State of the Art and Future Directions in Musical Sound Synthesis

Xavier Serra
Music Technology Group
Universitat Pompeu Fabra
Barcelona, Spain
xserra@iua.upf.edu

***Abstract*— Sound synthesis and processing has been the most active research topic in the field of Sound and Music Computing for more than 40 years. Quite a number of the early research results are now standard components of many audio and music devices and new technologies are continuously being developed and integrated into new products. Through the years there have been important changes. For example, most of the abstract algorithms that were the focus of work in the 70s and 80s are considered obsolete. Then the 1990s saw the emergence of computational approaches that aimed either at capturing the characteristics of a sound source, known as physical models, or at capturing the perceptual characteristics of the sound signal, generally referred to as spectral or signal models. More recent trends include the combination of physical and spectral models and the corpus-based concatenative methods. But the field faces major challenges that might revolutionize the standard paradigms and applications of sound synthesis. In this article we will first place the sound synthesis topic within its research context, then we will highlight some of the current trends, and finally we will attempt to identify some challenges for the future.**

Keywords—sound and music computing; sound synthesis.

I. Research Context

Musical Sound Synthesis is part of what is now called Sound and Music Computing (SMC). A good overview of this research field is the Roadmap funded by the EU and elaborated by the S2S2 consortium [1]. This Roadmap [2] covers quite a number of issues, but one of the main contributions is the definition of the actual field of research. In the Roadmap it is stated that the SMC research approaches the whole sound and music communication chain from a multidisciplinary point of view, and that by combining scientific, technological and artistic methodologies it aims at understanding, modeling and generating sound and music through computational approaches.

The sound and music communication chain covers all aspects of the relationship between sonic energy and meaningful information, both from sound to sense (as in musical content extraction or perception) and from sense to sound (as in music composition or sound synthesis). The disciplines involved in SMC cover both human and natural sciences. Its core academic subjects relate to musicology, physics (acoustics), engineering (including computer science, signal processing and electronics), psychology (including psychoacoustics, experimental psychology and neurosciences) and music composition. Most SMC research is quite applied and current areas of application include digital music instruments, music production, music information retrieval, digital music libraries, interactive multimedia systems, auditory interfaces and augmented action and perception (e.g. bionic ears, digital prosthesis and multimodal extensions of the human body).

Fig. 1 depicts the relationships between the different SMC research areas. It makes a basic distinction between research that focuses on sound (left hand side), research that focuses on music (right hand side) and the research topics that address the interaction between the two. For each research topic, there is an analytic and a synthetic approach. The analytic approach goes from encoded physical (sound) energy to meaning (sense), whereas the synthetic approach goes in the opposite direction, from meaning (sense) to encoded physical (sound) energy. Accordingly, analytic approaches to sound and music pertain to analysis and understanding, whereas synthetic approaches pertain to generation and processing. In between sound and music, there are multi faceted research topics that focus on interactional aspects. These are performance modeling and control, music interfaces, and sound interaction design.

All of the SMC research is very interdisciplinary and full of interdependencies. In this article we focus on sound synthesis but we will have to make connections with many of the topics depicted in Fig. 1.

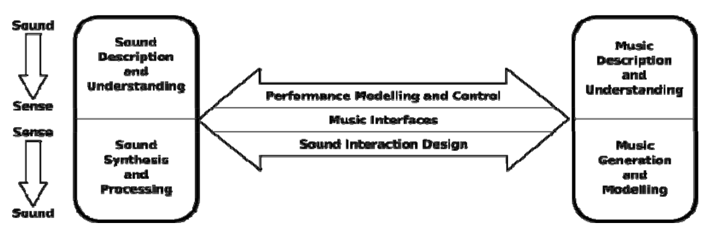


Fig. 1: Diagram of the SMC research areas (from [2]).

The topic of digital sound synthesis can be traced back to the work of Max Mathews in the 1950s at Bell Labs and the

Music V synthesis language that was developed by him in the late 1960s. Since then many synthesis techniques have been developed. For example in the 1970s and 1980s abstract algorithms, like FM and waveshaping, were developed, but they are currently considered obsolete. Then in 1990s computational approaches were developed that aimed either at capturing the characteristics of a sound source, known as physical models [3], or at capturing the perceptual characteristics of the sound signal, generally referred to as spectral or signal models [4].

