, which repeats the waveform with a possibly time-varying amplitude and frequency. Usually, in digital synthesis the waveform value at a particular instant is not computed anew for each sample. Rather, a table, containing the period values computed in equally spaced points, is built beforehand. Obviously, the more points in the table, the better the approximation. The *wavetable oscillator*, or *table look-up oscillator*, works by circularly accessing the wavetable at multiples of an increment, proportional to the instantaneous frequency, and reading the wavetable content at that position.

The oscillator output is multiplied by an amplitude envelope. The instantaneous frequency of the oscillator can be varied enabling the production of a tremolo and, with wider variations, of a glissando or melodies. The waveform is fixed, while the amplitude and frequency vary (Fig. 5). The partials are exact multiples of the fundamental, and they all behave the same. By changing the table, signals with different waveform can be generated (Mathews et al., 1969).

It is employed when good sound quality is not required. The constant waveform gives the sound a mechanical, dull, and unnatural character, which soon annoys the audience. Thus, in musical applications, fixed-waveform synthesis is not very effective when used alone. It is employed for its simplicity when timbre variety is not required or as a basic building block for transformation techniques, such as sinusoidal generators for frequency modulation. It is used also for generating control functions by storing in the table the whole function and reading it once.

An extension of this model is the *piecewise linear function generator*, where the relevant break points are stored in the table, and the continuous function is obtained by

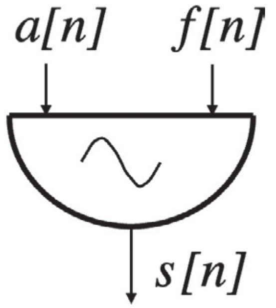

Figure 5. Wavetable oscillator, where amplitude and frequency are time-varying. The waveform to be repeated is indicated inside the module.

linear interpolation. It often used as envelope generator or as amplitude and frequency control function in additive synthesis.

It is possible to add several wavetable units together with independent vibrato and slightly detuned fundamental frequencies to obtain a chorus-like effect. Cross fading between wavetables is a convenient way to obtain an evolving timbre (*multiple wavetable synthesis)*.

Noise generators Signals whose behavior at any instant is supposed to be perfectly knowable are called deterministic signals. Besides these signals, *random signals* of unknown or only partly known behavior may be considered. For random signals, only some general characteristics, called statistical properties, are known or are of interest. The statistical properties are characteristic of an entire signal class rather than of a single signal. A set of random signals is represented by a random process. Particular numerical procedures simulate random processes, producing sequences of random (or more precisely, pseudorandom) numbers.

Random sequences can be used both as signals (e.g., to produce white or colored *noise* used as input to a filter) and as control functions, to provide a variety in the synthesis parameters most perceptible to the listener. The most common algorithm to generate uniformly distributed random numbers is the so-called *linear congruential generator*. To obtain low-pass noise or sequences that vary more slowly, one can generate a new random number not at each time instant and keep the output constant in the interval (holder) or perform a linear interpolation between the two generated values. The longer the interval between two generated numbers, the lower the cutoff frequency.

In the analysis of natural sounds, some characteristics vary in an unpredictable way; their mean statistical properties are perceptibly more significant than their exact behavior. Hence, the addition of a random component to the deterministic functions controlling the synthesis parameters is often desirable. In general, a combination of random processes is used because the temporal organization of the musical parameters often has a hierarchical aspect. It cannot be well described by a single random process, but rather by a combination of random processes evolving at different rates. For example, this technique is employed to generate $1/f$ noise (*pink noise)*, which is characterized by a power spectrum that fall in frequency like the inverse of the frequency.

3.1.2 Time-segment based models

Sampling synthesis Finding a mathematical model that faithfully imitates a real sound is an extremely difficult task. If an existing reference sound is available, however, it is always possible to playback the recorded sound. Such a method, though simple in its principle, is widely adopted by digital sampling instruments or samplers. Samplers store a large quantity of examples of complete sounds, usually produced by other musical instruments. *Sampling synthesis*, also called *wavetable synthesis*, means recording, processing and playback of sounds. When we wish to synthesize a sound, we just need to directly play a sound from the stored repertoire. 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 (Fig. 6, left).

The most common modification is that of varying the sampling rate (speed) when reproducing the sound, which results in a pitch transposition: slowing down the sound, lowers the pitch. However, substantial pitch variations are generally not very satisfactory as a temporal waveform compression or expansion results in unnatural timbre modifications. It is thus necessary to allow only pitch variations to take place for the synthetic sound to be similar to the original one. On the other hand, what makes the method interesting the most is certainly the variety of sounds available (Borin et al., 1997).

From the implementation viewpoint, computational simplicity and limited amount of information to be stored are two contrasting needs for samplers. In fact, in order to reduce the data to be stored, it is possible to adopt "looping" techniques with almost any stationary portion of sounds. One method of improving the expressive possibilities of samplers is the interpolation between different sounds, often referring to "piano" and "forte" playing modes.

Nowadays, sound sampling is the most widely used technique for sonification in multimedia applications. A big advantage of sound sampling is that we can obtain

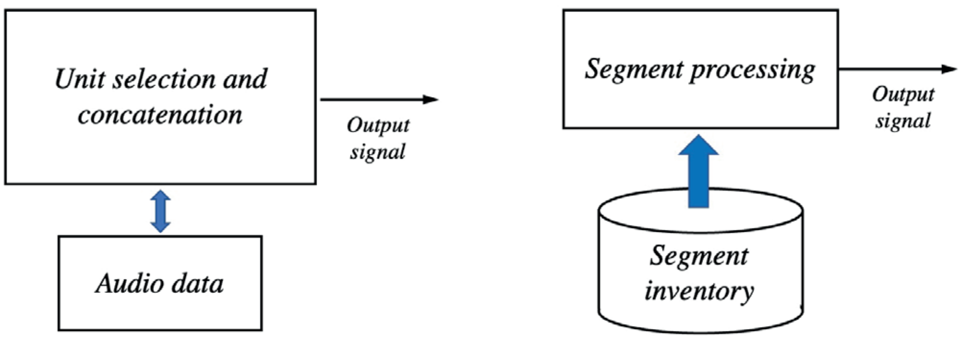

Figure 6. Synthesis from prerecord material or large sound database: sampling and concatenative synthesis (left), synthesis by time segment generation and processing as in classic electroacoustic techniques (right).

sounds which are practically indistinguishable from the real ones, with an almost negligible computational cost. Nevertheless, it is clear that in multimedia environments we have the same limitations found in musical instruments, with an additional unsatisfactory feature: the absolute separation between image and sound, due to the lack of a real sound model. It is interesting to propose an analogy between sound synthesis by sampling and static-image reproduction. Simple sound transformations, like amplitude scaling and frequency transposition, are analogous to size scaling of an image and to variations in color saturation. As a frequency transposition does not necessarily correspond to a different note played by the same instrument, in the same way a simple modification of color saturation does not necessarily correspond to a variation in scene illumination.

A major problem with sampling synthesizers, that strive to imitate existing instruments, is their lack of what we might call "prosodic rules" for musical phrasing. Individual notes may sound like realistic reproductions of traditional instrument tones, but when these tones are played in sequence, all of the note-to-note transitions – so important in instruments such as saxophones and the human voice – are missing.

Notice that sampled sounds can also be obtained synthetically or through the modification of other sounds, which is a way of widening the range of possibilities of application of samplers. From the composer's viewpoint, the use of samplers represents a practical approach to the so-called *musique concrète*.

Granular synthesis *Granular synthesis*, together with additive synthesis, shares the basic idea of building complex sounds from simpler ones. Additive synthesis, as we shall see, starts from the idea of dividing the sound in the frequency domain into a number of simpler elements (sinusoidal). Granular synthesis, instead, starts from the idea of dividing the sound in the time domain into a sequence, possibly with overlaps, of short acoustic elements called *grains*. Granular synthesis constructs complex and dynamic acoustic events starting from a large quantity of grains. The features of the grains and their temporal location determine the sound timbre.

In music, the use of granular synthesis techniques arises from the experiences of tape electronic music. In the early years of electronic music, the tools that composers had at disposal (e.g., fixed waveform oscillators and filters) did not allow for substantial variations of sound timbres. However, they were able to obtain dynamic sounds by cutting tapes, where real or synthetic sounds were recorded, into short sections and then putting them together again. The rapid alternation of acoustic elements provides a certain variety to the resulting sound. Granular synthesis offers unique opportunities to the composer and suggests new ways of organizing musical structure as clouds of evolving sound spectra (Cavaliere and Piccialli, 1997).

Sound granulation Two main approaches to granular synthesis can be identified: the former based on sampled sounds and the latter based on abstract synthesis. In the first case, a sound is divided in overlapping segments and windowed. Such a process is called *sound-granulation* and is quite similar to what happens in motion pictures, in which a fast sequence of static images produces a sensation of motion. By changing the order and speed of the windowed segments, however, a variety of sonic effects can be achieved. Grains can be extracted from different sounds to create cross-fading from one texture to another (Fig. 7). In any case, special attention should be paid to how to align time segments to avoid artifacts. To address this problem, several *Synchronous OverLap and Add* (SOLA) methods were developed.

Synthesis methods conceptually similar to granular techniques have received a new impulse due to the availability of ever larger databases of sounds. Various definitions are used in the literature, including *concatenative synthesis*, *audio mosaicing*, and *musaicing* (neologism from music and mosaicing). All works in this direction share the general idea that a target sound can be approximated by concatenating (linking together) sound segments taken from a pre-existing corpus of sounds. An appropriate segment description and selection strategy should be developed (Fig. 6, left).

Synthesis by time segment processing When longer sound files are used, we have *synthesis by time segment processing* (Fig. 6, right). We can include in the category of time domain methods also classic electroacoustic music techniques before the voltage-control era, where sounds were generated and recorded on a magnetic tape and then the tape was cut in pieces, which were edited, transposed, recombined and spliced to obtain an evolving and more complex sound. This approach was also used in computer music when the computing resources did not allow easy and fine mixing under the musician's control of the sound elements. The composer designs and synthesizes the various sound segments, which are stored in separate files. This constitutes the basic material, which is then placed over time.

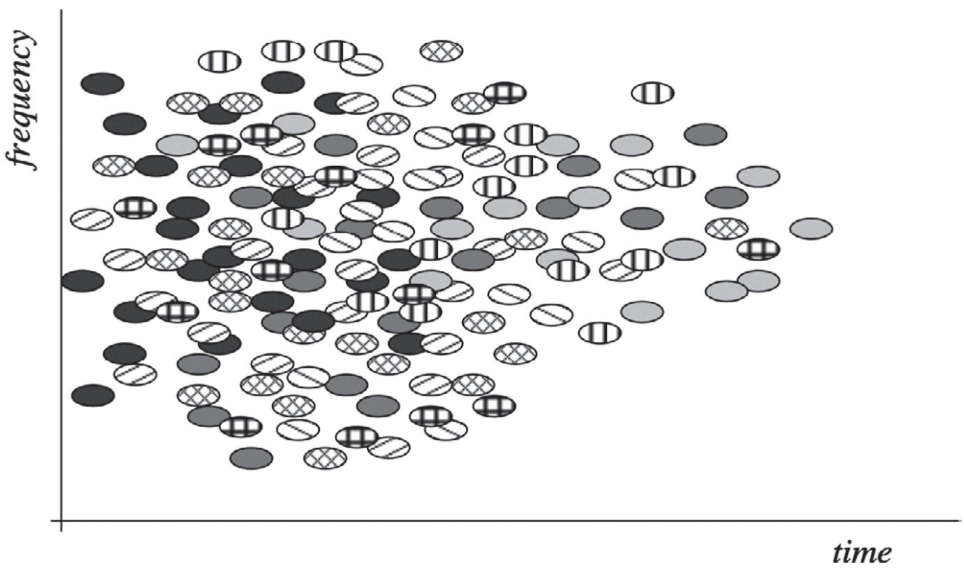

Figure 7. Representation of granular synthesis where grains derived from different sources are randomly mixed.

Micro-granular synthesis The second approach is based on *synthetic grains* typically consisting of a windowed sinusoidal which has the property of locating the energy both in frequency and time domain (Fig. 8). One possible analogy is with the mosaic technique, where the grains correspond to individual small monochromatic tiles and their juxtaposition produces a complex image. When the grains are scattered irregularly in the time-frequency plane, ”clouds” of microsounds, or sound textures, are obtained, that can simulate natural noisy sounds in which general statistical properties are more important than the exact sound evolution. Typical examples include the sound of numerous small objects (e.g., rice or sand) falling onto a resonating surface (e.g., a metal plate), or rain sounds composed by the accumulation of a large amount of water droplet micro-sounds, or even scratching/cracking sounds made by the accumulation of thousands of complex micro-sounds not necessarily deterministic. In general, we can expect these types of sounds to occur in the real world when they are the result of multiple realizations of the same event or the same phenomenon. In computer music, when the grains are irregularly distributed over time, this technique is also called *Asynchronous Granular Synthesis* (Roads, 1991).

Another peculiarity of granular synthesis is that it eases the design of sound events as parts of a larger temporal architecture. For composers, this means a unification of compositional metaphors on different scales and, as a consequence, the control over a time continuum ranging from the milliseconds to the tens of seconds. There are psychoacoustic effects that can be easily experimented by using this algorithm, for example crumbling effects and waveform fusions, which have the corresponding counterpart in the effects of separation and fusion of tones.

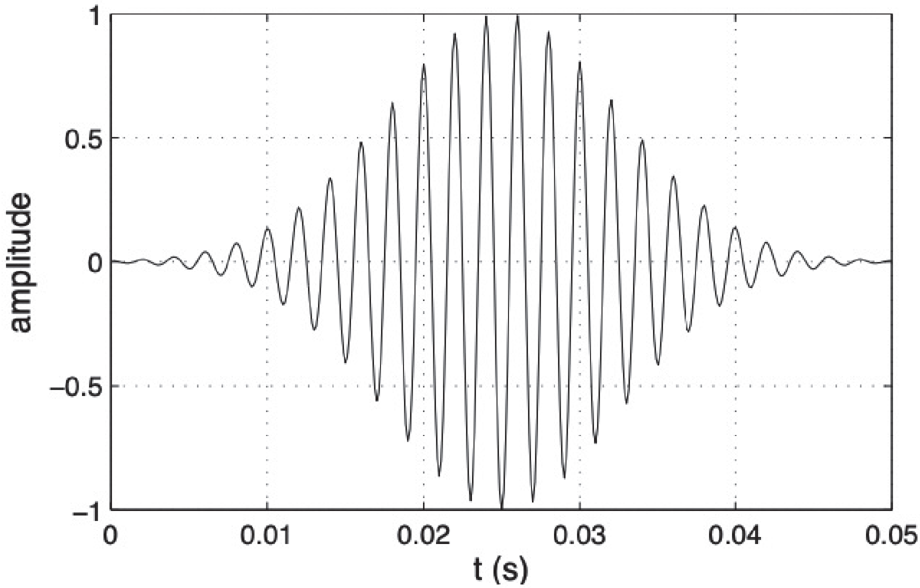

Figure 8. Example of a synthetic grain waveform, locating the energy both in frequency and time domain, for micro-granular synthesis.

Synchronized granular synthesis Other methods of synthesis, based on particular forms of elementary waveforms, have been proposed in computer music, especially as a way of realizing subtractive synthesis. The first realizations are *VOSIM* (Kaegi and Tempelaars, 1978) and *formant-wave function (French: forme d'onde formantique, FOF)* (Rodet, 1984) methods. These methods can also be considered as particular types of granular synthesis. In this case the temporal position of the grains is directly related to the pitch of the sound, and their waveform determines the spectral envelope. In computer music, when the grains are synchronized with the pitch period, it is called *Pitch Synchronous Granular Synthesis*, which is a way to implement in time domain the source-filter model excited by a pulse train (De Poli and Piccialli, 1991).

When the grains are aligned to a grid superimposed on the time-frequency plane, granular synthesis becomes the implementation of an inverse transform derived from time-frequency representations such as the *Short-Time Fourier Transform* (STFT) or *Wavelet Transform* (De Poli et al., 1991).

3.2 Frequency domain models

In this section the models which are best interpreted in the frequency domain are presented. In computer music often they are called *spectral models*.

3.2.1 Additive synthesis

In *additive synthesis*, complex sounds are produced by the superimposition of elementary sounds. In certain conditions, the constituent sounds fuse together and the result is perceived as a unique sound. This procedure is used in some traditional instruments, too. In an organ, the pipes generally produce relatively simple sounds; to obtain a richer spectrum in some registers, notes are created by using more pipes sounding at different pitches at the same time. The piano uses a different procedure. Many notes are obtained by the simultaneous percussion of two or three strings, each oscillating at a slightly different frequency. This improves the sound intensity and enriches it with beatings.

In order to choose the elementary sounds of additive synthesis, we first note that the Fourier analysis model enables us to analyze sounds in a way similar to the human ear and so to extract parameters that are perceptually significant. When we analyze a real, almost-periodic sound, we immediately notice that each partial amplitude is not proportionally constant, but that it varies in time according to different laws (Risset and Mathews, 1969). In the attack portion of a note, some partials, which in the steady state are negligible, are often significant. Frequency can also be variable over time (Fig. 9).

Any almost-periodic sound can be approximated as a sum of sinusoids. Each sinusoid's frequency is nearly multiple that of the fundamental, and each sinusoid evolves in time. For higher precision, the frequency of each component can be considered as slowly varying. Thus, additive synthesis consists of the addition of some sinusoidal

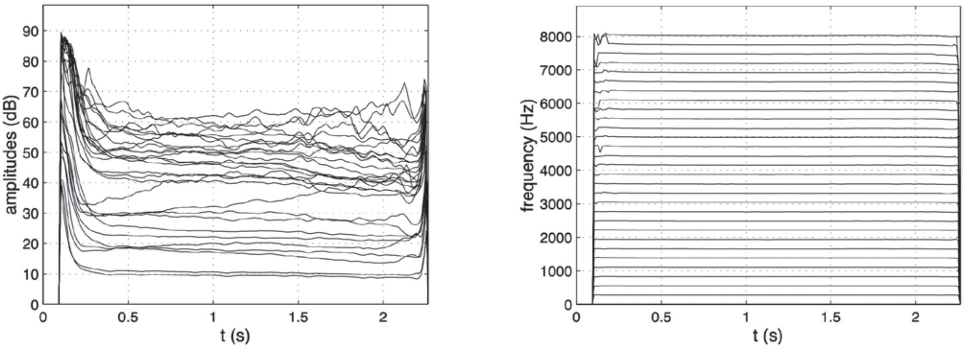

Figure 9. Fourier analysis of a saxophone tone: amplitude envelopes (left) and frequency envelopes (right) of the sinusoidal partials, as functions of time.

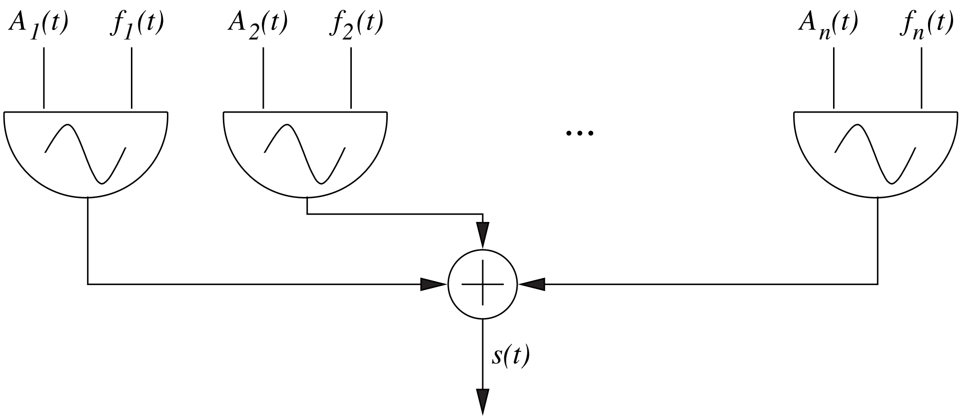

Figure 10. Additive synthesis: sum of sinusoidal oscillators with time-varying amplitudes and frequencies.

oscillators, whose amplitude and frequency are time varying (Fig. 10). The additive-synthesis technique also provides good reproduction of nonperiodic sounds, presenting in the spectrum the energy concentrated in some spectral lines. For example, Risset (1969) imitated a bell sound by summing sinusoidal components of harmonically unrelated frequencies, some of which were beating. In Risset's example, the exponential decaying amplitude envelope was longer for the lower partials. Additive synthesis provides great generality. But a problem arises because of the large amount of data to be specified for each note. Two control functions for each component have to be specified, and normally they are different for each sound, depending on its duration, intensity, and frequency. The possibility of data reduction has been investigated (Risset and Mathews, 1969). The method works best when used for harmonic or near harmonic sounds where little noise is present.

