Current Perspectives
In the Digital Synthesis of Musical Sounds

Xavier Serra
Audiovisual Institute, Pompeu Fabra University
Rambla 31, 08002 Barcelona, Spain
URL: http://www.iua.upf.es
email: xserra@iua.upf.es

[published in Formats, vol. 1. http://www.iua.upf.es/formats/**]**

When a musical sound is generated electronically, it is important to have a good model with parameters allowing an intuitive manipulation of the sound. The musician-user must be able to develop a musical intuition that will allow experimenting with the synthesis technique used. In this article we present a general view of digital synthesis, concentrating on two lines of research which will surely allow us to shatter the limitations which exist in methods used till now and finally manage an answer to the promises which computer music made in the 60's.

Introduction

The generation of sounds has been the music field where digital technology has had the greatest impact. But the evolution of digital synthesis techniques has been slower than first expected, and it has not been till the last few years that solutions for the future have been found which can give a new impulse to this development (Smith, 1991).

In a famous article in Science magazine from 1963, Max Mathews (Mathews, 1963), gave a very optimistic view of the computer as a musical instrument. The author, a pioneer of computer music, said that generating sounds from numbers was a completely general way to synthesize sound because the bandwidth and dynamic range of hearing are bounded and that therefore any sound we can perceive may be generated in this way. The promise of computer music was that the computer is capable of generating any sound that could ever come from a loudspeaker. A few years later, Max Mathews himself in his book on the technology of computer music (Mathews, 1969) wrote

"The two fundamental problems in sound synthesis are (1) the vast amount of data needed to specify a pressure function - hence the necessity of a very fast program - and (2) the need for a simple and powerful language in which to describe a complex sequence of sounds".

Problem (1) has been largely resolved by technological development, as the speed of digital processors has increased exponentially in the last twenty years. Problem (2) still has no satisfactory solution, as it is impossible to describe sounds if it is necessary to define each and every one of the numbers that represent an acoustic wave. We must be able to describe sounds from less numbers or start from recorded sounds. Fortunately, most waves are not musically interesting, and many sounds that are physically different are perceptually equal. Therefore, it is not necessary to generate all possible waves and the aim is to find a reduced group of synthesis and control techniques that will allow us to explore the whole timbre space of musical interest.

The traditional approach to this problem has been to try to generate sounds by combining simple synthesis elements. But in the last few years it has become clear that it is difficult to generate complex sounds with musical interest in this way, and that to compete with acoustic instruments in terms of sound complexity and expressive control, we need a new focus in digital synthesis. The focus proposed here starts out from the study of the sound reality surrounding us and looks for ways of extending this reality to a new, virtual world of sound.

Tradition of Digital Synthesis

Techniques of digital synthesis inherited the knowledge developed for synthesis by analog means. The first digital synthesis system, Music V, (Mathews, 1969) developed by Max Mathews during the 60's, introduced the concept of the unit generator as a digital version of the modules of analog synthesizers. A unit generator accepts numerical control entries and generates a signal, which is also numerical, which can be used as an entry to another unit generator or it can be a sound. Examples of unit generators are: oscillators, filters, adders, multipliers, envelop generators, and random number generators. From the combination of these elements (Fig. 1), synthetic sounds can be created similar to those obtained with the voltage-controlled modules of analog synthesizers, but with a more precise control. Most systems developed since then have started out from these concepts. Synthesizers based on the MIDI protocol are one example.