With the physical modeling approach sounds are synthesized by describing the behavior of the elements that make up a music 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 that surrounds them. It is this disturbance that arrives to our hearing system and is perceived as sound. Historically, physical models have been carried out by algorithms based on numerical integration of the equation that describes wave propagation in a fluid and by some efficient solutions in order to have real-time implementations.

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, what is described is the timbre or spectral characteristics of the string sound. Then, sound generation is carried out from these perceptual data, thanks to diverse mathematical and computational procedures. 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 analyzed sound.

With the physical modeling approach we generate sounds describing the behavior of the elements that make up a music 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 that surrounds them. It is this disturbance that arrives to our hearing system and is perceived as sound. Historically, physical models have been carried out by algorithms based on numerical integration of the equation that describes wave propagation in a fluid and by some efficient solutions in order to have real-time implementations.

II. CURRENT TRENDS

A recent issue of the IEEE Signal Processing magazine [5] was dedicated to sound synthesis and included a good collection of articles representing the current trends in physical models, spectral models and what is being called corpus based concatenative methods. Here we will just highlight some basic ideas.

The current research approaches to physics-based modeling can be classified into five categories: mass–spring, modal, wave digital, finite difference, digital waveguide and source–filter models [3]. The digital waveguide method, together with its variations, is currently the most popular modeling approach, because it leads to computationally efficient algorithms that enable real-time synthesis of many simultaneous voices. Fig. 2 shows a particular example of such an approach. Some of the current efforts are aiming at modeling ethnic and historical musical instruments and also at modeling non-musical sound sources.

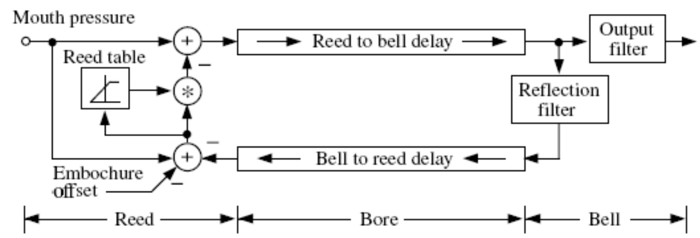


Fig. 2: Digital waveguide model of a single-reed woodwind instrument with cylindrical bore (from [3]).

Corpus based concatenative methods [5] take advantage of the rich and large databases increasingly available to assemble sounds by content-based selection and concatenation. Fig. 3 shows a general diagram of the concatenative methods. They make use of a variety of sound snippets in a database, source sounds, to assemble a desired output sound according to a target specification given by some descriptors or symbolic score, or by an example sound, audio score. With ever larger sound databases easily available, together with a pertinent description of their contents, these methods are increasingly used for instrument synthesis and other more abstract synthesis applications.