Additive synthesis is most practically used either in synthesis based on analysis (analysis/synthesis), often transforming the extracted parameters, or when a sound

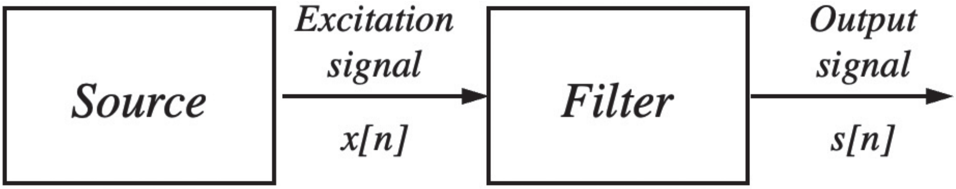

Figure 11. Source filter model.

of a precise and well-determined characteristic is required, as in psychoacoustic experiments. In any case, in order to familiarize musicians with sound characteristics and frequency representations, the technique is also useful from a pedagogical point of view. This method, developed for simulating natural sounds, has become the "metaphorical" foundation of a compositional methodology based on the expansion of the time scale and the reinterpretation of the spectrum in harmonic structures.

3.2.2 Source-filter models

Some sound signals can be effectively modeled through a feed-forward *source-filter* structure, in which the source is in general a spectrally rich excitation signal, and the filter is a linear system that acts as a resonator and shapes the spectrum of the excitation. A typical example is voice, where the periodic pulses or random fluctuations produced by the vocal folds are filtered by the vocal tract, that shapes the spectral envelope. The vowel quality and the voice color greatly depend on the resonance regions of the filter, called formants. The source-filter model exhibits a feed-forward structure of several blocks some of which generate signals (or acquire them as an input) and some transform such signals. If the transformation is linear, it is best interpreted in the frequency domain as a filter (Fig. 11).

In general, the division between the generator and the transformation gives rise to the possibility of controlling separately both the source and filter characteristics. There is, therefore, a greater flexibility of control and better interpretation of the parameters, as well as greater fusion in the class of sounds that can be obtained.

Subtractive synthesis In computer music, source-filter models are traditionally grouped under the label *subtractive synthesis*. Most analog voltage-controlled synthesizers in the 1960's and 1970's made use of subtractive synthesis techniques in which audio filters were applied to spectrally rich waveforms. It consists of filtering a spectrally rich signal source, which is what happens when the vibrations of violin strings are transmitted to the resonant body. The resulting spectrum is given by the product of the input signal spectrum multiplied by frequency response of the filter, so that some frequencies will be attenuated (subtracted) while others will be enhanced. According to the frequency response of the filter the general trend of the spectrum can be varied or, for example, a small portion of the spectrum of the signal can be extracted. If the filter is static, the temporal features of the input signal are maintained. If, instead, the

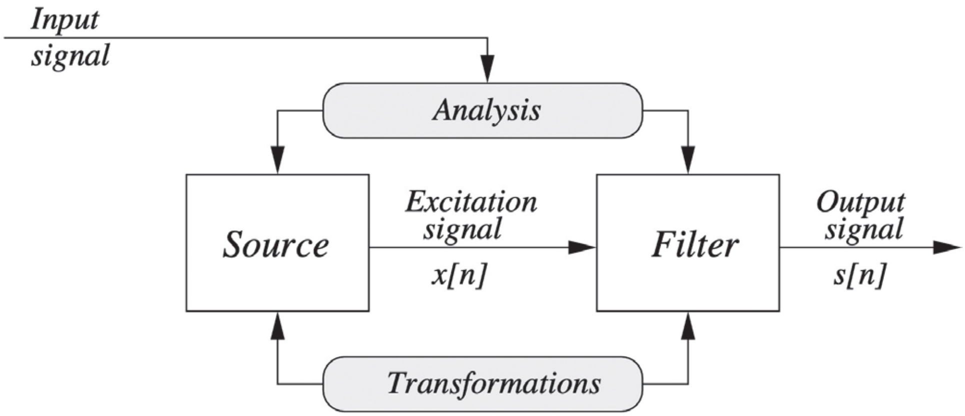

Figure 12. Source filter model in an analysis-synthesis framework: the analysis phase estimates the model parameters and the transformation step modify the parameters according to the musical desires.

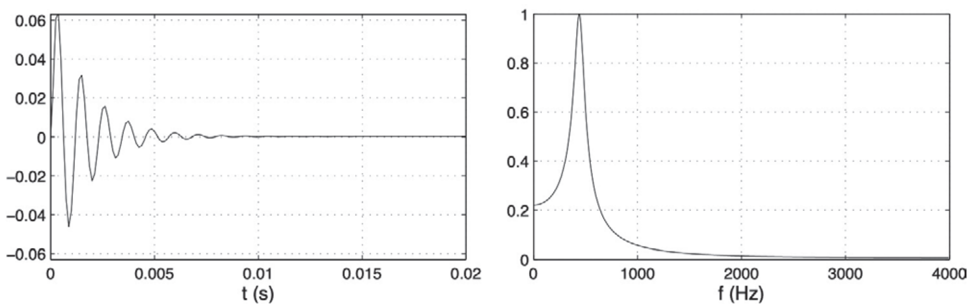

Figure 13. Example of a 2nd order resonator tuned on the center frequency *fc* = 440 Hz and with bandwidth *B* = 100 Hz: impulse response (left), magnitude response (right).

filter coefficients vary slowly over time, the characteristics of the sound will be a combination of those of the source and those of the filter, just as in singing.

This technique is most suitable for implementing slowly varying filters (acoustic response of a specific environment, spatialization) as well as filters that are subject to fast variations (muting effect, spoken or sung voice, sounds characterized by significant timbre dynamics). Notice that subtractive synthesis does not use specific assumptions on the periodicity of the source signal, therefore it can be successfully used for generating non-pitched sounds, such as percussions, in which case noise sources characterized by a continuous (non-discrete) spectrum are employed.

Source-filter models are often used in an analysis-synthesis framework, in which both the source signal and the filter parameters are estimated from a target sound signal, that can be subsequently resynthesized through the identified model. Moreover, transformations can be applied to the filter and/or the excitation before the reconstruction (see Fig. 12). One of the most common analysis techniques is Linear Prediction, that we will address in Sec. 3.2.2.

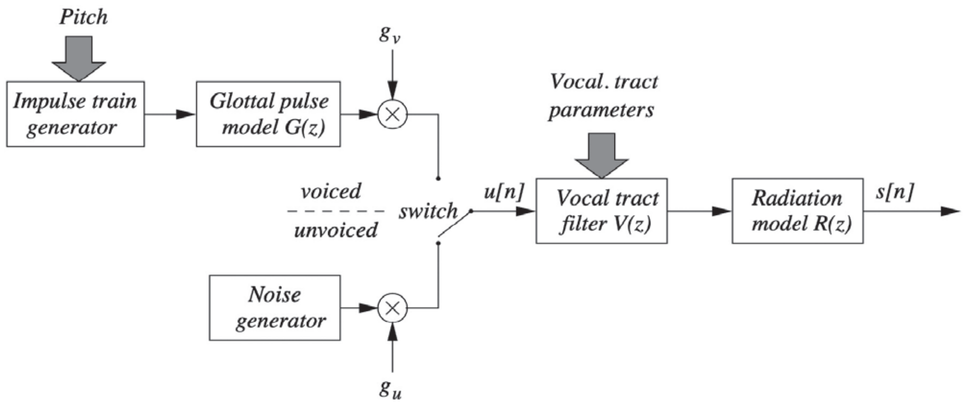

Figure 14. A general model for formant synthesis of speech.

Formant synthesis When the filter has prominent resonances, which enhance the spectrum in some frequency regions, we have *format synthesis*, which is very useful for sound characterized by such behavior as the vowel in voice synthesis. A class of filters that is widely used in subtractive synthesis schemes is that of *resonant filters:* the 2nd order IIR resonant filter is the simplest one, where center frequency and bandwidth can be controlled (Fig. 13). To separately control the frequency and bandwidth of each formant, a parallel structure (filterbank) of filters is advisable. When we want to control the overall trend of the spectrum, a single filter, as in LPC synthesis, is sufficient.

Formant synthesis of voice realizes a source-filter model in which a broadband source signal undergoes multiple filtering transformations that are associated to the action of different elements of the phonatory system. Depending on whether voiced or unvoiced speech (see above) has to be simulated, two different models are used. If the speech segment is a voiced signal, the input is a periodic pulse train whose period coincides with the pitch of the speech. If the speech segment is an unvoiced signal, vocal folds do not vibrate and turbulences are produced by the passage of air through a narrow constriction (such as the teeth). The turbulence can be modeled as white noise. The complete transfer function may or may not include vocal fold response depending on whether the sound is voiced or unvoiced. The block structure of the resulting model is shown in Fig. 14.

Linear Predictive Coding In linear source-filter models, if we can make simplified hypothesis about input, it is possible to estimate both the parameters of the source and the filter of a given sound. The most common procedure is *Linear Predictive Coding* (LPC) which assumes an impulse source or noise and a recursive filter (Fig. 15). By analyzing brief sequential segments of the sound, time-varying parameters can be extracted that can be used in resynthesis. Note that LPC-derived methods are in common use for voice transmission in cell phones.

Since a parametric model was used, the data obtained by the analysis have an exact interpretation in terms of the model. This fact supplies reference criteria for

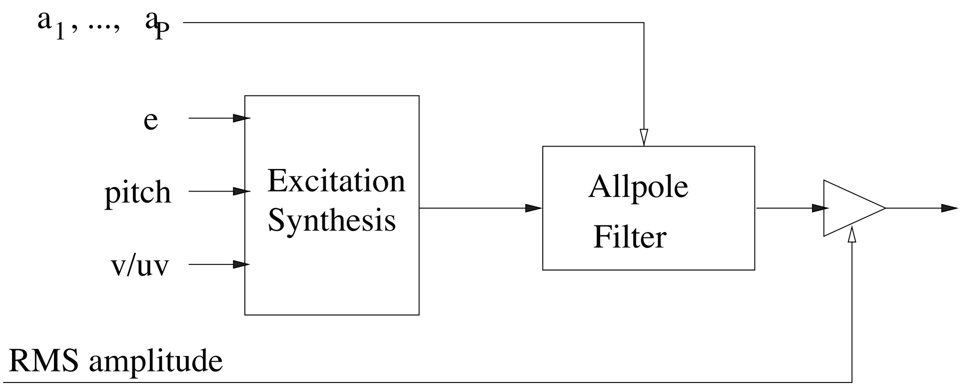

Figure 15. LPC synthesis.

their modification. Therefore, it is possible to control and modify separately different features of the signal: as an example, the pitch of a speech sound depends on the excitation and can be controlled separately from the formant structure, which instead depends on the filter. For example, when the excitation frequency is increased for the voice, the sound pitch is raised, without altering the position of the formants, and thus not affecting the voice quality and message. Or using the estimated filter parameters of a voice, that are perceptually robust, for filtering another sound, e.g. an orchestra, that has a dense time structure. As already mentioned, the time-frequency features of the two sounds combine, resulting in an orchestra that sings. In computer music, this application is called *cross-synthesis*.

3.3 Hybrid models or time-frequency models

Time domain and frequency domain models can be combined to have a more flexible and effective sound generation. Additive and subtractive syntheses are somewhat complementary, in the sense that the first one naturally reproduces sounds having a dominant periodic or quasi-periodic content, while the second one is better suited to reproduction of sounds having a dominant random content. In the second category we might put percussive, transient, or noisy sounds, like the consonants in speech. In sounds as they are found in nature, the two components, periodic and random, are almost always simultaneously present. As an example, consider the sound of a wind instrument: the deterministic signal results from self-sustained oscillations inside the bore, while the residual noisy signal is generated by the turbulent flow components due to air passing through narrow apertures inside the instrument. Similar considerations apply to other classes of instruments, as well as to voice sounds, and even to non-musical sounds. Therefore, it makes sense to try to separate the two components from the analyzed sounds, and then reproduce each component with the most appropriate algorithm. *Hybrid models* (i.e. time domain and frequency domain models) are used to this purpose.

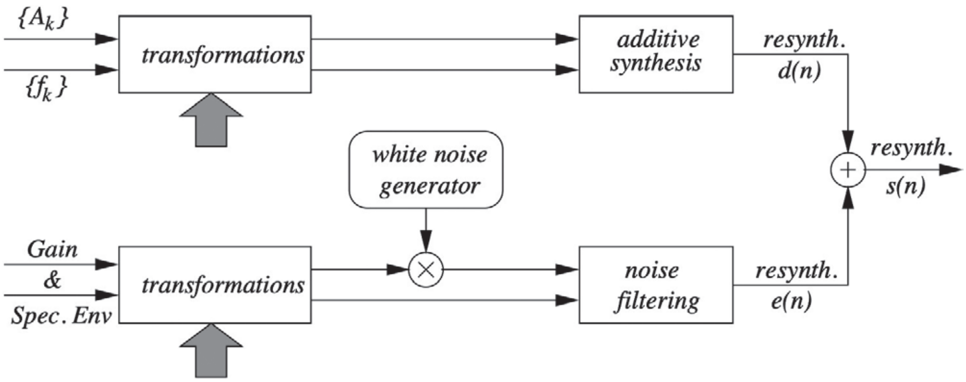

Figure 16. Spectral modeling synthesis (SMS) using additive and subtractive synthesis, with possible transformations of the analysis parameters.

3.3.1 Spectra Modeling Synthesis

Different models can be combined in order to have a more flexible and effective sound generation. One approach is *Spectral Modeling Synthesis* (SMS) that considers sounds as composed by a sinusoidal part and a residual noise part (*Sinusoidal* + *Noise* model). The fundamental assumption behind the sinusoidal + noise model is that sound signals are composed of slowly-varying sinusoids and quasi-stationary broad-band noises. A clever analysis procedure was proposed, to separate the sinusoidal from the noisy part of the signal and estimate their parameters from a given sound (Serra and Smith, 1990). During the synthesis the sines are generated by additive synthesis and the noise by subtractive synthesis. A more sophisticated approach considers a third impulsive component for representing transients, giving rise to the *Sinusoidal* + *Noise* + *Transient* model (Verma and Meng, 2000). Notice that the “sinusoidal + noise” additive model sounds good except for attacks, the “sinusoidal + noise + transients” additive model preserves attacks, but not the spatial image of multi-channel sounds.

Both additive and subtractive syntheses can be referred to as spectral models. Spectral models can be implemented in time-domain as additive synthesis for the sinusoidal component and subtractive synthesis for the noise component (Fig. 16), but also both can be jointly computed in frequency domain by using the inverse Short-Time Fourier Transform, which in computer music is usually called *Phase vocoder*, and can be efficiently computed through block processing.

3.4 Nonlinear models or abstract models

The transformations seen above, since they are linear, cannot change the frequency of the components that are present. Instead, when nonlinear transformations are used, frequencies can be even drastically changed. Thus, it is possible to vary substantially the nature of sounds in input.

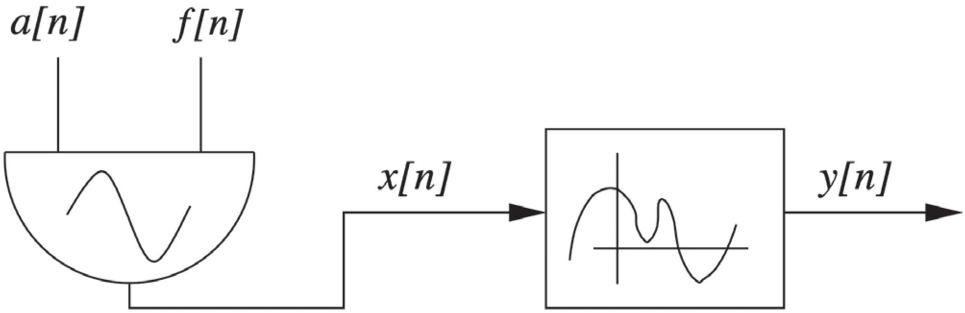

Figure 17. Sound synthesis by non-linear distortion (or waveshaping) of a sine wave.

Methods in this class are sometimes called *abstract algorithms*, since from mathematical equations generate synthesis sounds far from ”natural” sounds, but by manipulating these equations, we try to obtain sounds which allow a specific musical identity. In fact, the interpretation of nonlinear synthesis is not based on physical acoustics, but it comes from modulation theory applied to musical signals. Therefore, it inherits, in part, the analog interpretation as used in electronic music and is a new metaphor for computer musicians.

There are two main effects related to nonlinear transformations: spectrum enrichment and spectrum shift. The first effect is due to *non-linear distortion* of the signal and allows for controlling the brightness of a sound, while the second is due to its multiplication by a sinusoid (*ring modulation)* and moves the spectrum to the vicinity of the carrier signal, altering the harmonic relationship between the partials of modulating signal. The possibility of shifting the spectrum is very intriguing when applied to music. From simple components, harmonic and inharmonic sounds can be created, and various harmonic relations among the partials can be established. Often, the input amplitudes are varied by multiplying them by a constant or time-dependent parameter, called the modulation index. Thus, acting only on one parameter, the sound characteristics are substantially varied. Dynamic and variable spectra are easily obtainable. In additive synthesis, similar variations require a much larger amount of data, even when singles sinusoids are used.

The two classic methods for spectrum enrichment and spectrum shift, respectively *non-linear distortion* and *ring modulation* have, progressively, become less interesting, giving way to frequency modulation methods which combines both effects.

3.4.1 Non-linear distortion or waveshaping

Non-linear distortion (NLD) is an algorithm which had far less fortune than frequency modulation, even though it shares with it most of the advantages and drawbacks. In this method the idea is to feed a signal (typically a sine wave) into a function that maps the amplitude values in a non-linear manner, thus producing harmonically rich sounds (Fig. 17). Even for NLD it is possible to control the amplitude of partials when we are using special distorting functions and acting on the input sine amplitude (Arfib, 1979;

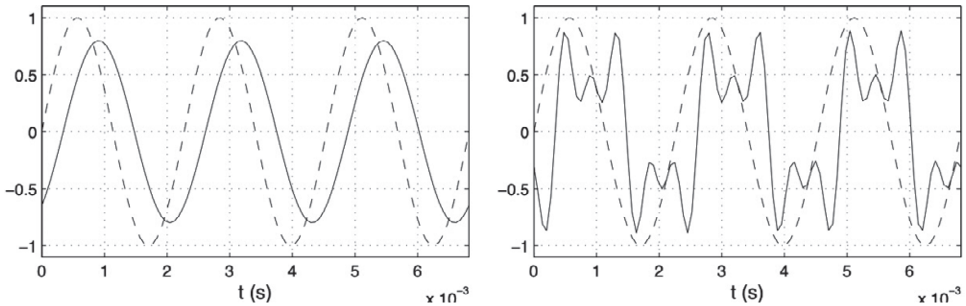

Figure 18. Example of output signals (solid lien) from a linear and from a non-linear memoryless system, in response to a sinusoidal input (dashed line): in a linear system the input and output differ in amplitude and phase only (left); in a non-linear system they have different spectra (right).

Brun, 1979). Besides being a synthesis algorithm by itself, NLD finds applications within other synthesis models, to obtain particular effects. In many real objects, non-linear saturation phenomena are found. They typically occur when the amplitude of vibrations is large enough. The acoustical effect is a spectral enrichment by addition of new spectral components (brightness), and it can be achieved by NLD (Fig. 18).

3.4.2 Ring modulation

Ring modulation (RM), sometimes called *multiplicative synthesis*, consists in the multiplication of the input (modulating) signal by a sinusoidal carrier signal (Fig. 19), which moves the spectrum to the vicinity of the carrier frequency. It derives from abstract mathematical properties of trigonometric functions as used in modulation theory applied to music signal. Therefore, it partially inherits and simulates digitally the processing blocks used in analog electronic music. Transformations that produce spectral shifts can produce very intriguing musical effects: complex harmonic and inharmonic spectra can be created starting from simple (sinusoidal) input sounds, and various harmonic relations among the partials can be established (Fig. 20). *Amplitude modulation* (AM) is a variant with similar characteristics.

