ESCHER - Modeling and Performing Composed Instruments in real-time

Marcelo M. Wanderley
Analysis-Synthesis Team
wanderle@ircam.fr

Norbert Schnell, Joseph Rovan
Real-Time Systems Group
schnell@ircam.fr, rovan@ircam.fr

Ircam
1, Place Igor Stravinsky
Paris, 75004 France

ABSTRACT

This article presents ESCHER, a sound synthesis environment based on Ircam's real-time audio environment *jMax*. ESCHER is a modular system providing synthesis-independent prototyping of gesturally-controlled instruments by means of parameter interpolation. The system divides into two components: gestural controller and synthesis engine. Mapping between components takes place on two independent levels, coupled by an intermediate abstract parameter layer. This separation allows a flexible choice of controllers and/or sound synthesis methods.

1. INTRODUCTION

High-quality sound synthesis methods are currently widely available, reflecting massive research over the last several decades. Today, however, the emphasis is shifting onto expressive control of these methods. Traditionally, real-time control of sound synthesis has been performed by instrument-like gestural controllers such as piano keyboards, wind controllers, guitars, etc., all usually based on the MIDI standard.

Nevertheless, gestural controllers may or may not follow the design guidelines of acoustic instruments, leading to two basic gestural controller families: *instrument-like controllers*, modeled after acoustic instruments and allowing the application of a previously learned gestural vocabulary; and *alternate controllers*, which allow the use of non-traditional gestural vocabularies. Instrument controllers have been designed in order to extend the performance possibilities of acoustic instrumentalists by applying their expert gestural control techniques to synthesis. Alternate controllers, on the other hand, provide a means of escaping from the idiosyncrasies of acoustic instrument playing techniques. They may therefore allow the use of any gesture or posture, depending on the sensing technologies employed. Thus, for the same synthesis result, one type of controller may be better suited than another, depending on the performer's background and specific musical intentions. [1]

Furthermore, the facile control of sound synthesis depends heavily on how the controller outputs are related to the available synthesis parameters. Since the outputs of the controller are not necessarily analogous to the inputs of the synthesis engine, an intermediate "mapping" stage is required, where controller variables are related to available synthesis variables [4] [5] [6]. Such mapping formalisms need to be devised in order to simplify the simultaneous control of many variables; an example would be the control of additive synthesis, where one needs to control the real-time evolution of hundreds of partials.

An ideal system designed for real-time sound synthesis should therefore be able to provide an environment where a performer may experiment with different gestural controllers and select the appropriate mapping strategies for the chosen controller and specific sound synthesis method. Our approach to this question is a system called ESCHER, described in detail in the following sections.

2. SYSTEM OVERVIEW

The ESCHER system is an audio environment designed to provide an intuitive control of sound synthesis in real-time. ESCHER's modular design permits easy adaptation to a wide range of different gestural controllers and/or different sound synthesis methods.

---

[1] For more information on gestural controllers, check [1], [2] and [3] or the address:
http://www.ircam.fr/equipes/analysis-synthesis/wanderle/Gestes/Externe/

Composed Instruments

In ESCHER the notion of a generic "composed instrument" is used to describe an instrument where the gestural controller is independent from the sound synthesis model, both related by intermediate mapping strategies. "Composed instruments" typically use two layers of parameter mapping on top of a more-or-less arbitrary synthesis engine to match various controller devices played by the performer to the sound synthesis result.

An abstract gestural model is used, which differentiates between continuous features of sound development and transitions between them. This model derives from acoustic instruments, where continuous modulations of the sound by the player (pitch variations, dynamics, etc.) alternate with abrupt changes (note attacks, bow changes, etc.), which cause transients.

The modular concept of composed instruments proves useful in choosing the interaction level desired by the user. Depending on the user's technical skills or specific musical aims, one can arrange the mapping in order to allow different degrees of control complexity. Being strongly dependent on the controller one uses (number of available output parameters, nature of the available parameters - continuous/discrete, range, etc.), the mapping can be configured in either a *micro mode*, where the performer has access to each individual parameter in detail, or in a *macro mode*, where some parameters may be kept constant to default values, allowing the user to concentrate on higher-level sound features. Examples of composed instruments have been implemented in ESCHER using both additive and granular synthesis.