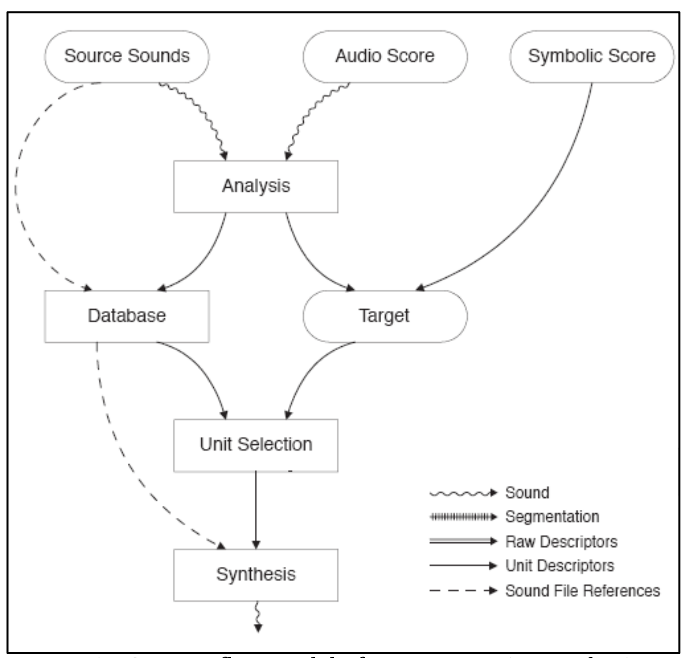
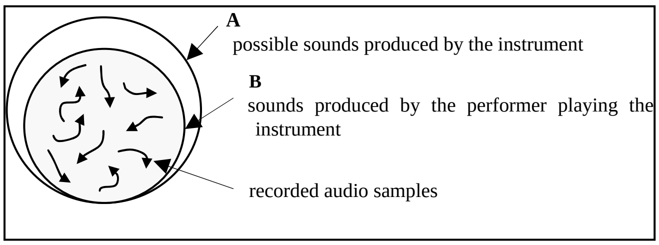
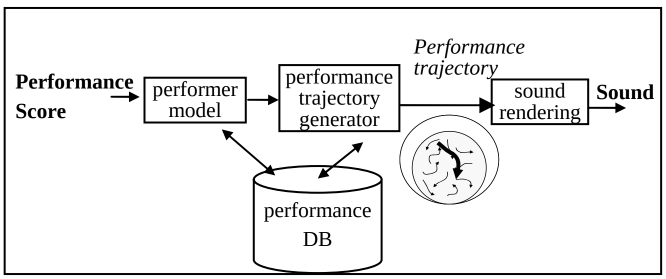


Fig. 3: Data flow model of a concatenative synthesis approach (from the article by Schwarz in [5]).

Generally, the spectral modeling approach [4] starts by sampling the sonic space (Fig. 4a) of a given instrument and analyzing the samples. These samples might be individual notes or short phrases, trajectories in the sonic space. Then the synthesis is based on transforming and concatenaing the samples from the information of the input controls, performance score (Fig. 4b). Thus we can basically consider the spectral modeling approach a type of concatenative synthesis. One of the key issues when spectral transformations have to be applied is the parameterization of the samples since it has to be the right one for the needed transformations. In the singing voice synthesis described by Bonada and Serra in [5] the parameterization is based both on perceptual and physical models of the voice, thus it is very much a hybrid modeling approach.



(a)



(b)

Fig. 4: General diagram of concatenative synthesis for instrument modeling. (a) Instrument sonic space, (b) Overall synthesis diagram (from the article by Bonada and Serra in [5]).

III. Open Issues

On a first approximation we could say that sound synthesis is a solved problem, that any sound can be obtained by digital means and therefore there is not much left to do. However on a second pass we can definitely find quite a number of unsolved open problems. Digital synthesizers are far from having the subtle nuances and expressivity of the acoustic instruments and there are many aspects to be improved in the current techniques. The SMC Roadmap [2] identifies a number of open issues related to sound synthesis of which we would like to emphasize some here.

A. Interaction centered sound modeling

It is widely believed that the main missing aspect in existing synthesis techniques is adequate controlled naturalness and expression which could be also understood as the lack of inclusion of musical gesture information within the synthesis step. The dynamic and expressive aspects of music and sound generation should be given greater weight in the design of sound synthesis techniques. Thus the open issue is how to analyze and model the expressive aspects of the musical gestures that control an instrument and how to use this information in the actual synthesis systems.

B. Physical modeling based on data analysis

To date, physical models of sound and voice have been appreciated for their desirable properties in terms of synthesis, control and expressiveness. However, it is also widely recognized that they lack the ability to fit with real observed data due to the high number of parameters involved, the fact that control parameters are not related to the produced sound signal in a trivial way and, in some cases, the radical non-linearities in the numerical schemes. All these issues make the parametric identification of physics-based models a formidable problem. Future research in voice and sound physical modeling should thus take into account the importance of models fitting real data, both in terms of system structure design and also of parametric identification. Co-design of numerical structures and identification procedures may also be a possible path to complexity reduction.