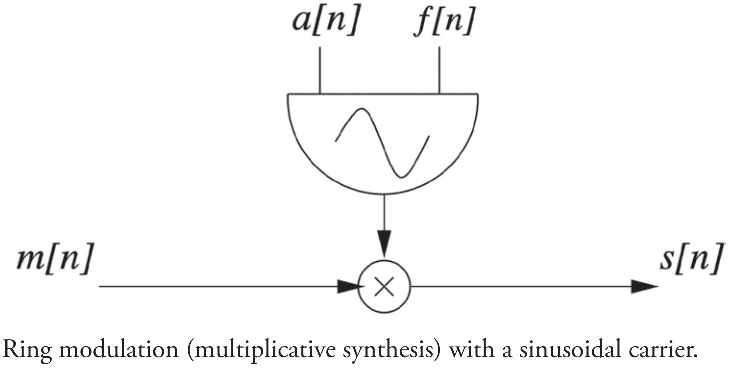

Figure 19. Ring modulation (multiplicative synthesis) with a sinusoidal carrier.

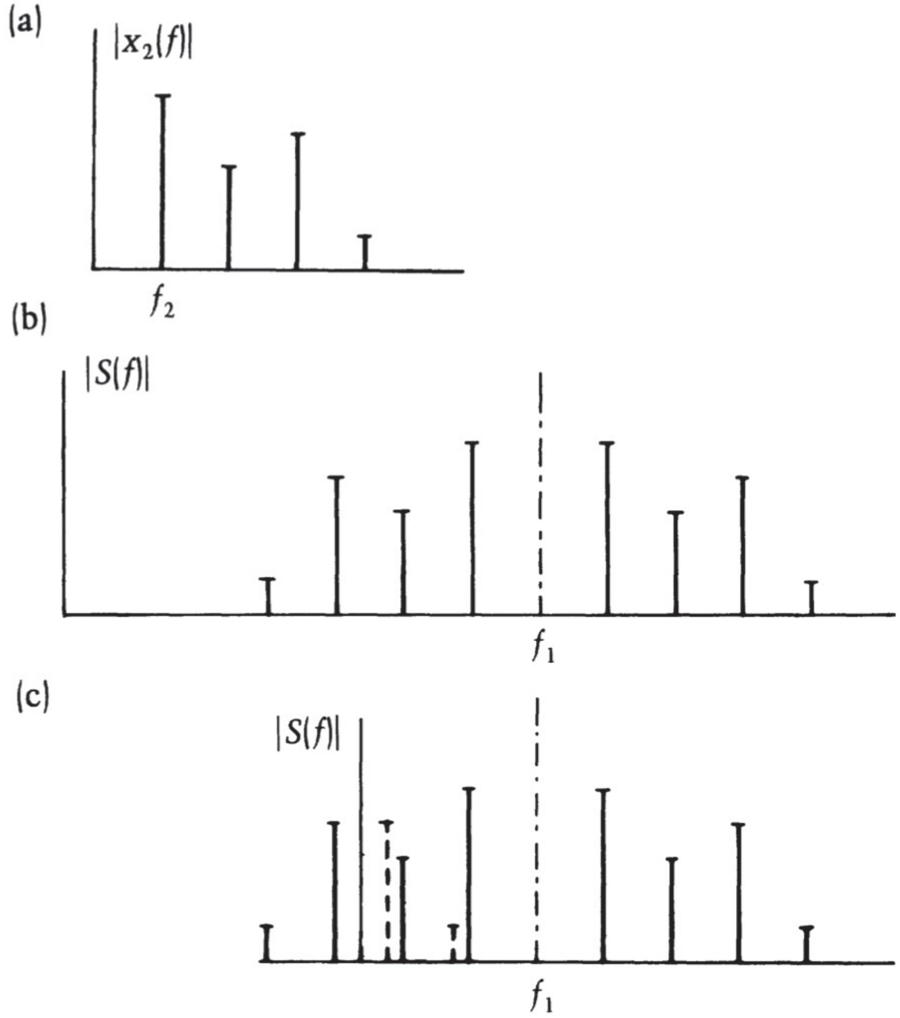

Figure 20. Multiplicative synthesis. Spectrum of a periodic signal x_2 with four harmonics (a). Resulting spectrum when the signal is multiplied by a sinusoid of frequency f_1, greater than its bandwidth ($f_1 = 7f_2$) (b). Resulting spectrum when x_2 is multiplied by a sine of frequency lower than its bandwidth ($f_1 = 2.6f_2$) (c). The components deriving from the folding of negative frequencies are shown as dashed lines.

3.4.3 Frequency modulation

Another very successful sound synthesis technique of the eighties had been the synthesis by *Frequency Modulation* (FM). John Chowning invented FM while experimenting on vibrato effects on digital oscillators. Actually, FM in its simplest formulation is nothing more than an audio-frequency vibrato effect (Fig. 21). If the frequency of a sinusoidal oscillator (carrier) is driven by another oscillator (modulator), new spectral components appear in the sidebands of the carrier frequency, spaced by in-

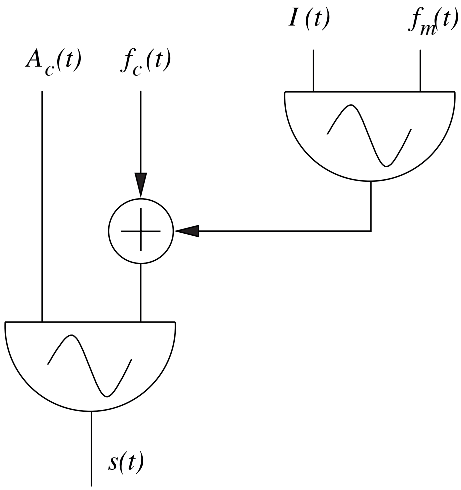

Figure 21. Frequency modulation synthesis.

tervals equal to the modulator frequency. When FM sound synthesis was first introduced, a complete mathematical theory was already available from the field of electrical communications. The real discovery was the possibility of generating complex sounds by means of very simple devices, like digital oscillators (Chowning, 1973).

The carrier/modulator frequency ratio determines the spectral content of sounds, and is directly linked to some important features, like the absence of even components, or the inharmonicity. The modulation index determines the bandwidth and is usually associated with a time curve (the so called envelope), in such a way that time evolution of the spectrum is similar to that of traditional instruments. For instance, a high value of the modulation index determines a wide frequency bandwidth, as it is during the attack of typical instrumental sounds. On the other hand, the gradual

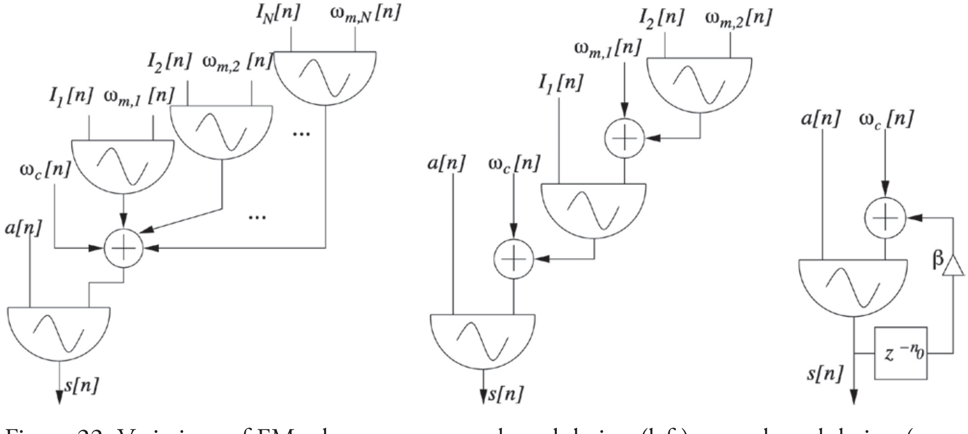

Figure 22. Variations of FM scheme: compound modulation (left), nested modulation (center), and feedback modulation (right).

decrease of the modulation index determines a natural diminution of the frequency bandwidth during the decay phase. Controlling the bandwidth of the produced signal gives the same effect as a dynamic filter, similar to the subtractive synthesis.

By combining various FM modules, richer spectra and a wider range of possibility for variation can be obtained. For example, when several carriers, or a complex periodic carrier, are used and modulated by the same modulator, side bands around each sinusoidal component of the carrier are obtained. The corresponding FM scheme is termed *compound modulation* (Fig. 22, left). This effect can be used to separately control different spectral areas of a periodic sound. It is also possible to use complex modulators. A similar effect is obtained when a sinusoidal modulator is itself modulated by a second one (*nested modulation* shown in Fig. 22, center). In this case, in fact, the carrier is modulated by an FM signal, thus rich in components. The resulting signal still maintains its frequency, as seen above in the case of parallel modulators, but with more energy in most of the lateral components. The last FM scheme that we examine is *feedback modulation* (Fig. 22, right), in which past values of the output signal are used as a modulating signal. Moreover, one may vary the delay length in the feedback, and observe emergence of chaotic behaviors for suitable combinations of the parameters (De Poli, 1983).

A point of strength of FM is its simplicity and efficiency, which allowed an immediate integration in low-cost chips. Moreover, the FM model offers a great freedom of action, since it can simulate real sounds while being a model open to the user through a few parameters at symbolic level. The fact that this symbolic description is far from the real-world experience is one of the fundamental limits of the algorithm. The control of an FM generator is far from intuition and requires a big deal of experience. Another feature, which can be seen both as an advantage and a drawback, is the fact that sophisticated users are able to recognize FM synthesis from the sound results. As a consequence, FM music instruments acquired their own identity of instruments "tout court", to the detriment of a more general usage.

Its main qualities, i.e. great timbre dynamics with just a few parameters to control and to low computational costs, are progressively losing popularity when compared with other synthesis techniques which, though more expensive, can be controlled in a more natural and intuitive fashion. The FM synthesis, however, still preserves the attractiveness of its own peculiar timbre space and, though it is not particularly suitable for the simulation of natural sounds, it offers a wide range of original synthetic sounds that are of considerable interest for computer musicians.

3.5 Models for sound and space

In visualization or sonification there is an essential process, which appears basically unvaried in the two domains of image and sound. It is the process of passing from the space of objects, thought of as entities provided by the model, to the space of images or sounds. In computer graphics, this passage is governed by the laws of perspective, illumination and visibility. In sound computing these concepts are respectively replaced by sound localization, radiation/diffusion and masking. Most of the aforementioned models consider sounds as monodimensional functions of time picked up at some point close to the source. While loudspeakers and voice can be considered as point-like sources, in many musical instruments sound is emitted from many points or from a radiating surface. Then, sound waves are diffused in the environment and they reach the ears of the listener, that we can consider as point-like pickups.

Several models have been proposed that take into account the effects introduced on pressure waves by propagation in air, interaction with surrounding objects, and enclosing surfaces. They are implemented as recursive filters featuring long delay-lines. These models are based on perceptual descriptions of acoustic scenes (e.g., reverberation, spaciousness), or on physical descriptions of the environment (e.g., geometry of the room, position of the sound source). The models can be implemented both in time or frequency domain. For an extensive treatment of the topic the reader might look at (Pulkki et al., 2011).

4. Sound modeling: physics-based models

The models seen so far have attempted to represent a sound as it reaches our ears. An alternative approach is that of modeling the source of a sound and obtaining its synthesis through a simulation of the physical phenomena that produce sound.

In the family of *physics-based models* we put all the algorithms generating sounds as a side effect of a more general process of simulation of a physical phenomenon. Physics-based models can be classified according to the way of representing, simulating and discretizing the physical reality. Hence, we can talk about cellular, finite-difference, and waveguide models, thus intending that these categories are not disjoint but, in some cases, they represent different viewpoints on the same computational mechanism. Moreover, physics-based models have not necessarily to be based on the physics of the real world, but they can, more generally, gain inspiration from it; in this case we will talk about pseudo-physical models. In this section, the approach to physics-based synthesis is carried on with particular reference to real-time applications, therefore the time complexity of algorithms plays a key role.

The aim is that of building physical models that can be used to produce sound and can be effortlessly employed by musicians (composers and performers). At the basis of the musical interest in these two models, there are two fundamental hypotheses:

- The complexity of sound is given by the complexity of the structure of the model and therefore by the generation algorithm;
- There exists a relationship between the effects of the actions on the source (that is to be simulated) and its model. In this way, the parametric control of the algorithm is simplified and has an inherent semantic interpretation.

As regards control parameters, we have parameters of the model, that depends on the physical properties of the source to be modeled, and parameters that affect

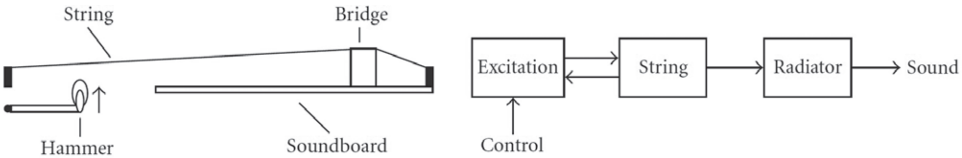

Figure 23. Physical model (left) and computational model (right) of a piano string.

the quality of the single sound. For example, in the first case, when modeling a piano string (Fig 23), tension, density, diameter, stiffness, etc. should be specified. The parameters that individuate a specific sound instance from the class of the sounds generable by a piano model are which key and key and key velocity. The sound of a piano string model depends on impact velocity and string state: it reacts realistically to pianist's touch and changes every time. In the second case the sound of a piano string model depends on impact velocity and string state: the physical model reacts realistically to pianist's touch and the sound changes every time. By comparison, in the case of additive synthesis the amplitude and frequency envelopes for all partials must be specified each time.

4.1 Functional blocks

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

4.1.1 Block decomposition

The first level of conceptual decomposition that we can devise for musical instruments consists in two functional blocks: an exciter and a resonator. The *exciter* is the place where energy is injected into the instrument, and it strongly affects the attack transient of sound, which is fundamental for timbre identification. The *resonator* sustains and controls the oscillation and is related with sound attributes like pitch and spectral envelope. We can argue that the exciter tends to define the timbre identity,

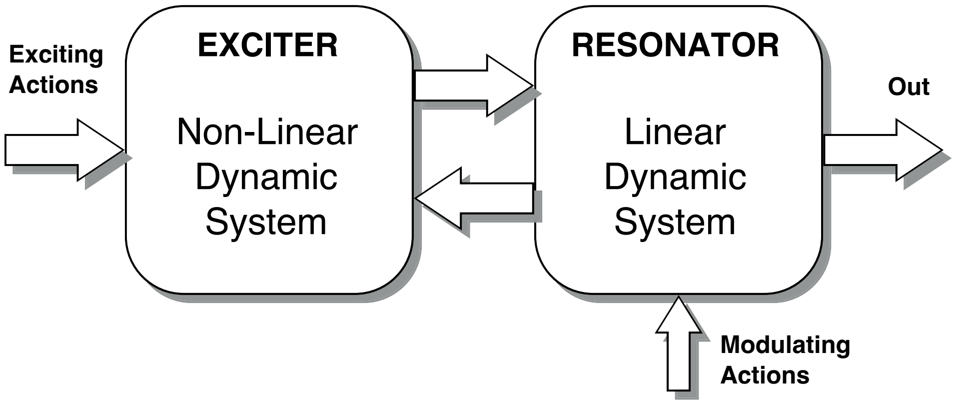

Figure 24. Exciter and resonator.

while the resonator tends to define the timbre quality. For instance, in a violin, one could separate the description of the bow (exciter) from the rest of the instrument (resonator). These two parts have corresponding mathematical descriptions typically consisting in a strongly non-linear system for the exciter and a (to a great extent) linear system for the resonator. In this conceptual scheme, the radiating element (bell, resonating body, etc.) is implicitly enclosed within the resonator.

The player controls the performance by means of inputs to the two blocks (Fig. 24). In a clarinet, for instance, we have a feedback structure where the reed is the exciter and the bore with its bell acts as a resonator. The player exerts *exciting actions* such as controlling the mouth pressure and the embouchure, as well as *modulating actions* such as changing the bore effective length by opening and closing the holes. In a plucked string instrument, such as a guitar, the excitation is provided by plucking the string, the resonator is given by the strings and the body, and modulating actions take the form of fingering. The interaction is only weakly feedback, so that a feed-forward scheme can be adopted as a good approximation: the excitation imposes the initial conditions and the resonator is then left free to vibrate.

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

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 modeling 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.

4.1.2 Model Structure

The interaction of exciter and resonator is the main source of richness and variety of nuances that can be obtained from a musical instrument. The interaction can be ”feedforward”, when the exciter doesn’t receive any information from the resonator, or ”feedback”, when the two blocks exert a mutual information exchange.

Feed-forward structure The simplest case is represented by the *feed-forward structure*. In this case, the exciter acts on the resonator without receiving any information from it: strictly speaking, therefore, one cannot talk of “interaction” between the two parts (Fig. 25). In many percussion instruments, for example, the excitation is a short impulse that is not affected by the feedback from the resonator. The feed-forward structure lends itself to a simple description of those instruments in which the excitatory mechanism imposes some initial conditions on the resonator, then letting it free to evolve over time. From the physical modeling point of view the source-filter model, we saw in sect. 3.2.2, is a sort of physics-based model with feed-forward structure. Alternatively, it can be seen as a signal generator. In this situation, the status of the exciter is controlled by the performer without any information from the resonator. For this reason, this model does not adequately simulate the transients.

In LPC synthesis, the LPC filter is not a perfect simulation of the vocal tract. The simple form of the filter is adopted because it is easy to solve equations to find its coefficients, given a frame of natural speech waveform. In estimating the LPC parameters, a residual signal can be obtained, which accounts for model imperfections and

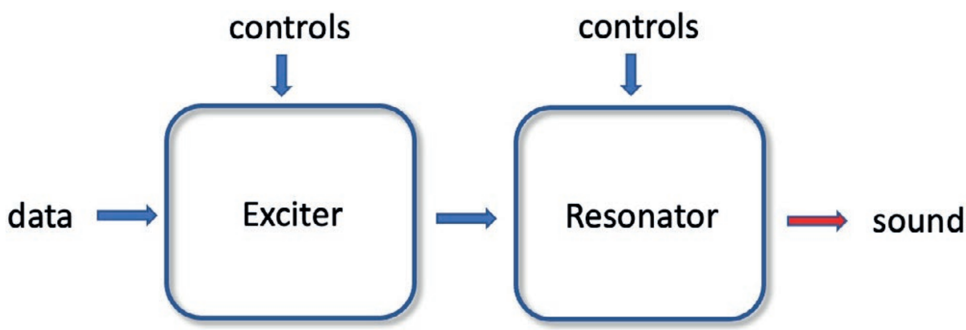

Figure 25. Feed-forward structure.

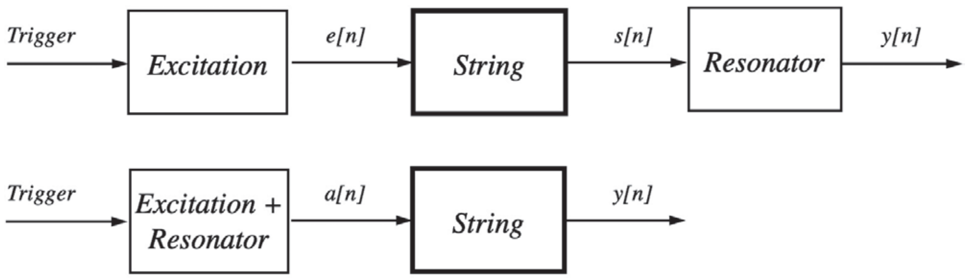

Figure 26. Commuted synthesis: Schematic diagram of a stringed musical instrument (above); the string input aggregates excitation and resonator impulse responses (below).

can be used as input to the filter during the synthesis to improve the sound realism. The limits of a detailed physical simulation are also found when we try to model the behavior of a complex linear vibrating structure, such as a soundboard; in such cases it can be useful to record its impulse response and include it in the excitation signal as it is provided to a feedforward interaction scheme. Such a method is called *commuted synthesis*, since it makes use of commutativity of linear, time-invariant blocks (Fig. 26). A valuable way of shortening the excitation table in commuted synthesis is to factor the body resonator into its most-damped and least-damped modes. The most-damped modes are then commuted and combined with the excitation in impulse-response form. The least-damped modes can be left in parametric form as recursive digital filter sections or can precomputed and stored in a wavetable. Typically, several excitation signals are used for one instrument. This method works very well for plucked or struck string instruments.