Control

A controller is connected to the ESCHER system through an initial mapping layer. This layer can consist of simple one-to-one relationships, instrument-like mappings, where one considers parameter dependencies specific to an instrument model, or any other metaphor for simplifying the manipulation of simultaneous parameters, such as the one presented in [6]. In general, this mapping involves the transformation and conjunction of the raw controller data via functions (typically by table lookups), conditions and linear filters.

This first mapping represents an adapter between a particular controller's outputs and a set of abstract parameters defined by the user during the composed instrument definition. These abstract parameters can be understood as the interface to the composed instrument synthesis module. While continuous values control the development of sound features according to continuous controller input, events trigger transition processes and abrupt changes. Events are often accompanied as well by abrupt changes of the continuous parameters and can thus have an additional parameterization.

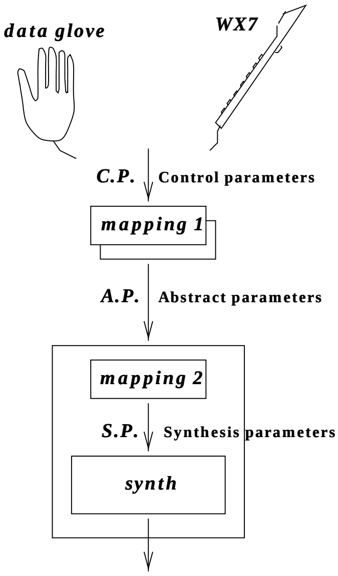

Figure 1: ESCHER system overview

In the context of a simple musical instrument model, continuous parameters could be pitch, dynamics, inharmonicity, density, formant specifications, etc. The beginning and end of a note would be discrete events, each having individual attributes in order to differentiate various transition types such as legato or staccato.

Parameter Spaces and Road-Maps

Timbre evolution in ESCHER is accomplished by navigating through different spaces by means of continuous movements and transitions. The second mapping layer thus consists of a linear interpolation of parameter sets determining the synthesized sound for continuous movements. A set of D continuous parameters forms a D-dimensional space, where the synthesis parameter sets are associated to the nodes of a grid each representing a particular quasi stationary synthesized sound. When the vector of values of the continuous abstract parameters point to a position in the D-dimensional parameter space, the actual synthesized sound is determined via linear interpolation of the synthesis parameters associated to the 2^D nearest nodes of the grid.

While continuous parameters of the composed instrument control the continuous movement within one parameter space in this manner, discrete events are handled by a "road-map" that governs navigation between multiple spaces, possibly triggering appropriate transitions. Thus a complete definition of the composed instrument for a particular synthesis engine consists of the abstract parameters, the distribution of the synthesis parameter sets to multiple parameter spaces and the road-map.

Sound Synthesis

Various sound synthesis methods can be used for composed instruments. Whatever the method, however, the synthesis should be defined in such a way that the continuous linear interpolation of the synthesis parameters causes a perceptually continuous development of sound features. Furthermore, the synthesis engine must implement the transitions for all possible paths defined in a road-map.

Because of their intuitive relationship between synthesis parameters and produced sound, synthesis methods using prerecorded and pre-analysed sound segments--like additive re-synthesis and granular synthesis[7]--are particularly interesting for building composed instruments. Additive synthesis is well-suited because of its attributes concerning time/frequency scaling as well as transformations and modulations in the frequency domain. In the context of a composed instrument controlled by a performer, additive re-synthesis can be used to sculpt prerecorded sound material in response to performer gesture.

For additive synthesis the synthesis parameters are the sets of partials (frequencies and amplitudes) corresponding to a certain sound considered as quasi-stationary. The partial sets are obtained from a previous analysis of recorded sound segments using Ircam's program "Additive."