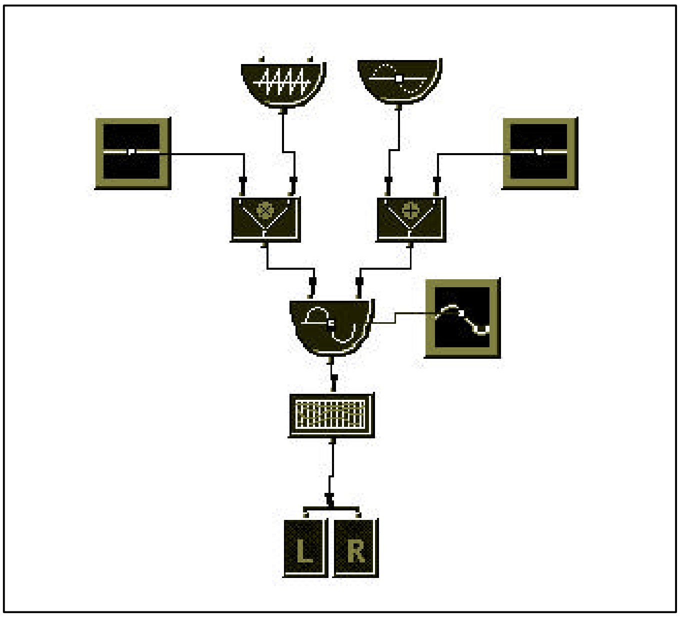

Figure 1: Synthesis algorithm by combining unit generators.

The implementation of synthesis algorithms has been done both at a software and a hardware level. The advantage of software implementations is its flexibility and unlimited complexity, as they are only bound by the programming language used; but for this same reason, they only work in real time with difficulty. On the other hand, hardware implementations work in real time but compete with difficulty in terms of flexibility and complexity. Given market characteristics, most commercial synthesizers use hardware implementations. However, thanks to the constant increase in computer processing speeds, the use of software based systems is more efficient every day and a large number of the better-known algorithms can currently work in real time.

Traditionally, digital synthesis techniques have been classified as: additive synthesis, subtractive synthesis, and non-linear synthesis. Additive synthesis is based on the sum of elementary sounds, each of which is generated by an oscillator. Subtractive synthesis is based on a complementary idea, that is to say filtering energy from a complex sound. Non-linear synthesis is a jumble in which a great number of techniques based on mathematical equations with non-linear behavior are included.

This classification is made from a very theoretical point of view which does not take us very far when trying to find new solutions with musical interest. Another, more useful taxonomy proposed by Julius Smith (Smith, 1991), organizes synthesis techniques as: processed recording, spectral models, physical models, and abstract algorithms.

One way of understanding Smith's classification is to speak of methods of synthesis as digital techniques which allow the achievement of a sonority continuum which goes from the reproduction of pre-existing sounds (recordings) to the generation of sounds from an absolute abstraction (imagined sounds), with all the intermediate steps. In this context, techniques based on processing recordings start out from the extreme of pre-existing sounds and try to create new, imagined, sounds, directly touching up real sound (as is the case of instruments known as samplers). At the other extreme of classification, there are abstract algorithms, which 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 communication (for example, with synthesizers based on frequency modulation techniques). Spectral and physical models are in the intermediate zone between these two extremes, and starting out from models or abstractions which describe pre-existing sounds and objects which generate sound, respectively, allow the exploration of most of the space between concrete sound realities and new, virtual, realities. We will go more deeply into these models further on.

Musical Objectives in Digital Synthesis

If we ask musicians working with technology what they would like to be able to do at the sound generation level, almost all would probably agree with the following answers: (1) to be able to create any imaginable sound and (2) to be able to manipulate any pre-existing sound in any conceivable way. These objectives are evidently utopia, not only because of the technological limitations, but also, and more importantly, because of the practical limits of our imagination. It is difficult to be able to imagine sounds without a reference to the world of sound surrounding us, and in fact our imagination always starts out from this referent. Thus, the great interest of techniques which have "sound reality" as a starting-off point.

Technological limitations make us value a series of compromises we must take into account when designing or using a specific synthesis technique. Specifically, we would like to mention four of these compromises:

> 1. 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.
>
> 2.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.
>
> 3.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.
>
> 4.Compute time. Compute time refers to the number of computer instructions needed to generate each of the sound samples synthesized. In this sense, frequency modulation synthesis is a very economical technique and additive synthesis requires much more compute time.

Ideally, we would want to maximize quality, flexibility, and generality and minimize compute time. This is not technologically possible and in each specific case we must evaluate what interests us most and choose the technique according to these considerations. For example, if we want maximum quality, we will surely have to renounce aspects of flexibility and generality, and if we want great flexibility we cannot have much quality. As we are looking more to the future and it is evident that the speed of digital processors will increase, we can allow ourselves the luxury, in our evaluation of synthesis techniques, of not worrying too much about compute time. Our priority will be maximizing quality, flexibility, and generality.