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

Feedback structure The structure that is most relevant to source models is the *feedback structure*, which also takes into account the action of the resonator on the exciter. This type of structure therefore requires a mutual exchange of information between the blocks making up the system (Fig. 27). Many musical instruments can be usefully described through a similar structure. An example that highlights the interdependence between the signals of exciter and resonator is that of a clarinet. In this instrument, vibratory phenomena take place in the bore. Within it, the perturbations are caused by variations of the incoming flow (action). On the other hand, the latter depends on the status of the opening of the reed which, in turn, is a function of the difference between the pressure existing in the mouth of the performer and that in the initial part of the bore (reaction).

This structure lends itself well to the simulation of persistently excited instruments. It should also be noted that, even in those cases in which the free evolution seems

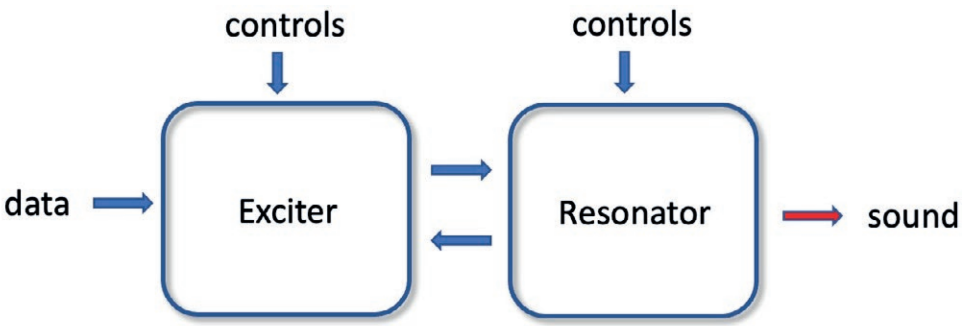

Figure 27. Feedback structure.

overwhelming, the interaction between exciter and resonator is however still clearly heard at the start of the sound, where it plays a fundamental role in giving the instrument its characteristic timbre.

While allowing a more accurate description of the instrument, the feedback structure also has some disadvantages too. In particular, because of its generality, it is not very easy to use in practice: in fact, it is not obvious how to specify the description of blocks and the way they interact. Another problem is linked to their non-computability. By using this structure, and therefore combining explicit equations, one can find oneself in the situation of having an implicit model equation that cannot be made explicit. More generally, one can observe that this situation arises, whenever both exit functions of the two blocks show an instantaneous dependence on the inputs: appropriate methods therefore need to be applied in its implementation. In this type of structure, temporal sound variations are caused mainly by the interaction between the parts. It is therefore not required to impose continuous variations in the parameters during synthesis, as is often necessary in signal models.

One interesting characteristic of this structure is the presence of non-linearity in the feedback loop, which gives rise to a complex behavior. This type of non-linear structure is very difficult to control and, because of this, it has been taken into consideration for musical instruments only quite recently. On the other hand, these algorithms exhibit a dynamic timbric character as intrinsic property of the structure. The synthesis for physical models is therefore interesting, as it allows an interpretation of this type of structure. The interpretation may give useful information for the control of model parameters and for the evaluation of their properties (stability, convergence and so on).

In digital implementations it is convenient to use the three block *EIR structure* (Fig. 28), where in between the two blocks exciter and resonator, a third block is found. This is an *interaction block* and it can convert the variables used in the exciter to the variables used in the resonator, or avoid possible anomalies introduced by the discretization process. The idea is to have a sort of adaptor for connecting different blocks in a modular way. This adaptor might also serve to compensate the simplifications introduced by the modeling process (Borin et al., 1992a).

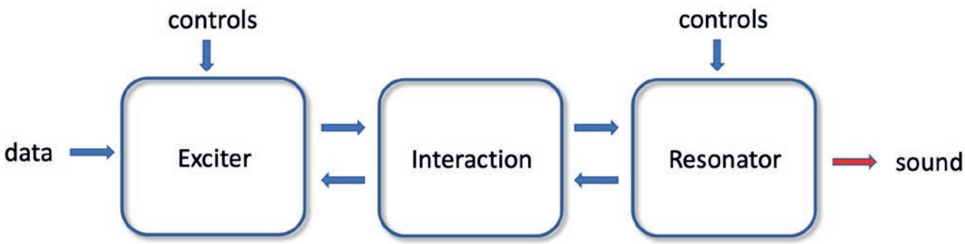

Figure 28. The three block EIR structure: the interaction block separates the exciter from the resonator, so that they can be designed independently.

4.1.3 Lumped vs. distributed models

When we consider the properties of the physical elements, physics-based sound modeling paradigms are often grouped into two broad categories, namely lumped and distributed models. *Lumped models* are used when a physical system can be conveniently described without explicitly considering its extension in space. The variables are function of time alone. As an example, a mechanical resonating body may be described in terms of ideal masses or rigid elements, connected to each other with spring and dampers, and possibly non-linear elements. The resulting models are naturally described in the time domain, in terms of Ordinary Differential Equations (ODE).

Distributed models are more often used for describing vibrating bodies or air volumes where forces and matter depend on (and propagate along) both time and space. All the variables are functions of time and one or more spatial variables. Strings, bars, acoustical bores, membranes, plates, rooms, can be treated as continuous distributed systems, and mathematically described by means of Partial Differential Equations (PDE).

In both cases the physical behavior can be represented by ordinary or partial differential equations, whose form can be learned from physics textbooks and whose coefficient values can be obtained from physicists' investigations or from direct measurements. These differential equations often give only a crude approximation of reality, as the objects being modeled are just too complicated.

4.2 Physical modeling techniques

The techniques used for physical modeling can be classified on the basis of the way they represent, simulate and discretize physical reality. The methods can be divided into five categories: (i) source–filter modeling, (ii) numerical solution of partial differential equations, (iii) vibrating mass–spring networks, (iv) waveguide synthesis, and (v) modal synthesis.

These categories are not disjoint, but in some cases represent different perspectives of the same generation mechanism. More generally, one can observe that models need not necessarily represent existing physical objects but may instead simply draw inspi-

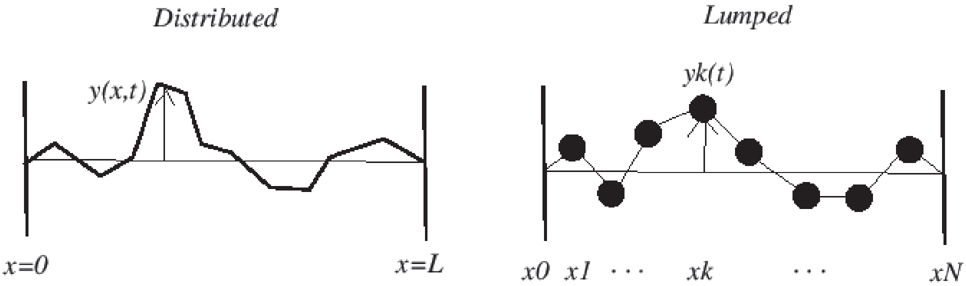

Figure 29. Distributed vs. lumped model of a string.

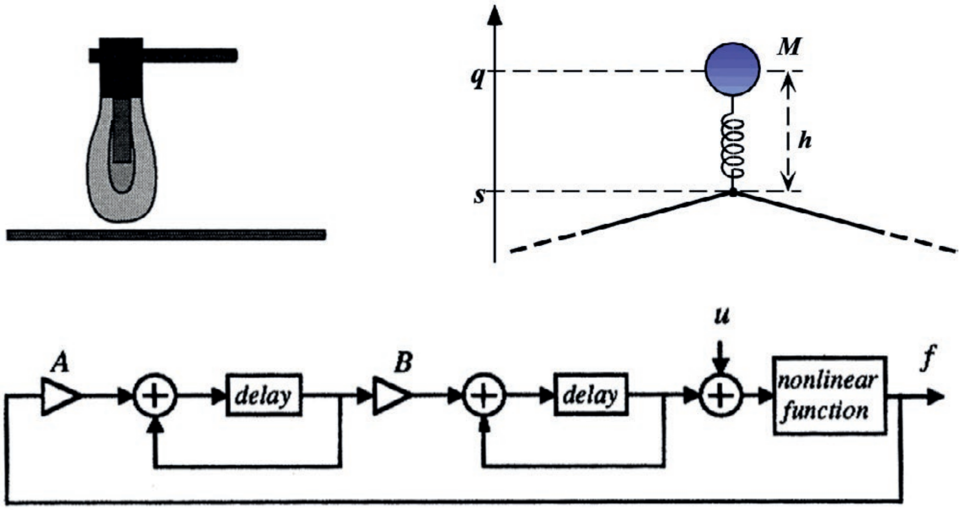

Figure 30. An example of a mechanical model of an exciter, the hammer of a piano: physical hammer (above, left), model (above, right), finite difference algorithm (below).

ration from them. In order to develop the model, one first needs to formulate specific hypotheses on the internal structure of the source. Then, from the knowledge of the physical laws that regulate its evolution in time, a system of equations is written out whose solution represents the signal of interest. There are several ways of arriving at these equations.

4.2.1 Digitizing differential equations: finite differences

A first class of models is represented by the equations of the dynamics of vibrating objects: there will be ordinary differential equations for the rigid elements and partial differential equations for the flexible and spatially distributed elements (Fig. 29). These equations regulate the variation of a physical quantity in space and time; in the simpler cases (absence of loss, perfect elasticity) it is possible to obtain analytical results, while in more complex cases, it is necessary to use the numerical solution. In this case, it is necessary to discretize the equation in time and space using numerical analysis methods, thus obtaining a system of *finite difference* (FD) equations. Methods in common use are finite difference approximation and bilinear transform. Fig. 30 shows the computable discretization scheme of the non-linear hammer, obtained from the straightforward finite difference approximation of the dynamics.

One of the most popular ways of solving partial differential equations is finite differencing, where a grid is constructed in the spatial and time variables, and derivatives are replaced by linear combinations of the values on this grid. Classical techniques from numerical analysis can be used to solve the equations, by numerical integration of difference equations. Moreover, several new techniques, suited for real-time synthesis, from signal processing are being developed to this purpose. This approach is closely related to the so-called Finite Element Method (FEM) or Boundary Element Method (BEM) that are used in mechanical engineering to simulate vibration of structures.

4.2.2 Cellular models

A second type of models is represented by *cellular models*. This method requires the decomposition of physical objects into simple mechanical elements (springs, masses, frictions) that are linked to one another. While suitable for the simulation of vibrating bodies such as plates, bars, ropes or non-uniform membranes, this method is hardly suitable in the simulation of acoustic pipes or wind instruments in general. This model has a very high computational cost. In fact, it takes into consideration the motion over time of all points in which the source has been decomposed into, while it is often enough to observe the motion of a few important point for the purposes of gaining musically interesting information.

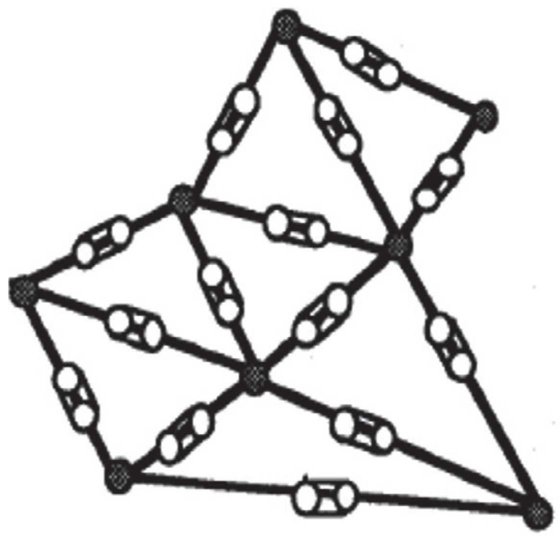

Figure 31. An example of a cellular model composed of an interconnected network of simple mechanical elements.

Several approaches to physics based sound modeling can be recast in terms of cellular automaton, the most notable being the CORDIS-ANIMA system introduced by Cadoz and his co-workers, who came up with cells as discrete-time models of small mass-spring-damper systems, with the possible introduction of nonlinearities (Fig. 31). The main goal of the CORDIS-ANIMA project was to achieve high degrees of modularity and parallelism, and to provide a unified formalism for rigid and flexible bodies. Moreover, in the case of the multiplicity of micro-objects, it has shown good effectiveness for joint production of audio and video simulations (Cadoz et al., 1984; Florens and Cadoz, 1991).

4.2.3 Wave and waveguide models

When discretizing physical systems, a key role is played by the efficiency and accuracy of the discretization technique. Namely, we would like to be able to simulate simple vibrating structures and exciters with no artifacts (e.g. aliasing, or non-computable dependencies) and with low computational complexity. Due to its good properties with respect to these two criteria, one of the most popular ways of approaching physical modeling of acoustic systems makes use of wave variables instead of absolute physical quantities.

A particularly efficient model for one dimensional resonator is the *digital waveguide* (DWG) model. It is based on the analytic solution of the equation that describes the propagation of perturbations in a medium and in the discretization of the solution. For instance, the propagation of the pressure variation in a cylindrical tube can be expressed as the sum of two waveforms propagating in opposite directions (Fig. 32). To model a discontinuity, we can insert a special junction that models the wave scattering. Low-pass and all-pass filters are used to simulate dissipative and dispersive effects in the medium.

We can thus model and implement such wave propagation by using simple delay lines. Therefore, by utilizing the *wave variables*, instead of the physical ones (such as velocity/force, flow/pressure), we obtain a model – made of delay lines, junction elements and filters – that strictly corresponds to our perception of physical reality. Moreover, the simulation algorithm is particularly simple and effective, allowing simulating many types of musical instruments in real time.

Let us consider deviations from ideal propagation due to losses and dispersion in the resonator. Usually, these linear effects are lumped and simulated with a few filters

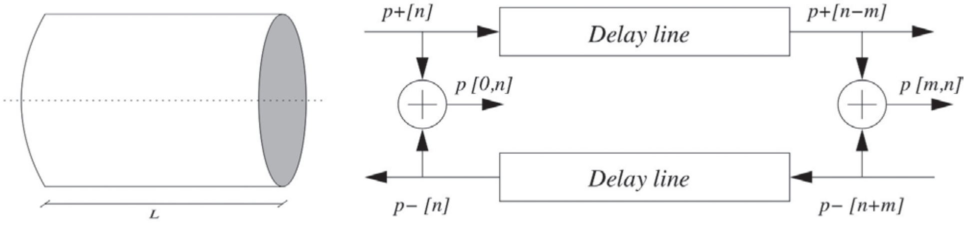

Figure 32. Wave propagation in a cylindrical tube (left) and the corresponding waveguide model (right).

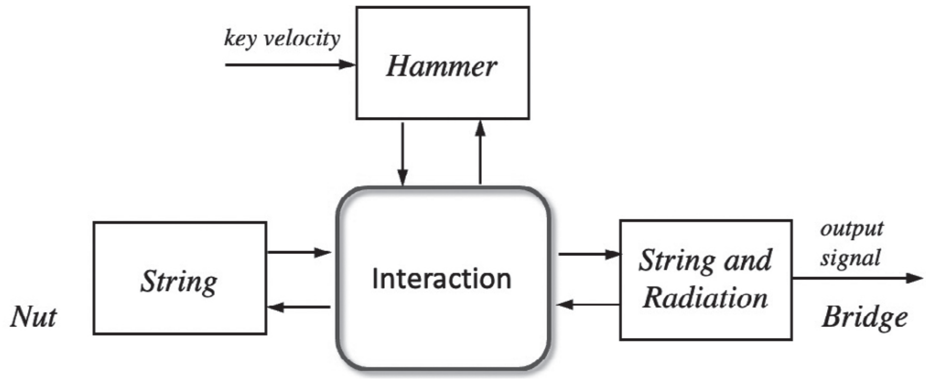

Figure 33. An example of a computational model of a piano implemented using the EIR (Exciter-Interaction-Resonator) three block structure.

which are cascaded with the delay lines. Losses due to terminations, internal frictions, etc., give rise to gentle low pass filters, whose parameters can be identified from measurements. Wave dispersion, which is often due to medium stiffness, is simulated by means of allpass filters whose effect is to produce a frequency-dependent propagation velocity. In order to increase the computational efficiency, delay lines and filters should be lumped into as few processing blocks as possible. When wave variables are adopted in the digital domain for representing lumped components this approach is called *Wave Digital Filter* (WDF). It is important to notice that Digital Waveguides and Wave Digital Filter are fully compatible with each other, which makes it very desirable to explore new physical models based on the wave representation of physical phenomena.

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 discretize it in time. Moreover, the solution can be computed only for the points of interest and not for each point of the structure, as when digitizing partial differential equations, thus allowing for great computational efficiency.

For example, figure 33 shows a computational model of a piano using the EIR structure of Fig. 28, where the hammer is modeled as in Fig. 30 and the strings are modeled using the waveguide approach as in Fig. 32.

Delay-line based oscillators As an example, let us consider a rough model of clarinet. The non-linear block representing the reed can simply be an instantaneous non-linear map. This use of instantaneous nonlinearities gives rise to a general scheme of nonlinear oscillator, depicted in Fig. 34, which is composed of an instantaneous map $y = G(x, u)$, possibly dependent on input parameters or signals u, a delay-line section, which determines the periodicity, and a linear reflection filter $R(z)$ which can be tuned to give the desired spectral dynamics.

In particular, if the reflection filter is reduced to a constant, and the input signal u is constant, the system evolution is described by an iterated map. This formula-

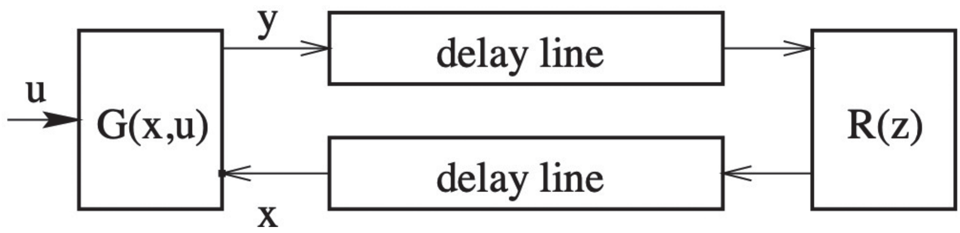

Figure 34. A computational model for non linear oscillators, useful for musical sounds.

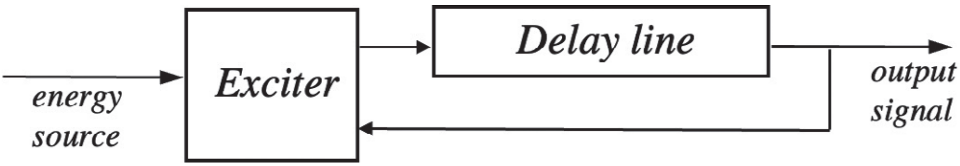

Figure 35. The simplified model for non linear oscillators as iterated maps, useful to explain nonlinear oscillations in musical instruments.

tion permits us to introduce qualitative reasoning about the conditions for establishing oscillations and the periodic, multiperiodic, or chaotic nature of the oscillations themselves (Fig. 35). This model allows demonstrating some of the basic nonlinear behavior of the clarinet, violin, and flute families oscillations (McIntyre et al., 1983).

The simplest case of the scheme of Fig. 34 is when the exciter is a unitary map and there is only the reflection filter connected in feedback to the delay line. If the reflection is also constant and unitary, we have a periodic repetition of the waveform initially stored in the delay line (Fig. 36, left), i.e. it results the table look-up oscillator from a different perspective.

If the reflection filter is moderately low-pass, the upper harmonics will decay faster than the lower ones. We can thus obtain simple sound simulations of the plucked strings, where the delay line serves to establish oscillations. This method is suitable to model sounds produced by a brief excitation of a resonator, where the latter establishes the periodicity, and the interaction between exciter and resonator can be assumed to be feedforward. This method is called *long-term prediction* or *Karplus-Strong synthesis* (Fig. 36, right).

A complete model of the clarinet is shown in Fig. 37, where the reed model can be an instantaneous map or modeled as a damped harmonic oscillator to take into ac-

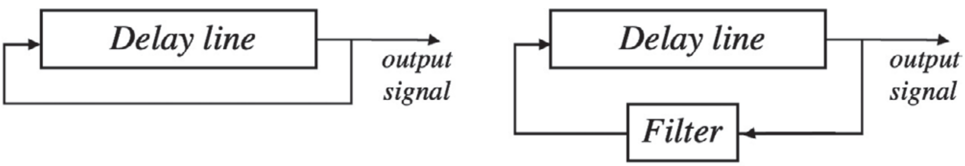

Figure 36. Feedback delay line as a table look-up oscillator (left) and as Karplus-Strong synthesis (right)

count the reed dynamics. The delay line represents the wave propagation in the bore; the reflection filter represents the tone hole lattice and the radiation filter represents sound emission from the tone holes and the bell.