C. Integration of control with sound generation

The traditional separation between gesture controllers and sound generators has some significant negative consequences, the most obvious being the reduction of the “feel” associated with producing a certain kind of sound. Another frequent criticism is the inherent limitations of the current protocols that connect these two components of the instrument chain. Currently it is impossible to design highly sophisticated control interfaces without a profound prior knowledge of how the sound or music generators will work. Thus the open issue is to how to design controllers that are aware of the sound they influence or how to design sound generators that are aware of their input controllers.

D. Feedback systems

When musicians play instruments, they perform certain actions with the expectation of achieving a certain result - a musical performance. As they play, they monitor the behavior of their instrument and, if the sound is not quite what they expect, they adjust their actions to change it. In fact, in most instruments, the sound produced has a direct mechanical feedback that is felt by the player and that is important in the monitoring process. Thus a relevant improvement of the digital instruments would be if they could provide appropriate feedback to the player, not just the auditory one. Current research attempts to incorporate into the musical instruments both haptic and visual feedback in addition to the auditory one.

IV. Beyond Sound Synthesis

In the previous section we have identified some of the open issues in digital sound synthesis; however they are still based on the traditional paradigm of a musical instrument. We should be able to go beyond that.

The first of the five big SMC challenges identified in [2] is to design better sound objects and environments. The idea is that by improving the sounds produced by the objects present in our environment we will enhance the affective and emotional character of these objects and consequently our quality of life. This challenge goes beyond the open issues just mentioned and also beyond the traditional signal processing centered approach in sound synthesis, thus beyond the concept of a musical instrument.

Most of the electronic artifacts and devices which surround us often come with speakers and with artificially designed sounds that are poorly suited to their function and aesthetics (think of mobile phones, for instance). Due to this widespread availability of electronic sound, we live in "schizophonic" environments, where sound is separated from its source. A sound is independent from the mechanical object that could have produced it. In addition, public and personal environments tend to be cluttered with unwanted sound and music.

This entire new sonic world generated and controlled by electronic devices offers many new challenges which at this stage are difficult to pinpoint. It is clear that the challenges are not just of an engineering nature, but at the same time the engineers will have to provide ideas and technical solutions to these challenges. The ideas and methodologies that will solve the problems and will improve our quality of life will come from many disciplines. Thus interdisciplinary research and collaboration between disciplines will be very important for the advancement of the field.

V. CONCLUSIONS

Despite its widespread use in today's electronic devices, not just in electronic music instruments, sound synthesis is quite an unknown engineering topic. Its research context has been a special one for many years and the engineers working on it have had to deal with many issues that are not typically considered engineering subjects. The topic has grown and developed to a quite mature situation. The Sound and Music Computing field has given a good framework to this research, offering constant challenges to overcome and proposing many applications to be developed. But now the field is changing and new paradigms arise that are revolutionizing the technological approaches used and the contexts in which they are applied. The opportunities for having an impact in the new Information Society are big but the problems to be solved are so many that decisions have to be taken about where to concentrate the research efforts and how to approach the problems to be solved. In this article I made an attempt to identify the current trends and challenges related to sound synthesis with the goal to raise the awareness for it within a broader engineering research community.

ACKNOWLEDGMENTS

A good part of this article has been taken from [2], and thus I would like to acknowledge the contributions from the rest of the $S2S^2$ consortium and the funding of the EC for this project.

REFERENCES

[1] $S2S^2$ project. http://www.s2s2.org/

[2] X. Serra; M. Leman, and G. Widmer, Eds., *A Roadmap for Sound and Music Computing* [Online]. 2007 ed. The S2S Consortium. Available: http://www.soundandmusiccomputing.org/roadmap/.

[3] V. Välimäki et al., "Discrete-time modeling of musical instruments." *Reports in Progress in Physics*, Vol. 69, pp. 1-78, 2006.

[4] J. O. Smith, *Spectral Audio Signal Processing* [Online]. March 2007 draft, Available: http://ccrma.stanford.edu/~jos/sasp/..

[5] V. Välimäki et al., Eds., "Signal processing for sound synthesis." Special section. *IEEE Signal Processing Magazine*, Vol. 24, No. 2, 2007.