Besides evaluating this compromise when choosing a system, it is necessary to consider that for a synthesis technique to be useful to a musician, its control must be intuitive and must thus start out from an existing sound and/or musical reality. There are two sound realities that musicians are used to handling and which are good starting points for the creation of new sounds. One of these realities is that of physical objects that generate sound mechanically, for example, traditional instruments. 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 reality surrounding us.

The other sound reality interesting to musicians is that of perception, that is to say the sound that the listener perceives. From this reality the musician is able to describe sounds and imagine new ones. We can represent perceptual characteristics of a specific sound by spectral models and at the same time we can manipulate these characteristics to obtain new perceptual realities.

We now present some basic ideas behind the digital synthesis techniques based on physical and spectral models. These techniques might be the answer to many current problems of sound synthesis and allow a very favorable choice within the aforementioned compromises, especially at a quality, flexibility, and generality level. We must remember, however, that because of the fact of being fairly complex algorithms, we can only obtain low compute times with difficulty.

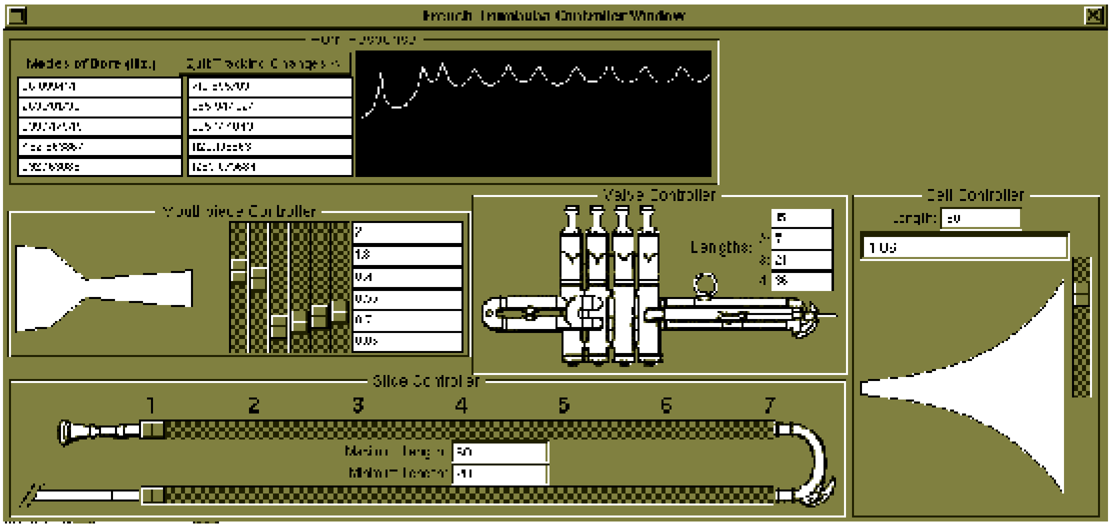

Figure 2: Graphic interface for a physical model of a musical instrument (by Perry Cook).

Synthesis with Physical Models

These methods of synthesis generate sounds describing the behavior of the elements which make up a musical instrument, such as strings, reeds, lips, tubes, membranes and resonant cavities. All these elements, mechanically stimulated, vibrate and produce disturbances, generally periodic, in the air which surrounds them. It is this disturbance which arrives to our hearing system and is perceived as sound.

Acousticians have been involved in finding models, mathematical relations, to describe these vibratory systems. These models have mainly been used to understand specific physical phenomena, but since Hiller and Ruiz's pioneering work (Hiller and Ruiz, 1971), they have also been used for synthesis. The first step in implementing these models consists in defining and measuring the physical characteristics of the object to be reproduced. For example, to generate the sound of a string, we must consider its length, thickness, density, etc. Once we introduce these physical measurements, the model allows us to reproduce the movement of the string numerically inside the computer and, at the same time, convert this movement into sound.