4.2.4 Modal synthesis

In the panorama of methods for physical modeling, modal synthesis represents a rather interesting approach, particularly for implementing resonators. Modal synthesis uses system theory principles (modal decomposition) in order to implement a linear system with the parallel of second order systems. By doing so, a certain modularity and structural regularity are maintained (Adrien, 1991).

The impulse response of a resonating object is represented as a linear combination of the outputs of *N* damped oscillators, each of which represents one *mode* of oscillation (resonance) of the object excited by a driving force or acoustic pressure. In this sense modal synthesis can be interpreted a source-filter approach in which the source is the driving signal and the filter is a bank of second-order bandpass filters (Fig. 38). In the feed-forward structure, it can be realized (and interpreted) in both time and frequency domain, respectively as additive synthesis or as source filter models. By choosing a different frequency and damping factor for each oscillator (or conversely center frequency and bandwidth of each filter), it is possible to account for a set of partials and decay times of the resonator spectrum. The input modal parameters can be chosen to match those of an arbitrary object. Moreover, morphing between different shapes and material can be obtained by designing appropriate trajectories for these parameters.

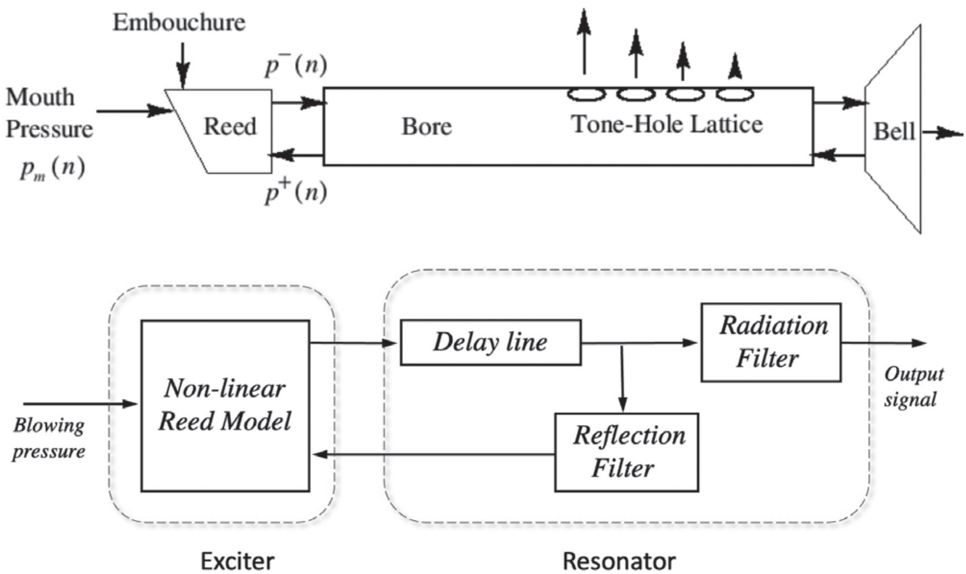

Figure 37. The clarinet (above) and a simplified computational model (below) implemented using the waveguide approach and the feedback structure.

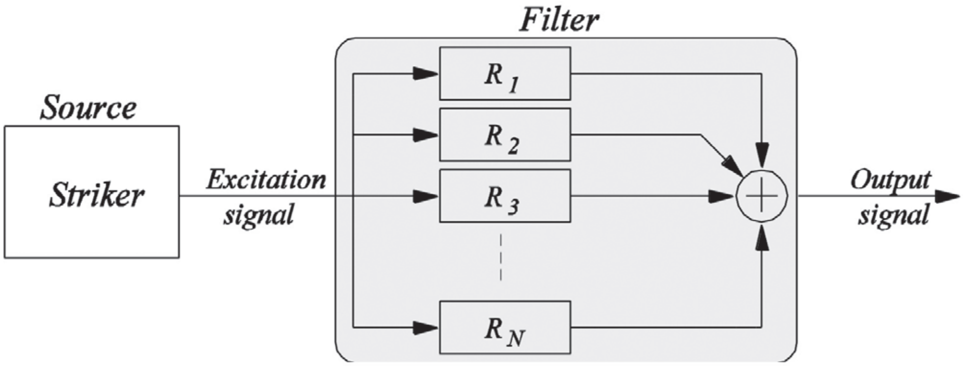

Figure 38. Modal synthesis implemented as a filterbank of 2nd order resonating filters.

Modal parameters can be analytically estimated for simple shapes or experimentally estimated from a recorded audio signal. In practice one will strike the object and then analyze the recorded response in air, with consequent spatially distributed interactions, and sound radiation through air. Modal synthesis is both effective and efficient for recreating the sounds of objects that exhibit a relatively small number of strong decaying modes. The time domain implementation (i.e. as sum of damped oscillators) is effective for impact sounds, while frequency domain implementation (i.e. as a filterbank) is effective for continuous excitation. It is often used to generate sounds for rigid bodies. The perception of size and shape of an object greatly depends on the distribution of the resonating modes.

What stimulates the interest of researchers and musicians in this method, besides the modularity of the resulting structures, is the frequency-domain representation. In fact, nonlinear physical models are normally represented in the time-domain, while sometimes frequency-domain conceptual representations can be more desirable for musicians. Another attractive characteristic of modal synthesis is its robustness. In fact, even though data are not obtained from theoretical considerations or experimental results, the system always produces acoustically coherent results, while other methods often behave in an unpredictable and unpleasant way when parameters are inappropriately chosen.

Notice that the modal formalism incorporates a "spatial" representation (e.g. it is possible to inject a force in a specific point of a modal resonator, or to measure its displacement in a specific point). Thus, the resonator can give a feedback information to the exciter for a proper coupling in the feedback scheme of Fig. 27.

4.3 Non mechanical sources

4.3.1 Virtual analog

In cases where the sound-producing (or sound-processing) device to be simulated is an analogue electronic system, rather than a mechanical or acoustic system, the term *virtual analog synthesis* is commonly used to refer to physics-based sound synthesis of these devices. Virtual analog is a type of physical modeling, which imitates the electronic properties of the circuit components rather than the mechanical or acoustical qualities of some device. It is used to emulate digitally the circuitry found in analog electronic devices or analog synthesizers, by implementing mathematical models of analog circuitry. These models can simulate accurately the non-linearity of the circuitry and the interactions that take place between all the components under different conditions. Because their parameters can be modified in real time by the user, the resulting synthesizers both sound and behave almost identically to a real analog synthesizer. The nonlinearities, which imitate the behavior of analog components, bring about pleasing distortion and compression, familiar characteristics of nostalgic music equipment. For example figure 39 shows the original electronic scheme (left) and the electric scheme digital simulated by CSC (right) of the device *Selezionatore di ampiezza* of Studio di Fonologia musicale of Milan.

4.3.2 Pseudo-physical models

One can observe that the physical model synthesis refers to the real world not just to draw inspiration from it to build conceptual models, but also to identify model parameters and to evaluate the results. It is often the case that the results of the model are compared with reference to sounds in the real world for the sake of a qualitative evaluation. If one refers to real-time simulation of musical instruments, the compromise required by the simplification of the model and its efficient implementation render the results (even when timbre-rich) less than satisfactory from the point of view of imitation. This approach has a scientific motivation, as it allows confirming the hypotheses relating to its functioning that are at the basis of the model, but it fails to exhaust the possibilities of their musical use.

Musicians have a clear intuition about the relationship between physical objects and the sound they produce. With this knowledge, the design of new objects is possible, and, therefore, of new sounds. Synthesis with physical models allows us to start out from this reality and create virtual acoustic objects that go beyond the physical

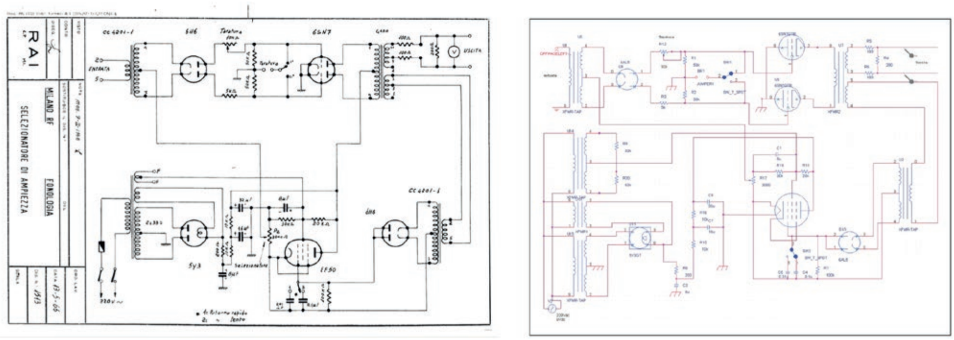

Figure 39. Original electronic scheme (left) and simulated scheme for virtual analog (right) of *Selezionatore di ampiezza* of Studio di Fonologia Musicale of Milan (Canazza et al., 2011).

reality surrounding us. In fact, the synthesis for physical models can be considered, from a more abstract viewpoint, as the generator of a musical reality in itself. It gives musicians projecting the sounds the possibility of drawing inspiration from the real world to create their own metaphors, without being bound to the usual physical laws in their experimentation. This allows them to focus mainly on the structures and only marginally on traditional metaphors. Structures, in fact, are responsible for the generation of classes of homogeneous sounds, while their conceptual interpretation may be considered useful or limiting, depending on the specific situations.

A typical case in which the physical interpretation is important is the parametric control of an instrument. In fact, the algorithms derived from physical models have the property of control "robustness". This property is derived from the fact that the physical model of an instrument has characteristics of energy passivity and stability that allows evaluating what the parametric variations allowed on the algorithm are and forecasting their effects on the final result. Therefore, the physical interpretation of structure has the advantage of facilitating the control of the instrument, while giving the musician a conceptual model of control that is very close to gestural experience. This leads to consider synthesis for physical models as a starting point and guide for the exploration of a whole class of algorithms, without limiting ourselves to traditional sound generation modes.

4.4 Models of objects for auditory interaction

In the common practice of multimedia systems, sound has a secondary role, and it is almost always subordinated to visual information. In real life, the hearing perception is very important and can be of use, other than in conveying acoustic messages, as a complement to sight perception, or as a reinforcement to messages coming from other senses.

Sounds, like images, adjoin a large amount of information when associated with the dynamic evolution of the objects they represent. Much of the communicative power of a picture or a sound is in the possibility of associating it to an *object*. Thus, it is evident how the availability of common models for sounds and images can increase the communicative properties of multimedia systems. In the case of visual communication, a model allows transformations of the image in order to get many different views. In the same way, for sounds we feel the necessity of building models allowing interaction with the sound object and overcoming the slavery to the "frozen" sounds. When the models for visual and acoustic "views" are alike, the communicative potential of the multimedia system is increased.

What stated above holds for the communication "to" the user, but it is even more important when we are dealing with communication "with" the user. For instance, a slider can be thought of as a static image, but if we are planning that the user will act on the slider, then we have to make an abstraction of the slider object, and associate to it a context-dependent view. By analogy, the acoustic event, which is produced when rubbing two objects, can be represented by a prerecorded sound. But if we want the user to be able to determine the contact pressure between the two surfaces, then we have to design a "rub" object and a model accounting for the possible interactions. What we said is particularly important when adopting new forms of human-machine interaction, involving touch or force information exchange. In this case the acoustic data can improve the realism in interaction and increase the resolution. A high-level model for all the multi-sensorial information is a key step towards a better coherence among the perceived signals. The idea is to shift the focus from sound models to *sounding object models*.

For all the purposes outlined herein, sound models should be accurate in their audio quality, versatile in their response under different conditions, and they must respond to the expectations of the user. These three features are essentials for object models for auditory interaction and an approach based on physical modeling is generally desirable.

We want to insist on a particular aspect of sensorial communication in multimedia systems: the possible unification between sound and image models. This unification is determinant for achieving a high coherence of aural and visual cues. Moreover, the effort for finding out unitary models can be worth doing, because it can simplify the analysis and engineering process in multimedia development. This approach is strengthened by the experience of artistic research, which has moved along the same direction during this century. Without necessarily aiming to a total art, many artists have been looking for forms of unification among different media, and they found them in generalized models. A significant example can be found in Kandinsky's effort for finding common perceptual models of color and musical timbre. Modern technologies allow proceeding farther than the perceptual unification, and they make it possible to look for a "source" unification, i.e. to base the multimedia models on common generative mechanisms (De Poli and Rocchesso, 1998)

5. Control models

We saw that the behavior of a synthesis model depends on its internal structure and on the input (or control) data. The problem of control in synthesis refers to everything that goes from the symbolic description of sounds, as expressed in the score, to the sound, using synthesis models. Traditionally, the score is a series of notes (symbols that describe a sound and its properties on an abstract level), and it is up to the player, with the help of an instrument, to turn this into sound. Therefore, control in synthesis occurs by coordinating symbolic information, discrete in time, and information that can be thought of as varying continuously (*control signals)*. Such control signals (often in computer music called *control functions)* can be generated by appropriate procedures or acquired from gestural interfaces. For a deep discussion on gestural control interfaces see (Wanderley and Depalle, 2004) and for the related research at CSC see (Canazza et al., 2022a).

Two levels of abstraction in control can, in general, be distinguished and which correspond to two different time scales. The first level, *sonological control*, controls the timbre evolution of the note and determines the passage from expressive parameters to the underlying algorithm. In this case the signals vary during the evolution of the note and operate along the time scale of its duration. Random and periodic frequency variations, in order to obtain a vibrato effect, are an example of this. The second, *expressive control*, involves the player as the interpreter. It refers to the passage from symbols to actions in order to choose and render the desired expressive effects. Generally, this does not mean just the simple transformation from symbol to symbol, but determines rather, the continuous variation of a set of parameters. Variations in the duration and amplitude of the note in order to emphasize the grouping of the phrase, is an example. This level consists, therefore, in the generation of signals that vary along the time scale of the musical phrase. The musician, thus, directs and shapes the flow of musical sound which form the entire work and acts in a similar way as the conductor of an orchestra. The idea of the quality of the timbre, i.e. the capacity of the instrument to produce "beautiful sounds", is associated to the first level. At the second level, the "playability" property, i.e. the possibility that the player is given to interact satisfactorily with the instrument is given priority.

Control signals differ from acoustic signals in several respects. For example, their frequency analysis does not seem to have any significant interpretation, therefore control synthesis and manipulation techniques are more suitable to be developed and described in the time domain. Despite this lack of parallelism, some sound synthesis techniques do have a counterpart in the synthesis of control signals.

5.1 Time domain control models

5.1.1 Abstract models

The control functions can be obtained with several procedures. When we have not reference to real signals, we have abstract models. One it to use arbitrary shapes, for instance some composers have tracked the shapes of mountains or urban skyline. In another case the functions can be generated by composition programs according to the composer needs.

5.1.2 Time-segment models

Sampling When, however, reference is made to real signals, then control signals can be derived from analysis of given sounds. These are successively used in resynthesis and with typical manipulations in the time domain. In a certain sense, this approach is similar to the use of sampling in synthesis. Control synthesis techniques based on recording-and-reproduction are characterized by the timbral richness of natural sounds and the expressiveness of acoustic instruments but, similarly to sound synthesis techniques based on sampling, they suffer from a certain rigidity in their usage. In particular, when expressive control signals are derived from the analysis of acoustic samples, all gestural actions are recorded, including those that are characteristic of the performer. Data reduction techniques can be used to abstract the desired and most perceptually relevant characteristics (Risset, 1969). Even though the possibility of modifying control signals appears as being minimal in the case of recording-reproduction, it is always possible to use such signals in a creative way, for instance redirecting some control signals to different control inputs. For example, the pitch envelope could be used for controlling the bandwidth. Once a variety of control signal samples are available, their impact on the timbre quality needs to be evaluated and interpreted in order to be able to use them in combination with other time-domain techniques such as cut and paste, amplitude or time scaling, etc.

Composite controls A simple synthesis model of control signals consists of combining several elementary signals through superposition or chaining or partial overlapping, as we saw for granular and additive synthesis. For example, in the case of sound frequency control, it is possible to add a signal that describes the general trend of the frequency (pitch) to a periodic slow oscillation (tremolo) and to other random or fractal variations. As far as timbral control is concerned, a control signal can be generated as a chain of different waveforms, each of which describes a different portion of timbral evolution and is selected among a collection of typical trends. For example, the attack-sustain-decay-release (ADSR) is a common model for control signals commonly used for modifying the amplitude envelope. This method consists of a chain of four signals that describe characteristic portions (attack, decay, sustain, release) of the timbre.

Interpolation Interpolation, both linear and nonlinear, is often applied in the synthesis of control signals. Since synthesis can be considered as a process that maps ”little” localized information into a continuous variation of a multitude of parameters, the concept of interpolation seems to be quite a natural way of approaching the problem of control parameter synthesis. Starting from the specification of a few significant samples of the control signal (for example, the pitch of the notes that constitute a musical phrase) an interpolation model generates an analog signal that is characterized by properties of smoothness and regularity characteristic of the interpolator.

5.1.3 Source-filter approach

A typical solution for the synthesis of parametric fluctuations through statistical models consists of filtering white noise with an appropriate linear filter (source-filter approach). The parameters of the filter can be estimated by analyzing given acoustic samples. This solution can be generally employed whenever it is not possible to make specific hypotheses on the control structure, although it is possible to extract a statistic description of it.

5.1.4 Hybrid models

As we said earlier, the reproduction of control signals has the same problems as those typical of sound synthesis based on sampling. In particular, the fact that the whole control function needs to be stored makes this approach not particularly versatile. In order to avoid the intrinsic rigidity of this method, one can think of modeling the control signal with a combination of a deterministic signal that models the average trend of the control signal and a random process that describes its fluctuations. It should be quite clear that, in this case, the statistical properties of the random portion are of crucial importance.

5.1.5 Fractal models

In some cases, variations in the control parameters are to be attributed to chaotic behavior of the acoustic mechanism of sound production (such as in an organ pipe driven by a jet of air). When this happens, control signal generators are well described by *fractal models*. What makes a fractal model interesting is the fact that it captures an important temporal characteristic of natural signals, *self-similarity*, that is, the statistical similarity of some temporal characteristics when viewed from different time scales. In particular, a single generator could be employed for simultaneously generating several signals that evolve on different timescales. For this reason, source-filter models are more suitable for modeling self-correlation on a short timescale, while fractal signals are better for modeling self-correlation on a longer timescale.

5.2 Physical models for control

Physical models can also synthesize control signals. In this case, the system is slowly varying and provides the dynamics for the evolution of the control signal. So far, however, this approach has been rarely used for the synthesis of control signals. Most of the available examples are meant for obtaining descriptive physical metaphors for musical processes, rather than for modeling existing mechanisms. For example, Todd (1995) suggested a model of a ball accelerating along a surface with several holes in it, for describing the expressive acceleration or slowing down of the musical tempo. Sundberg and Verrillo (1980) suggested the analogy between the final slowing down of a musical piece and a person that stops walking. Such models generate parameter variations that can be cognitively perceived as plausible and recognized as natural.

It should be noted that cellular models, when some elements evolve on different time scales, are particularly useful for obtaining physically plausible dynamic sound evolution, even at the level of performance control (Florens and Cadoz, 1991). By its principle, the physical modeling approach offers new ways of complex and relevant structuring with an intimate relationship between different hierarchical levels (Cadoz, 2009).

5.3 Gestural control

In its early days, computer music was produced offline by mainframes. In batch computer music, the only way to input data to the machine was through alphanumeric symbolic languages that represented the synthesis/processing algorithm and the sound parameters. Each expressive intention needed to be explicitly formalized in the input data. The performer's gesture was not taken into account: the composition of the sound and the composition with sounds replaced the traditional performance with instruments. The musician had to select and adjust the synthesis model and to formalize the control parameters and signals. Generally, it can be observed that the synthesis of control signals occurs very often at low levels of abstraction or using rather simple procedures. Few models have been put forward to describe and generate control signals. Not much research has be done to formalize what knowledge and experience gained in synthesis techniques and to identify models that are general enough to allow the musician to turn her/his attention to controlling the control signals and, therefore, operate at a more abstract level.