Generally, all instrumental models are made up of two kinds of elements: exciters and resonators (Borin et al., 1992). Exciters are the elements that cause, and sometimes maintain, the vibratory phenomenon, while resonators are where the vibrations with musical interest take place. In the case of a violin, the bow works as an exciter and the combination of the string with the wood cavity acts as resonator. In the computer, we can have models of the different elements, and by combining them and specifying their various physical characteristics, the user creates "musical instruments". With these means we can design and listen to the sound of a mechanical instrument that may be physically impossible to build. For example, coupling a double-reed model to create a sound stimulus, with a calf-skin membrane model to serve as resonator, we will generate sound. I doubt that anyone can construct an instrument such as this outside the computer.

The control of these methods of synthesis is done by means of the mechanical controls that an acoustic instrument of the same characteristics would have. Thus, we must give a model of a brass instrument, besides the stimulus characteristics produced by the lips, the different measurements of tubes and cavities that make up the instrument (Fig. 2).

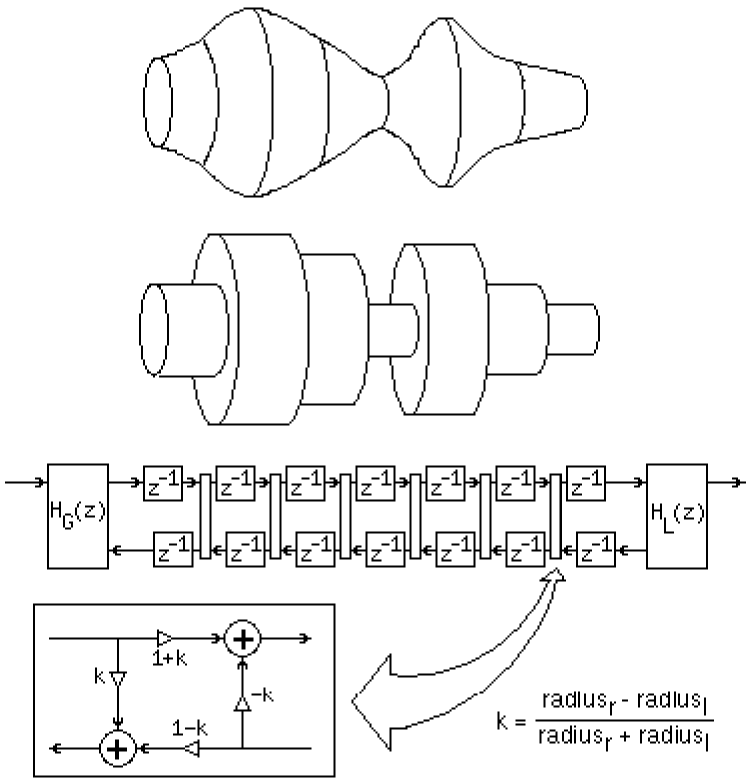

Figure 3: An acoustic tube (first drawing) can be sampled in space the same as sounds are sampled (second drawing). The sampled tube can be converted directly to a digital filter (third drawing).

Historically, physical models have been carried out by means of very complex algorithms that can hardly work in real time with current technology. These implementations have been based on numerical integration of the equation that describes wave propagation in a fluid (Fletcher and Rossing, 1991). Recently, more efficient solutions have been found for this problem (Smith, 1992) (Fig. 3) and systems have begun to appear with interest for musicians.

These physical models can be considered "reality generators", not only in the sense of imitating traditional instruments, but also for helping to conceptualize this reality and create structures which do not have a physical parallel. In this case the physical reality is used as a source of inspiration and not as a reference of the quality of the sound produced.

Synthesis with Spectral Models

Spectral models are based on the description of sound characteristics that the listener perceives. To obtain the sound of a string, instead of specifying the physical properties, we describe the timbre or spectral characteristics of the string sound. Then, sound generation is carried out from these perceptual data, thanks to diverse mathematical procedures developed in the last few years.

One advantage of these models is that techniques exist for analyzing sounds and obtaining the corresponding perceptual parameters. That is to say, by analyzing a specific sound we can extract its perceptual parameters. From the analysis, it is possible to synthesize the original sound again and the parameters can be modified in the process so that the resulting sound is new but maintains aspects of the sound analyzed.

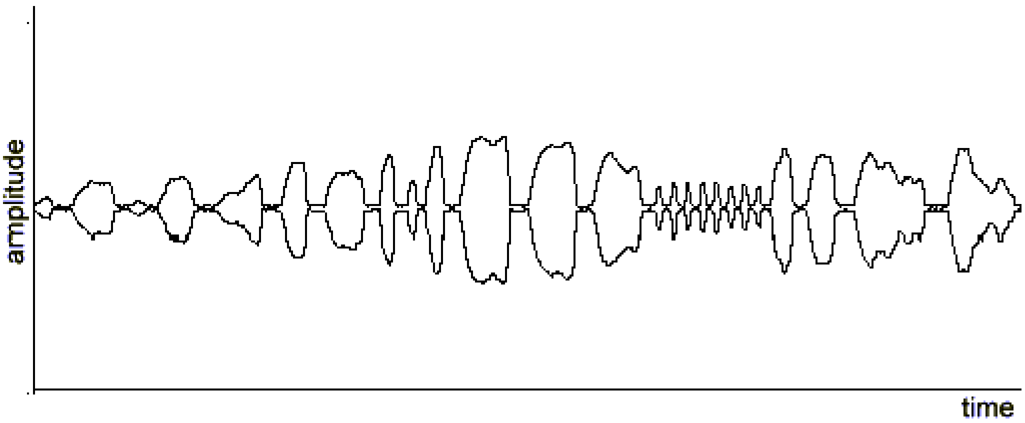

Figure 4: Temporal representation of the sound of a bird.

For a musician, synthesizing sounds with this kind of model is quite intuitive. For example, we can begin with the sound of a clarinet, previously analyzed, and modify it with instructions such as: make the sound more inharmonic, make the attack more percussive, mix in some timbre characteristics of a voice, make the timbre brighter, etc. The sound is converted into a plastic material that we can manipulate as we like.

The Fourier transform is the first step towards a perceptual modeling of sound. With this technique, a sound (Fig. 4) is decomposed into its harmonics or frequency components, whose evolution in time can be studied (Fig. 5). One step further in the Fourier transformation is decomposing the sounds into sinusoids (partial) and noise (residual component), that is to say, analyzing sounds with this model and generating new ones from analysis data (Serra, 1995; Serra and Smith, 1990). Analysis detects partials by studying spectral characteristics of a sound and represents them as sinusoids. These partials are subtracted from the original sound and the "residue" left over is represented as filtered white noise. The synthesis part of the system is a combination of additive synthesis for the sinusoidal part and subtractive synthesis for the noise part. This strategy of analysis and synthesis can be used both to generate sounds (synthesis) and to transform pre-existing sounds (processing). To synthesize sounds, we generally wish to model a whole family of timbres, for example, an instrument, and this can be achieved by analyzing isolated notes and transitions between notes played with an instrument, and building a data base which characterizes a whole instrument or any family of timbres, from which we synthesize new sounds. In the case of sound processing, the objective is the transformation of any sound, that is to say, not restricting ourselves to isolated notes nor requiring a previously built analytical data base.

Spectral models can be considered "transformers of reality". As with physical models, we can reproduce pre-existing realities, in this case a perceptual one, and modify it to obtain new sounds. Comparison between these two kinds of models (Table 1) shows a great complementarity between them.