In the 1980s, real-time music systems have begun to be designed, which allowed an increasingly effective interaction between the performer, the machine and the listener and fostered new performance practices. The availability of real-time synthesis gradually shifted the attention to gestures as a source of control functions. Gestures are an easy and natural way for controlling sound generation and processing. A precursor was the Groove system by Mathews and Moore (1970) that allowed real time control and editing of performer actions described (graphically or symbolically) by time functions. In digital instruments, the physical interface of the control device is disjointed from the sound generation and processing system. This offers the designer and the musician the possibility of establishing the correspondence between the interface and the generation system that best suits their musical ideas, as well as to vary it according to their own expressive needs. In this way, the complete behavior of the instrument can be changed. The focus of the research became the *mapping* of detected actions of the performer to input parameters of the synthesis engine. While the input devices determine the mechanical and acoustic constraints of the system and the sound engine affects the sound characteristics, the mapping defines the performative qualities of the instrument (Wanderley and Depalle, 2004).

It is advisable to decouple input gestural data from sound model control data in two (or more) steps through layered mapping, which includes an intermediate representation of parameters. The intermediate information can be at different representation levels, such as physical, perceptual and even conceptual (Wanderley and Depalle, 2004). We are still in the early stages of understanding the complexities of how the mapping layer affects the perception (and the playability) of an electronic instrument by its performer.

6. CSC research on sound models

Since 2022 marks the 800th anniversary of the University of Padova this work also aims to be a review of research in the field of music technologies at Padova University, focusing on scientific and musical research on sound models during the Seventies and the Eighties.

6.1 Background

The first attempts to use electronic technology to produce music at the University of Padua date back to the late Fifties, when an innovative photoelectric organ was designed by Giovanni Battista Debiasi, in which oscillations were produced by a rotating disk with slits that periodically modulate the light reaching a photodiode. The different envelopes of each harmonic and organ stop were produced by modulating the light intensity through a sliding window, in a sort of additive synthesis (Debiasi, 1959). In the late Sixties the Group *Nuove Proposte Sonore* (NPS), led by Teresa Rampazzi, introduced analog electronic music in the Conservatory of Music in Padova, and then began collaborating with the university in the field of computer music, where a group of researchers and musicians were working on speech synthesis and computer music research (Dashow et al., 1978; Debiasi et al., 1984). In 1979 these activities have been institutionalized by the university with the creation of the Centro di Sonologia Computazionale (CSC). The founding members were Debiasi, Giovanni De Poli, Graziano Tisato and Alvise Vidolin. Since the beginning, CSC has been a leading research center in the field of computer music. Musical creativity has been a stimulus to pursue new paths in scientific and technological research. Over the years, CSC's research has addressed several topics, described in detail in (Canazza and De Poli, 2020), which include sound processing, expressiveness, multimodal interaction, and musical cultural heritage. On the other hand, theoretical and scientific achievements have been constantly applied to music production. More than two hundred musical pieces have been made at CSC and have received wide recognition (http://csc.dei.unipd.it/multimedia-works/).

In 1979 a course on computer music (*Musica all'elaboratore elettronico*) was instituted by the Faculty of Engineering, one of the first such courses in the world dedicated to technology students. The course evolved with different names and programs, keeping up with the state of the art of international computer music research. The need to teach sound synthesis to advanced students has given impetus to sound modeling research.

In 1980, CSC and Venice Biennale (on strong drive of Mario Messinis, director of the Music Sector) converged into the creation of the LIMB (Laboratorio per l'Informatica Musicale della Biennale), a structure that offered for a decade the most fruitful opportunities for scientific research, music production and dissemination of computer music. The focus of basic scientific research in this decade was on instrumentality, i.e., allowing interaction in real-time processing and the categorization of sound classes by new synthesis algorithms and perceptual timbre spaces.

Since the 1990s, with the advancement of real-time synthesis by special processor and later by software the new focus of CSC research was on the exploration of expressiveness and performance. The goal was to overcome the rigidity typical of early computer music, to render the many expressive nuances introduced by a performer while playing a piece of music. A general account of CSC scientific and musical research can be found in (Durante and Zattra, 2002; Canazza and De Poli, 2020; Canazza et al., 2022a).

6.2 Scientific and musical research at CSC

6.2.1 Systems for computer music

In the Seventies, the first research objective of CSC was to develop a system for computer music that would provide researchers the opportunity to operate in an integrated manner both at the score level and at the sound level. The first musical sounds from computer that could be heard at CSC in the beginning of 1974 were a melody played by an organ simulation by time-segment processing.

The main concern was to create a complete system, easy to use and flexible in application, for producing music with the equipment of Computing Center of the university, an IBM mainframe connected to an IBM System/7 for high-quality four-channel D/A conversion. For batch synthesis MusicV and Music360 programs were used (Dashow et al., 1978). Languages for alphanumerically traditional scores encoding and for computer-aided composition were developed. Tisato (1976) created the ICMS (Interactive Computer Music System) in 1976 for interactive synthesis. The system operated in a multi-programming environment and its principal purpose was to develop a single environment suitable for the processing from any sound source, being it acoustic or synthesized, vocal or instrumental. It provided real-time synthesis, editing, and mixing of selected musical material, reverberation and spatial distribution on four channels, LPC sound analysis and synthesis using any sound source as the stimulus. Particular care was given to man-machine interaction through simple commands and graphic visualization. ICMS provided an easy introduction to computer music particularly for non-specialists. The system was successfully used in the production of many music works, for acoustic and psycho-acoustic research and for educational purposes.

In the 1980s CSC, together with the Institut de Recherche et Coordination Acoustique/Musique (IRCAM) in Paris and the Laboratorio per l'Informatica Musicale della Biennale (LIMB) in Venice, developed the 4i System, designed by Giuseppe Di Giugno. The system is based on digital processors for live electronics with four DACs, two ADCs, and a control interface for performance parameters (Canazza et al., 2022a). This system was used to play real time sound synthesis in one of the most important musical works of the second half of the 20th century, Luigi Nono's *Prometeo* (1984) and for Salvatore Schiarrino's *Perseo e Andromeda* (1989) which is the first musical theater work with only real-time digital sounds.

6.2.2 Models and algorithms

Time-domain synthesis research CSC has been a pioneer in time-segment synthesis. Research on speech synthesis began in the early 1960s at the initiative of Debiasi. At that time the most common approach was frequency-domain simulation of the phonation process by time-varying filters, such as the vocoder. Computers had very limited computing power and memory space, however. Thus, Debiasi's approach was concatenative synthesis in the time domain. Debiasi's idea was to develop a real-time text-to-speech synthesis for the Italian language by experimentally identifying a minimum set of elementary speech segments (phonemes or parts of phonemes) which, when appropriately recombined, allowed the synthesis of any message, keeping intelligibility as the main goal. With about a hundred speech units, the system produced very intelligible speech synthesis, even if of a robotic quality (Francini et al., 1968). Later the system was expanded for German, Greek and Serbo-Croatian languages (Stathopoulou et al., 1980). The speech unit system was employed by De Poli and Vidolin in the first computer piece of CSC *Consonantico* for computer voice and electronic processing (1975). Moreover, a concatenative synthesis of singing was realized by resampling the vocal units to obtain vowels with appropriate pitch and duration (De Poli, 1975).

A system for the automatic translation from any Italian text into naturally fluent speech was developed in the 1980s. It was built up around a phonological processor, which mapped the phonological rules of Italian into prosodic structures, and of a synthesizer, which processed and joined LPC coded diphones, derived from the previous research on concatenative speech synthesis (Delmonte et al., 1984). An interesting application of the system involved referring to the timbre characteristics of the diphones archive as a spectral vocabulary for controlling musical sound synthesis (Mian and Tisato, 1986).

In music research additive synthesis was used by Fausto Razzi, in the piece *Progetto Secondo* (mono tape, 1980), where he developed the sound element of the sine wave to attain an accomplished formal structure. The composer's requirement was to work with sound spectra made up of three frequencies, organized like triplets of a piano's strings (beating each other). In relation to the location in the range of musical pitches, the values of the fundamentals were determined. He used additive synthesis on a quarter-tone scale, with frequency values such as to generate first-order beatings (Razzi, 1981).

Winter Leaves (stereo tape, 1980) by Mauro Graziani is an investigation of the relationships between harmony and timbre. The basic technique is to construct sounds whose spectra have harmonic valence. One of the systems to achieve this is to create sounds whose spectral components have chordal relationships. *Winter leaves* has been realized through additive processes starting from simple sounds (sines); it is a system that creates complexity through the superposition of simple elements. Each of the spectra, generated with a simple additive synthesis with 5 oscillators, is then enriched from the acoustic point of view thanks to a ring modulation process among the five basic components taken two by two. In this way, other frequencies are obtained which are used both as spectral components and as 'notes' on which to articulate the melodic / harmonic discourse. The piece is controlled both in its composition and in its synthesis by simple patterns. It progresses from the generation and accumulation of fundamental materials to their definitively structured organization at the end (Zattra, 2003a, 2004).

Agostino Di Scipio had from the very beginning a need for a blending of musical form and sound matter. He saw the potential of fractals and in their numerical methods as a useful front-end for granular synthesis At CSC he implemented, with

Graziano Tisato, granular synthesis and granular processing on the local mainframe computer system, using several nonlinear maps as a front-end control structure (Di Scipio and Tisato, 1993). Then in the composition *Fractus* (for viola and computer-generated sounds, on stereo tape, 1990) he experimented the non-linear iterated maps, which in science allow modeling of chaotic systems, for the formal construction of the piece (Di Scipio, 1990). The electronic sounds were generated with a form of synthesis by frequency modulation and support the instrumental part, reflecting on the whole the dialectics of order and disorder despite the economy of a rather reduced set of musical elements.

In the same period, Bernardi et al. (1997) used analysis methods based on chaos theory to study the self-similarity properties of music and control signals. The local fractal dimension (LFD) yields a good description of the fluctuation of sound waveform, exploiting the geometrical characteristics of its time graph. The local fractal dimension proves its usefulness in its scale- and time-varying formulation. Analysis of musical signals in phase space with chaos theory was shown to provide important information on the sound production mechanism.

Among the very systematic works we find *Parafrasi* (for computer processed voice, stereo tape, 1981) by Aldo Clementi composed using canon techniques, starting from recorded material (by the soprano Liliana Poli). The source material consists of a melodic module, sung in three different tonalities and with three different metronomes. In the synthesis the elements are transposed, retrograded (by reading backward the sound files), concatenated, and finally superimposed with various input delays, until the whole sound is made up of six canons in three voices each, three in straight motion and three in retrograde motion (Clementi, 1982; Graziani, 1982). The technique is in a way a musical experimentation of sampling with transformations we have seen in Sect. 3.1.2. He notes that the computer allows microvariations due to vertical coincidences and ever-changing horizontal proximities or departures: the sounds very close to each other determine variable and changing frictions around a single cluster and, stylistically, an equally changeable state of continuous sound vertigo (Clementi, 1982).

Time-segment models In time domain, various synthesis techniques have been investigated, in particular those utilizing VOSIM-type oscillators or non-linear techniques. The problem of finding parameters for a generalized VOSIM oscillator and for waveshaping has been studied. For this, equivalence relations were determined for shaping polynomials which produce the same spectrum, but have different dynamic behavior (De Poli and De Poli, 1976, 1979; De Poli and Faccio, 1989).

De Poli collaborated with Aldo Piccialli of the University of Naples working on time-domain algorithms for sound synthesis. Different strategies were proposed to bring methodologies and techniques of digital signal processing into a granular synthesis context (De Poli and Piccialli, 1988, 1989). With this goal in mind, they developed several waveform design and transformation techniques for granular models (*Pitch Synchronous Granular Synthesis)* to produce sounds with time-varying formant regions (De Poli and Piccialli, 1991).

Non-linear models The non-linear distortion model was studied for input consisting of the sum of two sinusoidal signals and a particular waveshaping was examined in which the spectrum produced also depends on the input frequency (De Poli, 1984a). The use of a distorting function specified by the ratio between two polynomials was then proposed (De Poli, 1984b). Finally, frequency modulation was also an object of investigation. The use of phase or frequency series and feedback modulators was studied (De Poli, 1983), as well as special discrete modulation with phase distortion (De Poli and Piccialli, 1984).

Music compositions with non linear techniques Non linear techniques has been a source of inspiration in music research. Among the early CSC pieces, a first attempt, in terms of control, to overcome the limits of the 1970s computer technologies is well represented by Teresa Rampazzi's *With a light pen* (stereo tape, 1976), composed on the computer with the Interactive Computer Music System (ICMS) of CSC, which used a light-sensitive computer input device on a CRT display to create the synthesis sounds and organize them in time (Tisato, 1976). The system allowed real-time FM synthesis, spectral analysis, sound editing and mixing (Fig. 40).

In 1978 Teresa Rampazzi composed *Computer dances* (4 tracks tape), a piece that studied timbre using frequency modulation synthesis. This work is the result of Rampazzi's tone research applied to rhythms at different speeds. The signals at the beginning are slow moving in order to let the transformations of tone modulations to be perceived. Only towards the end the signals become shorter and very articulated; the number of which reach voices or "dancers" superimposed. The two pieces obtained a special mention at the International Electroacoustic Music Competition in Bourges (France) in 1977 and 1978, respectively (Zattra, 2003b).

In the same period, James Dashow proposed a theory based on a timbral model in which "... the nonharmonic domain of frequency relationships may in some way contain a necessary system of hierarchical structural functions." This approach utilized various modulation techniques as a beginning point for generating "... a group of chords that have meaningful and possibly necessary functional relationships ..." (Dashow, 1980). A single pair of pitches, the generating dyad, is made to generate sev-

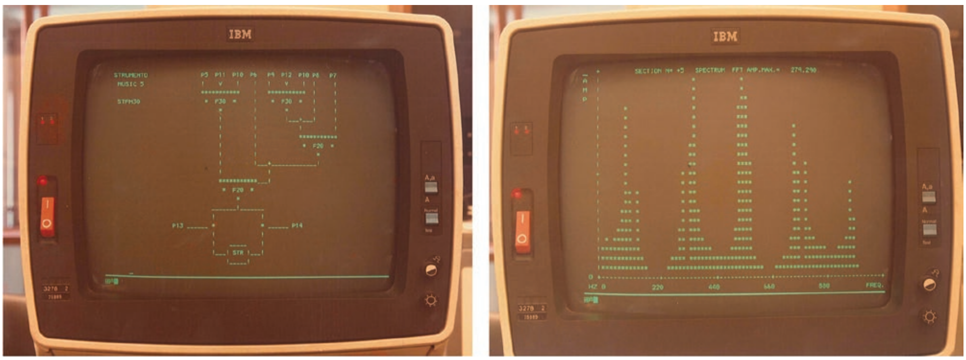

Figure 40. The ICMS system: FM synthesis (left), spectral analysis (right).

eral inharmonic partial series, all of which have the generating dyad in common. The resulting spectra can be interpreted as chords that "harmonize" the generating dyad.

This approach to timbre construction offers a means for associating timbral distinctions much more closely with other elements of compositional structure. The choice of timbre is in this way more organically related to other aspects of the composition insofar as the timbre is made to be the direct result of those pitches or frequencies articulated at a given moment (Dashow, 1978). Background structure can be composed in terms of invariant groups of discrete frequencies (traditional pitches, for example), while surface elaboration and prolongation can be achieved by generating inharmonic partial and scalar systems from the basic frequency or pitch structure grouped in pair. This technique was applied to ring modulation, amplitude modulation, and a controlled use of the foldover phenomenon.

Effetti collaterali (for clarinet and stereo tape, 1976) is the first piece that makes use of the notion of harmonizing dyads or triads of pitches with frequency-modulation spectra. *Conditional assemblies* (4 tracks tape, 1980) is a more complete application of this idea of starting with two or three pitches and working backwards to discover which frequency relationships between two signals would be necessary in order to realize a series of spectra harmonizing the generating pitches. The piece

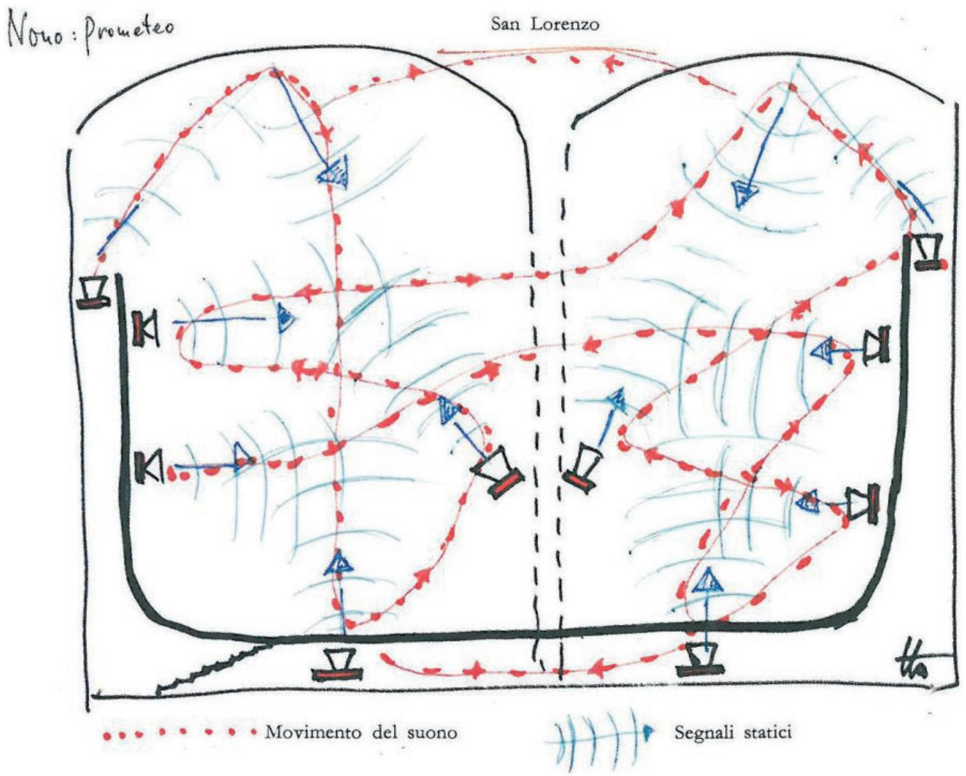

Figure 41. *Prometeo, tragedia dell'ascolto* by Luigi Nono: Sound movements inside the wooden resonating structure, as drawn by Hans Peter Haller, sound director.

was rewarded with a second prize at the 1981 Bourges International Electroacoustic Music Competition.

A real leap in the quality of CSC music works was achieved in the early 1980s when a collaboration with IRCAM, in Paris, and the Venice Biennale is started to build the 4i System, a processor capable of generating sounds in real time. The 4i System was immediately used by Luigi Nono as a musical instrument in his opera *Prometeo, tragedia dell'ascolto* (for vocal and instrumental soloists, choir, orchestra, live electronics, and 4i System, 1984), which is definitely one of the most important musical works of the second half of the twentieth century, and in which technology plays a dominant role.

At its first performance of the opera, which took place in the Church of San Lorenzo in Venice, the synthesis sounds of 4i System were the first to invade the acoustic space of the church, emerging from the bottom of the wooden structure to spread in all directions followed by the orchestral and chorus sounds, thus starting the piece (Fig. 41). Among the composer's various requests, a gesture-controlled performance environment was created that could simulate a "coro velatissimo" (very veiled chorus) used in different sections of the opera. Therefore, the computer did not have to play a simple melody, but rather generate several sets of sounds that moved over time, and could be aggregated around one or more musical pitches, and at the same time changed the timbre both in a harmonic and inharmonic sense. The fluidity required by the transformations led us to prefer the potentiometer as a sensor for controlling the musical gesture, and to choose the frequency modulation sound synthesis technique for high efficiency and versatility in timbre control (Vidolin, 1997; Canazza et al., 2022b).

Advanced algorithms for audio restoration The last four decades at CSC have seen the realization of many musical works. As a result, the problem of preserving these works for posterity arose. Researchers at CSC addressed the problem of improving existing algorithms for digital restoration, not only for simple denoising, nor purely for the aesthetic of digital silence, but rather to tackle the issue in terms of computational efficiency and quality of results, and extending their applicability to sounds and music that have been relatively neglected, such as electroacoustic and computer music (Canazza and Vidolin, 2001).

CSC developed a time-domain algorithm, based on the Extended Kalman Filter (EKF), that was optimized for music restoration and simultaneously solves the problems of broadband noise filtering, signal parameters tracking, and impulsive noise ("clicks") removal by properly reconstruct the lost signal (Canazza et al., 2010).

In frequency domain, an algorithm based on Short Time Spectral Attenuation method for broadband noise removal was proposed. Special consideration was paid to the perceptually relevant characteristics of the signal and the psychoacoustic masking effect of the ear. To filter the noise in a perceptually meaningful way, it was proposed to transform the audio signal from an "outer" to "inner" representation, i.e., into a representation that takes into account how the sound waves are perceived by the auditory system. By properly filtering the inner representation, only the audible noise components can be removed, preserving the signal from possible distortions caused by the restoration process (Bari et al., 2001).