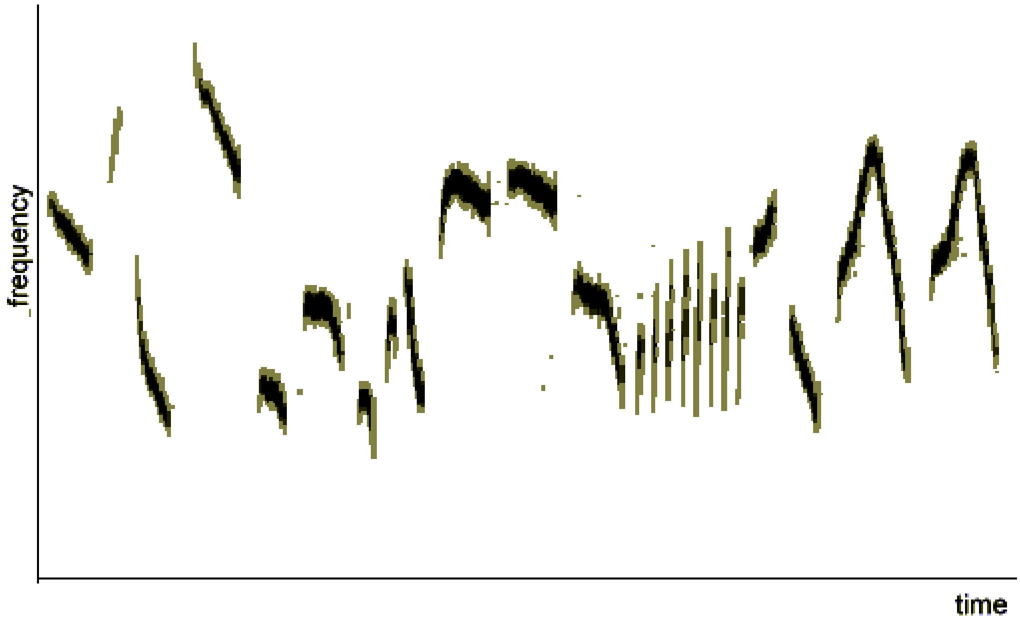

Figure 5: Spectral representation of the bird sound from figure 4.

Conclusions

The actual situation of digital synthesis is not yet at these levels; commercial synthesizers do not allow an approach to sound generation as has been presented here. But we are getting closer. Actually, not long ago the first commercial synthesizer based on physical models appeared (VL1 by Yamaha) and there are several computer systems, outside traditional commercial channels, which allow experimenting both with physical models and with spectral models (for example, software distributed by some research centers and universities). But even these programs are not yet at the level we would like. We hope research in this field will continue and that it will not be long before we can have these kinds of tools for musical creation. That the computer may finally come to musical maturity and keep the promises that were made in the 60's.

physical models	***spectral models***
based on physical reality	based on perceptual reality
controlled by physical parameters	controlled by perceptual parameters
specific models for each instrument	general models for all sounds
synthesis independent of analysis	synthesis starting from analysis
ideal for reproducing traditional instruments	less ideal for reproducing traditional instruments
less appropriate for reproducing natural sounds	ideal for reproducing natural sounds
high calculation time	high calculation time
allows choice between quality and flexibility	allows choice between quality and flexibility
closer to flexibility than to quality	closer to quality than to flexibility

Table 1: Comparison of characteristics of physical models to those of spectral models.

Bibliography

BORIN, G.; DE POLI, G.; SARTI, A. (1992) "Algorithms and Structures for Using Physical Models" in Computer Music Journal 16(4):30-42.

MATHEWS, M.V. (1963) "The Digital Computer as a Musical Instrument" in Science 142(11):553-557.

MATHEWS, M.V. et al. (1969) The Technology of Computer Music. Cambridge, Mass.: MIT Press.

FLETCHER, N.H.; Rossing, T.D. (1991) The Physics of Musical Instruments. New York. Springer-Verlag.

SMITH, J.O. (1991) "Viewpoints on the History of Digital Synthesis" in Proceedings of the 1991 International Computer Music Conference. San Francisco. International Computer Music Association.

SMITH, J.O. (1992) "Physical Modeling Using Digital Waveguides" in Computer Music Journal 16(4):74-87.

HILLER, L.; RUIZ, P. (1971) "Synthesizing Musical Sounds by Solving the Wave Equation for Vibrating Objects: Parts I and II" in Journal of the Audio Engineering Society 19(6):462-470 and 19(7):542-551.

SERRA, X. (1995) "Musical Sound Modeling with Sinusoids plus Noise", (manuscript).

SERRA, X. (1994) "Sound hybridization techniques based on a deterministic plus stochastic decomposition model" in Proceedings of the International Computer Music Conference, 1994.

SERRA, X.; Smith, J. (1990) "Spectral Modeling Synthesis: A Sound Analysis/Synthesis Based on a Deterministic plus Stochastic Decomposition" in Computer Music Journal 14(4):12-24.