Frequency domain research

Subtractive synthesis *Perseo and Andromeda* (for four voices and synthetic sounds, 1989) by Salvatore Sciarrino is one of the first works in the history of musical theater in which the traditional acoustic instruments are completely left out in favor of digital sounds (played in real time, not recorded on a fixed tape) by two musicians playing a network of four computers running a live electronic software developed by CSC. Particularly innovative was the sound generation technique based on subtractive synthesis. It was not intended as an adaptation to new media, but Sciarrino meant to design it with computers, creating a particular surreal estrangement, inseparable from the fantastic dimension of such music. Inseparable from the electronic part, the vocal part was organized as an extension of the ambient noise, but in such a way as to leave intact the intelligibility of the words (Vidolin, 1997).

The synthetic sounds are not an imitation of the traditional acoustic ones, they are, in fact, anything but orchestral. The aim is to create an abstract musical game or the sounds are designed to suggest the sonorous panorama of the island of Andromeda: the wind, the sea, the seagulls, the horizon, the pebbles, drops of water, etc. All the sounds arise from a single algorithm synthesis that, in its most elementary form, consists of a white noise filtered through a second order low-pass resonant filter. Therefore, amplitude, cut-off frequency and the filter bandwidth are the parameters of this algorithm, which can be conveniently varied according to a specific control function, written on the score (see Fig. 42). The performance environment, developed by the CSC, allowed

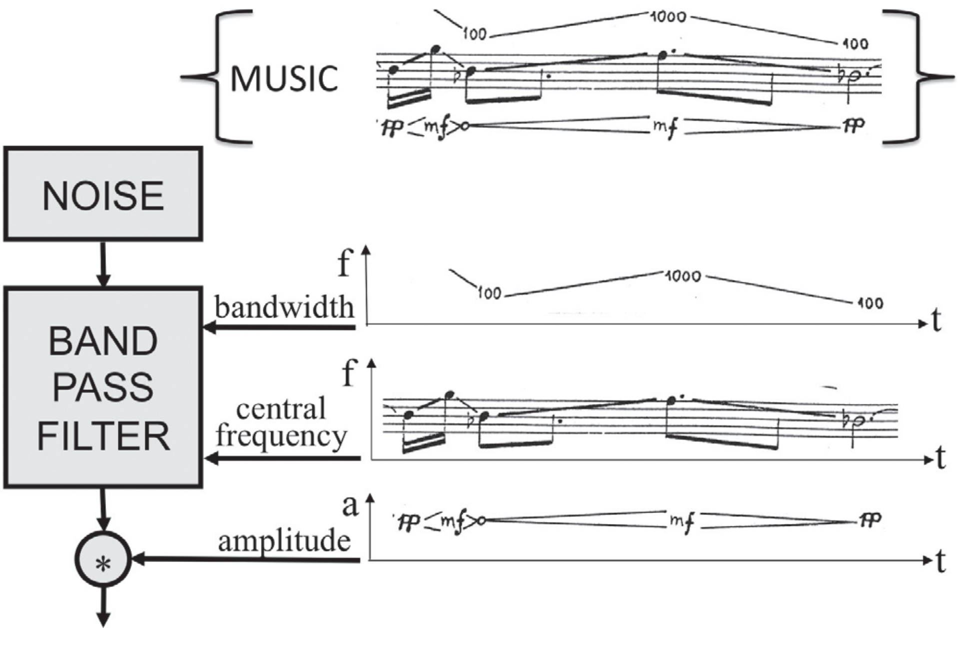

Figure 42. Subtractive synthesis algorithm employed in *Perseo and Andromeda*. The control functions are shown as written in the score.

for the movement of the sound in space. Most sounds, according to Sciarrino, should consist of sound objects passing over the listeners' heads, starting from a distant horizon in front of them and then disappearing behind them. In other cases, the sounds must completely envelop the listener, to give the impression that the sounds are coming from everything around them (Vidolin, 1991).

Richard Karpen realized *The vision* in 1985 (4 tracks tape) by exploiting the vocal sound data with temporal variations in the creation of a broad scale of dynamic vocal timbres. The piece used LPC voice synthesis on ICMS system and Music360 program, by transforming the speech parameters as in Fig. 15. He experimented with the relationship between text and music and the synthetic voices which resulted represented the confused memory of the writer. *The Vision* is in three parts, each preceded by a portion of spoken text taken from Doris Lessing's novel, *Briefing for a Descent into Hell*. The text simply (if dramatically) describes or recounts a vision or experience that has taken place. The music is concerned rather with the process of discovery or recognition. In the music of *The vision* the focus is elusive, but moments of clarity arise, only to fade again in the general fabric of things (Karpen, 1987).

Spatial audio For reverberation, circulant feedback delay networks (CFDN) were studied as a generalized model of a resonator. CFDNs retained the main advantage of physical modeling techniques, namely the availability of physically meaningful parameters like size, absorption, damping, diffusion, etc. At the same time, the CFDN model was sufficiently general that it could be used as instrument resonator, post-processing filter, or reverberator (Rocchesso and Smith, 1997). In the piece by Maurizio Pisati *Zone I (Zone-hack with virtual direction)* for alto flute, sound tube, MIDI electric guitar (1995), the system was used to divide the acoustic space of the hall longitudinally and, through a central corridor of speakers, a virtual sound path from the back of the stage to the back of the audience and vice-versa was created. The Doppler shift affecting noisy sounds was a desirable side-effect of the spatialization system, as it magnifies the direction and velocity of movements (Rocchesso, 2002).

Research activity at CSC included the development of innovative techniques for spatial audio synthesis with particular attention to binaural audio synthesis and real-time audio rendering. Spagnol et al. (2013) analyzed the contribution of the external ear in relation to specific and individual Head-Related Transfer Functions (HRTF) and modeled the physical features that had a perceptual interest for vertical localization of sound. Efficient real-time algorithms were developed for spatial sound rendering for a coherent simulation of complex multi-source acoustic environments where the spatial positions of both the listener and the sound source were expected to move dynamically.

Physical models

Algorithm structure To experiment with the effectiveness of physical modeling approach, researchers at CSC initially studied efficient algorithms for the simulation of specific musical instruments and the main mechanisms of sound excitation. Research on synthesis with physical modeling continued with the definition of the concept of a generalized exciter and resonator as a unifying element. This structure, shown in Fig. 43, inspired the realization of most of the classical mechanical and fluid dynamic exciters of musical instruments, and of pseudophysical exciters (Borin et al., 1992a).

Computational problems often occur in physical modeling synthesis. The feedback scheme is able to describe the behavior of an instrument very accurately because the interaction between exciter and resonator is the principal element responsible for the timbre dynamic evolution. CSC addressed the problem arising often in the feedback structure when in both blocks the output is instantaneously dependent on the input from the other block. In fact, in this case, there is a closed computation loop without delay, which gives rise to problems of noncomputability. This issue was addressed, and two novel methods were devised: the *K-method* (which used geometric transformations of nonlinearities and algebraic transformations of equations in the time domain, (Borin et al., 2000))), and a generalization of the formalism of the wave digital filters applicable to nonlinear elements (Sarti and De Poli, 1999). This latter proposal was well matched with the waveguide models that were widespread at that time in musical instrument simulations (and remain in use today).

Another problem that arises when analyzing the structural properties of the feedback scheme is how to guarantee real compatibility between the systems to be interconnected. In fact, normally, the feedback scheme does not allow blocks to be built independently of each other. To cope with this problem, the EIR structure with a third block, in between the two blocks exciter and resonator, was proposed (Borin et al., 1992a). This is an interconnection block and it can convert the variables used in the exciter to the variables used in the resonator, or avoid possible anomalies introduced by the discretization process. The idea is to have a sort of adaptor for connecting different blocks in a modular

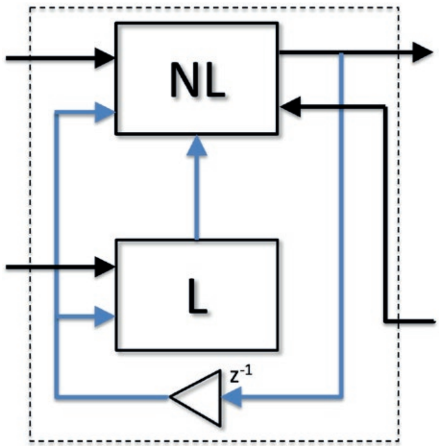

Figure 43. A model for a generalized exciter. The structure of the exciter is given by the interconnection of an instantaneous nonlinearity (NL) and a linear (L) dynamic system. The task of the linear part is to add "memory" to the nonlinearity. A unit delay (bottom element) ensures computability.

way (Fig. 28). This adaptor might also serve to compensate the simplifications introduced by the modeling and digitalization processes, and was successfully used in many acoustic instruments simulations, such as reed instruments (Balena and De Poli, 1985; De Poli and Puppin, 1989; Avanzini and Rocchesso, 2002; Avanzini and van Walstijn, 2004; van Walstijn and Avanzini, 2007), piano (Borin and De Poli, 1996; De Poli and Rocchesso, 2002; Bank et al., 2003), flute (Magalotti et al., 1995) and membrane instruments (Avanzini and Marogna, 2010; Avanzini et al., 2012).

In order to achieve a satisfactory audio quality of the source models, it is often necessary to use structures far more complex than that of the simplified clarinet of Fig. 37, below). A first fundamental step is considering the dynamics of the exciter; this implies the introduction of memory elements inside the non-linear block. When physicists study the behavior of musical instruments, they always use dynamic models of the exciter. A non-trivial task is the translation of these models into efficient computational schemes for real-time sound synthesis. A general structure has been proposed, that allows good simulations of large instrumental families. In figure 43, this structure is schematically depicted. The block *NL* is a non-linear function of several variables, while the block *L* is a linear dynamic system enclosing the exciter memory (Borin et al., 1992b). Studies and simulations have shown that reeds, air jets, bows and percussions, can all be represented by this scheme.

A new one-mass model of the vocal folds was proposed for articulatory speech synthesis. It requires significantly lower computational resources and about half of the control parameters with respect to the reference two-mass model, which makes it suitable for real-time implementation (Avanzini, 2008).

Few composers at CSC used physical modeling in their compositions. In *Dialodiadi* (1995), for flute, clarinet, and Yamaha VL-1 (the first physical model synthesis commercial instrument), Diego Dall'Osto doubled the two acoustic instruments into virtual instruments capable of humanly impossible virtuosity by exploiting the synthesis for physical models. The work, as the title tells, involves "dialogs" and "dyads" between real instruments and synthesis instruments that simulate the flute and clarinet. The synthesis instruments interact with the real ones, extending their possibilities without losing touch with their cultural reality.

Virtual analog by physical modeling There are several different kinds of musical instruments, each with peculiarities that have to be preserved and communicated. However, all of them share a common characteristic: to be understood they have to be played. In the 2010s the CSC defined a methodology for designing interactive multimedia installations for presenting old musical instruments (acoustic and electrophones) in museum settings, without losing their cultural context.

An example of such an approach is the installation, conceived by CSC, for the Music Instrument Museum in Milan, where the original electronic devices of the Studio di Fonologia Musicale (RAI, Milan, Italy) are conserved and exhibited. The Studio represented one of the European centers of reference for the development of electroacoustic music, that radically changed the way of producing music and listening to it. Despite the technology involved, electroacoustic music of that times has a non-obvious gestural component. To transmit the experience of how such music was conceived and created, the installation comprises a tangible copy of the devices, which replicate their physical appearance and behavior, and generates the sounds by virtual analog simulation of the internal electronic components. The visitors can feel the original experience of playing and producing early electronic music, by acting on switches and knobs of the installation in the same way as electronic musicians did in the 1950's and 1960's. Virtual audio allows to retain the idiosyncratic behaviors and possible imperfections of the original instruments (Avanzini and Canazza, 2015).

Music research inspired by sound properties Synthesis was the preferred technique at the CSC for work on the control of the microscopic dimensions of sound. Psychoacoustics, cognitive sciences and musical acoustics were sources of inspiration for many researchers and composers at CSC.

Music based on perceptual research During the '80s Roberto Doati realized many works investigating perceptual rules. Since the use of psychoacoustics has proven valid for sound perception research, an investigation employing psychology might provide insight into and understanding of the principles that control the perception of musical form. In *Gioco di velocità* (stereo tape, 1981) the descriptive principles formulated by the Gestalt school were used: closure, similarity, good continuation, etc. The law of good form, for example, favors the choice of simple, regular, symmetrical forms, and such are the structures used. The pure sounds that constitute them (sine waves with slight deviation in frequency) lose their identity to give rise to an organic "whole". The piece was selected at the first edition of Opera Prima (Teatro La Fenice, 1981) for young composers.

In composing *Una pulce da sabbia* (stereo tape, 1982) Doati turned his attention even more to timbre. Research on this extremely complex perceptual attribute at that time found its most powerful investigative tool in the computer. Through the synthesis models used (waveshaping and FM) Doati was able to control the three perceptual dimensions of the timbre space, i.e. attack time, brightness, spectral flux (Doati, 1983). Figure 44 shows a detail of the score. In this work the perceptual space of the timbre became the ordering principle for composition.

This musical approach prompted CSC researchers to investigate the physical and perceptual relationships that exist between sounds and to explain the main factors that differentiate timbres. The derived timbre spaces (based on physical or auditory features) supported the importance of the features of the steady-state portion when evaluating timbre quality and confirmed the importance of the attack for recognizing sound sources (Cosi et al., 1994; De Poli and Prandoni, 1997).

Music research: focus on spectral models The ICMS system allowed the spectral analysis of all the sound partials while also indicating the approximate pitch on the temperate scale. The relationships between the acoustic and musical properties of the sounds could thus be highlighted, thus fostering the composers' focus on spectral models. The 'analysis-resynthesis' at that time consisted of the spectral analysis of a

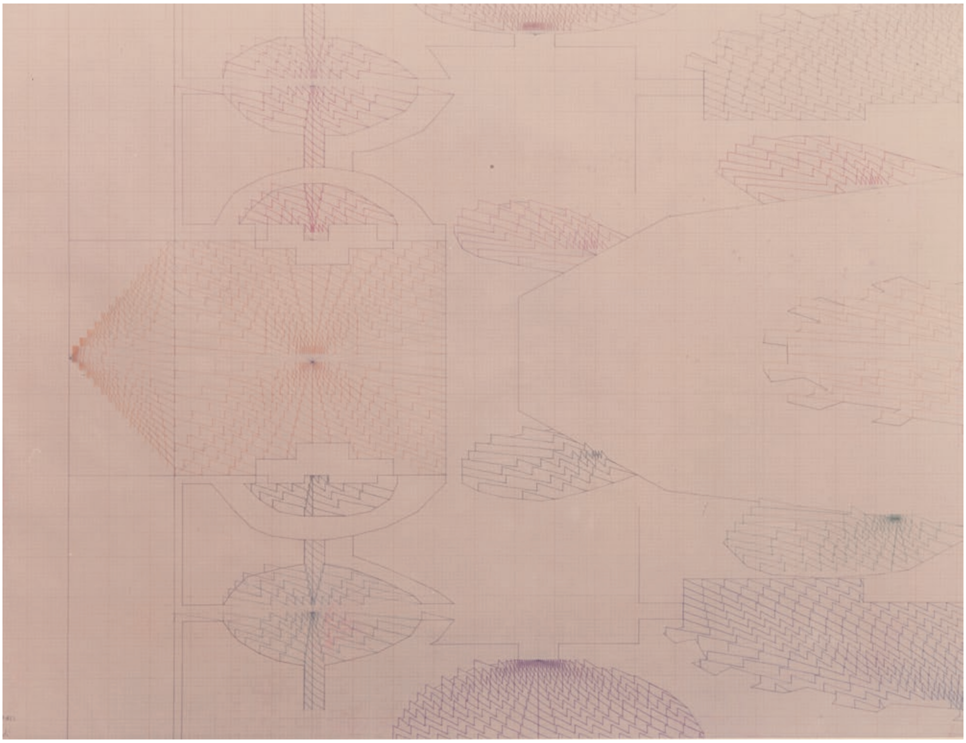

Figure 44. Detail of the score of Roberto Doati's *Una pulce da sabbia* (1982).

sound followed by a non-automatic resynthesis, that is, through the visual selection of certain parameters. In early compositions, mainly one single technique was experimented with. Then composers began to use all those available, as if it were an orchestra with many instruments.

The piece *Cadenza estesa e coda* by Claudio Ambrosini (for amplified flute and stereo tape, 1981) focuses not on the beginning phase of the sound, that "attack transient" that is so important in distinguishing one instrument from another, and on which all our attention usually converges, but on the opposite portion, the one in which the sound gradually loses energy and moves toward silence. From the timbre point of view, Ambrosini aims to develop the signifying capacity of the electronic medium, through a series of modes of articulation: initial cadence of the flute, orchestral response of the computer, coda in which the two instruments interact on a deeper plane of relationship (Ambrosini, 1982). The acoustic elements are bands of colored noise and sounds composed of spectra with non-harmonic components, also using the foldover (Torresan, 1982). Particular care was given to the spatial element: it is as if the center of the scene (the one where the flutist stands) is unharmed, natural, and safe. The sound there is beautiful, pure and lives its own existence. But then the sounds, moving through space thanks to the electronics, are pushed to the edges of the stage, which are instead "active," corrosive, sometimes destructive to the sound waves that come close to them, which are then distorted, as if burned by the contact. As if in a painting the colors are pushed toward the border marked by its frame (Ambrosini, 1982).

In *Sotto pressione* (for 2 oboes and computer generated sounds, 1982) Wolfgang Motz wanted to create many different relationships between the live instrumental sounds and the electronic sounds on tape. To this purpose, he studied the spectral structure of oboe multiphonic sounds to develop the musical material. These multiphonics usually arise from the anomalous opening of a hole in the middle of a fingering that covers several holes and are characterized by audible partials at the sums and differences of the two base frequencies. Interestingly, the structure of such multiphonic sounds is similar to that of sounds generated by means of the frequency modulation technique. This similarity also theoretically explains why these multiphonic oboe sounds often give an impression of electronic sound. Amplitude and frequency modulation are the techniques most used in this work. The fusion between electronic and oboe sounds seems to be the main characteristic of the work. The oboe often mixes with electronics and vice versa. This example, significant in many other ways as well, can demonstrate in what a fascinating and insightful way scientific and artistic work can complement each other and how it becomes possible to deeply penetrate sound matter to consciously open up new musical avenues (Motz, 1984). The work won the Honorable Mention in mixed category at the XI International Electroacoustic Music Competition (Bourges, 1983), and the First Prize at Stuttgart City Competition for young composers, 1983.

Marco Stroppa from 1983 to 1985 realized at CSC the cycle *Traiettoria* (*Traiettoria ... deviata, Dialoghi, Contrasti)* for piano and synthesized tape, employing the ICMS and the MusicV systems. He began composing *Traiettoria ... deviata* by working on piano resonances. Analyzing the resonance of the piano it emerged that by removing the attack transient of the sound what remained was a very little pianistic sound; in this no-man's-land it was possible to legitimately insert the intervention of the computer. *Traiettoria ... deviata* was the study of the two sound perspectives: that of piano sounds and that of synthesized sounds which were not limited to simulating acoustic sounds but interacted like a virtual orchestra. The piano was conceived as a tool that allowed the composer to explore the microscopic dimensions of sound (the disposition of partials, dynamic profiles, etc.) and produce sounds with broader spectral characteristics than piano sounds. Stroppa translated this idea by working on piano resonances, using the instrument (with trills, tremolos or pedaling) but also using sound synthesis (with a very detailed conception of the physical characteristics of sound) and composition techniques borrowed from electroacoustics (Tiffon and Sprenger-Ohana, 2011). For example, in the first part of *Traiettoria ... deviata*, the computer deals with creating an ambiguous substrate in which the chord can be transformed into a timbre and vice versa. The chord resonates on the piano, and the computer captures the resonance, in fading out, moving to an intermediate and ambiguous sound region, in which from a note goes back to a chord and from that to a timbre, elaborating and developing its density, until a cluster, which marks the moment of transition to the next phase (Tamburini, 1985). Additive synthesis and frequency modulation were used in the realization, producing hundreds of short sound events, which were then edited and mixed with ICMS system. These time-segment electroacoustic techniques, used as a model for composition, were of great importance to Stroppa (Tiffon and Sprenger-Ohana, 2011). The ICMS system was very useful to test the fusion of sounds in a quick and interactive way; to check if the result was perfect or to shift the sounds slightly by a few msec, until the blending was adjusted in an ideal way, and all this with an extremely musical reasoning. The micro-variations over time were not compositional but purely performative in nature (Durante and Zattra, 2002).

7. Conclusions

Since the time of the earliest experiments in computer music, many techniques have been developed for both reproducing and transforming natural sounds and for creating novel sonorities. Several models for sound synthesis were described, mostly from the user's point of view, outlining their main strengths and weaknesses.

Any number of techniques may be used to obtain a specific sound, even though some are more suitable than others. For musical use a versatile and efficient technique is not sufficient, but it is necessary for the musician to be able to specify the control parameters to obtain the desired result in an intuitive manner. It is then advisable for musicians to build their own conceptual models for a deep understanding of a technique, based on both theoretical considerations and practical experimentation. This process is necessary because a "raw" synthesis method does not stimulate either the composer or the performer. On the other hand, a solid metaphor for the sound-production mechanism can provide the composers with better stimulation and inspiration, and help performers improve their interpretive skills. An abstraction based on the model structure has been shown to be effective.

Traditional musical instruments are made of vibrating parts that produce sound and of parts that serve to control it, such as keys. The performer, while ignoring the mechanisms that create the sound, learns to obtain the desired sounds by acting on the control mechanisms. In dealing with virtual instruments, on top of the sound model and the corresponding algorithm, which constitutes the mechanism for sound production, there is a hardware and software interface for the control of the sound. This tends to specialize the generation mechanism by facilitating the possibility of obtaining a certain type of sounds. Musicians, by operating on this interface and basing themselves on the interpretation of the underlying mechanisms, form their own abstract idea (*conceptual model)* and execution procedure allowing them to obtain the desired sounds. In general, the same model may give rise to different interfaces and interpretations. The more these reflect the intrinsic properties of the production mechanism, the more useful they are for musicians in search of new sounds. Just as for traditional instruments the type of timbre is determined by the mechanism producing the sound (e.g. wind instruments sound the way they do because they are cylindrical or conical tubes and have a reed), similarly the timbre of virtual instruments is determined by the model and type of control available.

The computer music experience shows the joint research effort of scientists and engineers, together with the creative experimentation of musicians, can enrich the expressive possibilities of virtual instruments and give rise to new forms of art and human-computer communication. This scenario is the natural extension of the old tradition of cooperation and mutual intersection between science and music.

References

Adrien, J.-M. (1991). The missing link: Modal synthesis. In De Poli, G., Piccialli, A., and Roads, C., editors, *Representations of Musical Signals*, pages 269–298. MIT Press, Cambridge, MA, USA.

Ambrosini, C. (1982). Cadenza estesa e coda. In Vidolin, A. and Polo, N., editors, *Quaderno LIMB n. 2*, pages 82–92. La Biennale, Venezia.

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):757–768.

Avanzini, F. (2008). Simulation of vocal fold oscillation with a pseudo-one-mass physical model.

Speech Communication, 50:95–108.

Avanzini, F. and Canazza, S. (2015). Virtual analogue instruments: an approach to active preservation of the Studio di Fonologia Musicale. In Novati, M. M. and Dack, J., editors, *The Studio di Fonologia: A Musical Journey 1954–1983, Update 2008–2012*, pages 89–108. Ricordi Publications, Milan.

Avanzini, F. and Marogna, R. (2010). A modular physically based approach to the sound synthesis of membrane percussion instruments. *IEEE Transactions on Audio, Speech and Language Processing*, 18(4):891–902.

Avanzini, F., Marogna, R., and Bank, B. (2012). Efficient synthesis of tension modulation in strings and membranes based on energy estimation. *The Journal of the Acoustical Society of America*, 131(1):897–906.

Avanzini, F. and Rocchesso, D. (2002). Efficiency, accuracy, and stability issues in discrete-time simulations of single reed wind instruments. *The Journal of the Acoustical Society of America*, 111(5):2293–2301.

Avanzini, F. and van Walstijn, M. (2004). Modelling the mechanical response of the reed-mouthpiece-lip system of a clarinet. Part I. A one-dimensional distributed model. *Acta Acustica united with Acustica*, 90(3):537–547.

Balena, F. and De Poli, G. (1985). Un modello semplificato del clarinetto mediante oscillatore non lineare. In *Atti VI Colloquio di Informatica Musicale*, pages 111–138, Napoli.

Bank, B., Avanzini, F., Borin, G., De Poli, G., Fontana, F., and Rocchesso, D. (2003). Physically informed signal processing methods for piano sound synthesis: a research overview. *Eurasip Journal on Applied Signal Processing*, 2003:941–952.

Bari, A., Canazza, S., Poli, G. D., and Mian, G. A. (2001). Toward a methodology for the restoration of electroacoustic music. *Journal of New Music Research*, 30(4):351–363.

Bernardi, A., Bugna, G.-P., and De Poli, G. (1997). Musical signal analysis with chaos. In Roads, C., Pope, S. T., Piccialli, A., and De Poli, G., editors, *Musical Signal Processing*, pages 187–220. Swets & Zeitlinger, Lisse, The Netherlands.

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

Borin, G., De Poli, G., and Rocchesso, D. (2000). Elimination of delay-free loops in discrete-time models of nonlinear acoustic systems. *IEEE Transactions on Speech and Audio Processing*, 8(5):597–605.

Borin, G., De Poli, G., and Sarti, A. (1992a). Algorithms and structures for synthesis using physical models. *Computer Music Journal*, 19(4):30–42.

Borin, G., De Poli, G., and Sarti, A. (1992b). Sound synthesis by dynamic systems interaction. In Baggi, D., editor, *Readings in Computer-Generated Music*, pages 139–160. IEEE Computer Society Press, Los Alamitos, CA.

Borin, G., De Poli, G., and Sarti, A. (1997). Musical signal synthesis. In Roads, C., Pope, S., Piccialli, A., and De Poli, G., editors, *Musical Signal Processing*, pages 5–30. Swets & Zeitlinger, Lisse.

Cadoz, C. (2009). Supra-instrumental interactions and gestures. *Journal of New Music Research*, 38(3):215–230.

Cadoz, C., Luciani, L., 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.

Canazza, S., Avanzini, F., Novati, M., and Rodà, A. (2011). Active preservation of electrophone musical instruments. the case of the "Liettizzatore" of "Studio di Fonologia Musicale" (RAI, Milano). In Zanolla, S., Avanzini, F., Canazza, S., and de Götzen, A., editors, *Proc. of Sound and Music Computing Conference*, pages 304–309.

Canazza, S. and De Poli, G. (2020). Four Decades of Music Research, Creation, and Education at Padua's Centro di Sonologia Computazionale. *Computer Music Journal*, 43(4):58–80.

Canazza, S., De Poli, G., and Mian, G. A. (2010). Restoration of audio documents by means of Extended Kalman Filter. *IEEE Transaction on Audio, Speech, and Language Processing*, 18:1107–1115.

Canazza, S., De Poli, G., and Vidolin, A. (2022a). Gesture, music and computer: The Centro di Sonologia Computazionale at Padova University, a 50-year history. *Sensors*, 22(9):3465.

Canazza, S., De Poli, G., and Vidolin, A. (2022b). Tra musica e computer: il Centro di Sonologia Computazionale. *Padova e il suo territorio*, XXXVII(216):25–29.

Canazza, S. and Vidolin, A. (2001). Preserving electroacoustic music. *Journal of New Music Research*, 30(4).

Cavaliere, S. and Piccialli, A. (1997). Granular synthesis of musical signals. In Roads, C., Pope, S. T., Piccialli, A., and De Poli, G., editors, *Musical Signal Processing*, pages 187–220. Swets & Zeitlinger, Lisse, The Netherlands.

Chowning, J. (1973). The synthesis of complex audio spectra by means of frequency modulation. *Journal of the Audio Engineering Society*, 21(7):526–534.

Clementi, A. (1982). Parafrasi. In Vidolin, A. and Polo, N., editors, *Quaderno LIMB n. 2*, pages 78–79. La Biennale, Venezia.

Cosi, P., De Poli, G., and Lauzzana, G. (1994). Auditory modelling and self-organizing neural networks for timbre classification. *Journal of New Music Research*, 23(1):71–98.

Dashow, J. (1978). Three methods for the digital synthesis of chordal structure with non-harmonic partials. *Interface*, 7(2/3):69–94.

Dashow, J. (1980). Spectra as chords. *Computer Music Journal*, 41(1):43–52.

Dashow, J., De Poli, G., Tisato, G., and Vidolin, A. (1978). Computer music at University of Padova. In *Proceedings of the 1978 International Computer Music Conference (ICMC78)*, pages 486–493, Evanston, Illinois, USA.

De Poli, E. and De Poli, G. (1976). Identificazione dei parametri di un oscillatore VOSIM a partire da una descrizione spettrale. In *Atti III Colloquio di Informatica Musicale*, pages 161–177, Padova, Italy.

De Poli, E. and De Poli, G. (1979). Determinazione dei parametri VOSIM di un suono quasi periodico. In *Convegno Associazione Italiana di Acustica 79 (AIA 79)*, volume II, pages 41–43, Roma. Esa - Edizioni Scientifiche Associate.

De Poli, G. (1975). Esperimento di sintesi del canto per segmenti mediante modulazione del periodo fondamentale. In *Atti tavola rotonda sulla Sintesi della Parola*, pages 305–308. Centro di Fonetica del CNR.

De Poli, G. (1983). A tutorial on digital sound synthesis techniques. *Computer Music Journal*, 7(4):8–26. Reprinted in The Music Machine, C. Roads Ed., 429–47, MIT Press 1991.

De Poli, G. (1984a). Frequency dependent waveshaping. In *Proceedings of International Computer Music Conference (ICMC84)*, pages 91–101, Paris.

De Poli, G. (1984b). Sound synthesis by fractional waveshaping. *Journal of the Audio Engineering Society*, 32(11):849–861.

De Poli, G. and Faccio, A. (1989). Generalizzazione dell'oscillatore VOSIM per la sintesi del suono. In *Proceedings of the VIII Colloquio di Informatica Musicale*, pages 26–32, Cagliari, Italy. AIMI.

De Poli, G. and Piccialli, A. (1984). An efficient method of acoustical synthesis through phase distortion. *Alta Frequenza*, LIII(3):143–149.

De Poli, G. and Piccialli, A. (1988). Dynamic control of FIR filter for sound synthesis. In Lacoume, J., Chehikian, A., and Martin, N. Malbos, E., editors, *Signal Processing IV: Theories and Applications*, pages 559–562, (North-Holland). EURASIP, Elsevier Science Publishing Publishers B.V.

De Poli, G. and Piccialli, A. (1989). Waveform transformation techniques for granular synthesis. In Pravica, P., editor, *Proceedings of ICA89 International Congress on Acoustics*, volume 3, pages 113–116, Beograd.

De Poli, G. and Piccialli, A. (1991). Pitch-synchronous granular synthesis. In De Poli, G., Piccialli, A., and Roads, C., editors, *Representations of Musical Signals*, pages 187–219. MIT Press, Cambridge, MA, USA.

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

De Poli, G. and Prandoni, P. (1997). Sonological models for timbre characterization. *Journal of New Music Research*, 26(2):170–197.

De Poli, G. and Puppin, S. (1989). Un modello del clarinetto: canna e fori. In *Atti VIII Colloquio di Informatica Musicale*, pages 52–58, Cagliari, Italy.

De Poli, G. and Rocchesso, D. (1998). Physically based sound modelling. *Organized Sound*, 3(1):61–76.

De Poli, G. and Rocchesso, D. (2002). Computational models for musical sounds sources. In Assayag, G., Feichtinger, H., and Rodriguez, J., editors, *Mathematics and Music: A Diderot Mathematical Forum*, pages 257–288. Springer Verlag, Berlin.

Debiasi, G. B. (1959). Sulla riproduzione del suono dell'organo a canne con organi elettronici. *L'Elettrotecnica*, XLVI(11):754–765.

Debiasi, G. B., De Poli, G., Tisato, G., and Vidolin, A. (1984). Centro di Sonologia Computazionale C.S.C. University of Padua. In *Proceedings of the 1984 International Computer Music Conference (ICMC-84)*, pages 287–298, Paris, France.

Delmonte, R., Mian, G., and Tisato, G. (1984). A text-to-speech system for Italian. In *ICASSP '84. IEEE International Conference on Acoustics, Speech, and Signal Processing*, volume 9, pages 81–84.

Di Scipio, A. (1990). Composition by exploration of non-linear dynamic systems. In *Proceedings of the International Computer Music Conference (ICMC-90)*, pages 324–327, Glasgow, UK.

Di Scipio, A. and Tisato, G. (1993). Granular synthesis with the interactive computer music system. In *Proceedings of the X Colloquium on Musical Informatics*, pages 159–165, Milan. AIMI/LIM, University of Milan.

Doati, R. (1983). Regularity, direction, velocity. *Perspectives of New Music*, 22(1/2):61–86.

Durante, S. and Zattra, L., editors (2002). *Vent'anni di musica elettronica all'Università di Padova. Il Centro di Sonologia Computazionale*. Cleup, Padova.

Florens, J. L. and Cadoz, C. (1991). The physical model: Modeling and simulating the instrumental universe. In De Poli, G., Piccialli, A., and Roads, C., editors, *Representations of Musical Signals*, pages 227–268. MIT Press, Cambridge, MA.

Francini, G., Debiasi, G. B., and Spinabelli, R. (1968). Study of a system of minimal speech reproducing units for Italian speech. *Journal of Acoustical Society of America*, 43(6):1282–1286.

Gaver, W. W. (1993). What in the world do we hear? An ecological approach to auditory event perception. *Ecological Psychology*, 5(1):1–29.

Graziani, M. (1982). Esecuzione all'elaboratore di parafrasi. In Vidolin, A. and Polo, N., editors, *Quaderno LIMB n. 2*, pages 80–82. La Biennale, Venezia.

Kaegi, W. and Tempelaars, S. (1978). Vosim–a new sound synthesis system. *Journal of Audio Engineering Society*, 26(6):418–425.

Karpen, R. (1987). The vision (1985). *Perspectives of New Music*, 25(1/2):10.

Le Brun, M. (1979). Digital waveshaping synthesis. *Journal of The Audio Engineering Society*, 27:250–266.

Magalotti, R., Borin, G., and De Poli, G. (1995). Time domain simulation of the instruments of the flute family. In *Proceedings of ISMA 95 International Symposium on Musical Acoustics*, pages 132–137, Paris, France.

Mathews, M., Miller, J. E., Moore, F. R., Pierce, J. R., and Risset, J.-C. (1969). *The Technology of Computer Music*. MIT Press, Cambridge, MA.

Mathews, M. V. and Moore, F. R. (1970). Groove—a program to compose, store, and edit functions of time. *Communications of the ACM*, 13(12):715–721.

McIntyre, M. E., Schumacher, R. T., and Woodhouse, J. (1983). On the oscillations of musical instruments. *Journal of Acoustical Society of America*, 74(5):1325–1345.

Mian, G. A. and Tisato, G. (1986). Sound structuring techniques utilizing parameters derived from a voice analysis/synthesis system. In *Proceedings of the International Computer Music Conference (ICMC-84)*, pages 67–81, Paris.

Motz, W. (1984). Sotto pressione per due oboi e nastro quadrifonico sintetizzato all'elaboratore (1982). In Vidolin, A. and Polo, N., editors, *Quaderno LIMB n. 4*, pages 71–84. La Biennale, Venezia.

Pulkki, V., Lokki, V., and Rocchesso, D. (2011). Spatial effects. In Zölzer, U., editor, *DAFX: Digital Audio Effects*, pages 139–184. John Wiley & Sons, 2nd edition.

Razzi, F. (1981). Progetto secondo (1980). In Vidolin, A. and Polo, N., editors, *Quaderno LIMB n. 1*, pages 82–88. La Biennale, Venezia.

Risset, J.-C. (1969). An Introductory Catalog of Computer-Synthesized Sounds. Technical report, Bell Laboratories, Murray Hill, N.J.

Risset, J.-C. and Mathews, M. V. (1969). Analysis of musical-instrument tones. *Physics Today*, 22:23–30.

Roads, C. (1991). Asynchronous granular synthesis. In De Poli, G., Piccialli, A., and Roads, C., editors, *Representations of Musical Signals*, pages 143–186. MIT Press, Cambridge, MA.

Rocchesso, D. (2002). Spatial effects. In Zölzer, U., editor, *DAFX: Digital Audio Effects*, chapter 6, pages 137–200. John Wiley & Sons, Ltd.

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

Rodet, X. (1984). Time-domain formant-wave-function synthesis. *Computer Music Journal*, 8(3):9–14.

Sachs, C. (1940). *The history of musical instruments*. W. W. Norton & Co, New York, 1st edition.

Sarti, A. and De Poli, G. (1999). Toward nonlinear wave digital filters. *IEEE Transaction on Signal Processing*, 47(6):1654–1668.

Schaeffer, P. (1966). *Traité des Objets Musicaux*. Éditions du Seuil, Paris, France.

Serra, X. and Smith, J. O. (1990). Spectral modeling synthesis: A sound analysis/synthesis system based on a deterministic plus stochastic decomposition. *Computer Music Journal*, 14(4):12–24.

Spagnol, S., Geronazzo, M., and Avanzini, F. (2013). On the relation between pinna reflection patterns and head-related transfer function features. *IEEE Transactions on Audio, Speech and Language Processing*, 21(3):508–520.

Stathopoulou, P., Kokkinakis, G., and Mian, G. A. (1980). Synthesis of Greek speech with elementary speech segments. In Lainiotis, D. G. and Tzannes, N. S., editors, *Advances in Communications*, volume I, pages 506–515, Dordrecht. Springer Netherlands.

Sundberg, J. and Verrillo, V. (1980). On the anatomy of the retard: A study of timing in music. *Journal of the Acoustical Society of America*, 68:772–779.

Tamburini, A. (1985). Ritratto di poetica musicale: «Traettoria», per pianoforte e computer di Marco Stroppa. In *Quaderno LIMB n. 2*, pages 69–76. La Biennale, Venezia.

Tiffon, V. and Sprenger-Ohana, N. (2011). The creative process in Traiettoria: An account of the genesis of Marco Stroppa's musical thought. *Contemporary Music Review*, 30(5):377–409.

Tisato, G. (1976). An interactive software system for real-time sound synthesis. In *Proceedings of the 1976 International Computer Music Conference (ICMC-76)*, pages 135–143, Boston, MA, USA.

Todd, N. P. (1995). The kinematics of musical expression. *Journal of the Acoustical Society of America*, 97:1940–1949.

Torresan, D. (1982). Interpretazione ed esecuzione all'elaboratore della composizione 'Cadenza estesa e coda'. In Vidolin, A. and Polo, N., editors, *Quaderno LIMB n. 2*, pages 82–92. La Biennale, Venezia.

van Walstijn, M. and Avanzini, F. (2007). Modelling the mechanical response of the reed-mouthpiece-lip system of a clarinet. Part II: A lumped model approximation. *Acta Acustica united with Acustica*, 93(3):435–446.

Verma, T. S. and Meng, T. H. Y. (2000). Extending spectral modeling synthesis with transient modeling synthesis. *Computer Music Journal*, 24(2):47–59.

Vidolin, A. (1991). I suoni di sintesi di Perseo e Andromeda. In Doati, R., editor, *Orestiadi di Gibellina 1991*. Ricordi, Milano.

Vidolin, A. (1997). Musical interpretation and signal processing. In Roads, C., Pope, S., Piccialli, A., and De Poli, G., editors, *Representations of Musical Signals*, pages 439–459. Routledge, Lisse, NL.

Wanderley, M. and Depalle, P. (2004). Gestural control of sound synthesis. *Proceedings of the IEEE*, 92(4):632–644.

Zattra, L. (2003a). *Science et technologie comme sources d'inspiration au CSC de Padoue et à l'IRCAM de Paris*. PhD thesis, Universit´e de Paris IV–Sorbonne and Trento University.

Zattra, L. (2003b). Teresa Rampazzi. Pioneer of Italian electronic music. In *Proceedings of the XIV Colloquium on Musical Informatics*, pages 11–16, Firenze, Italy.

Zattra, L. (2004). Searching for lost data: outlines of aesthesic–poietic analysis. *Organised Sound*, 9(1):35–46.

Zölzer, U., editor (2011). *DAFX: Digital Audio Effects*. John Wiley & Sons, Ltd, 2nd edition.