As we already mentioned, ESCHER implements the concept of "timbral spaces" [8] where additive sound models are distributed in a virtual orthogonal D-dimensional space, each axis corresponding to an abstract parameter. The mapping is thus accomplished via the synthesis parameter values hard-coded in each additive model distributed in the space.

Examples of a similar systems using the concept of timbral spaces are the ISEE [9], where four predefined parameters define a timbral space in FM synthesis, and the system developed by L. Haken and his colleges at the University of Illinois [10]. The latter is based on the Kyma environment and uses traditional additive analysis/synthesis extended by the use of bandwidth information for each partial.

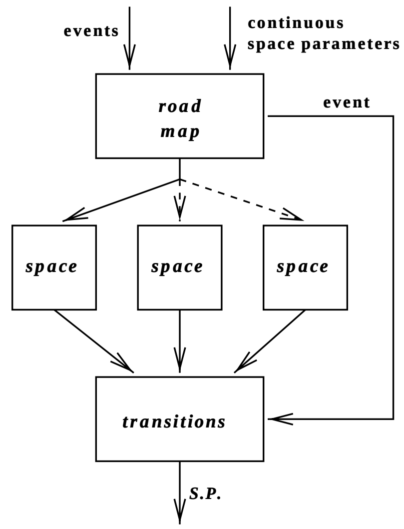

Figure 2: Events and Continuous Parameter Management

Time Warping

Since the parameterized segments after the analysis stage still carry unextracted features of the sound image -- including temporal modulations -- it is interesting to maintain the synthesis parameters as segments with an embodied temporal development when they are distributed in the parameter spaces.

In this case the processing of the interpolation of the synthesis parameter segments must already take into account the temporal treatment -- "time-warping" -- of the segments. Thus at least one more value will be added to the continuous parameters entering the composed instrument to control the time-warping of the segments. This might be, for example, the duration of a loop in which the segments in a particular space are read before the interpolation, or a parameter giving a continuous offset to the beginning of the segment to control "scratching" through the segments under control of the performer.

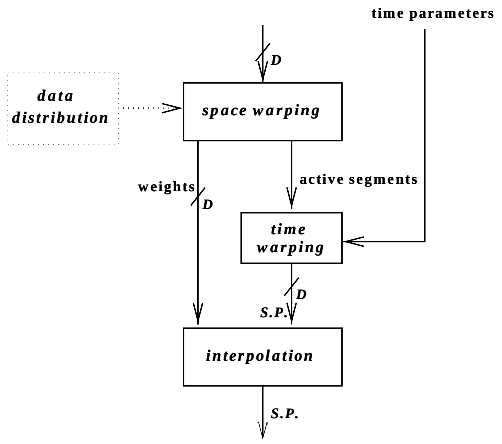

Figure 3: Space/Time Warping

3. APPLICATIONS

Wind Instrument Modeling
One application based on the ESCHER composed instrument model, an additive re-synthesis model of a clarinet, was built using a YAMAHA WX-7 wind controller. The controller device provided MIDI controller values for breath pressure, fingering and lip pressure, which were connected via various mappings to loudness, dynamic spectrum and pitch. Events determined the beginning of a note parameterized by the strength of its attack and whether it or not it was slurred or re-articulated.

Two different parameter spaces were used for the attack and sustain parts of the clarinet. Twenty-one appropriate pre-analysed segments--three different dynamics for each of seven different pitches-- were placed on a seven-by-three grid to cover the two dimensional parameter spaces. While the segments in the attack space were read straight through with a duration scaled inversely proportional to the strength of the note attack, the segments associated with the space of the sustain part were read in loops in order to keep the embodied modulations of the analyzed material.

Different transitions were defined between a sustaining note and a new attack (legato note) as well as for newly attacked notes and to smoothly join the attack to the sustain part (transition from attack space to sustain space). The road-map in this case basically consists of the mutual switching and initial parameterization of the two parameter spaces driven by the incoming note events. In addition, the sustain segments are automatically joined after the attack and each switch is associated to the triggering of an appropriate transition.

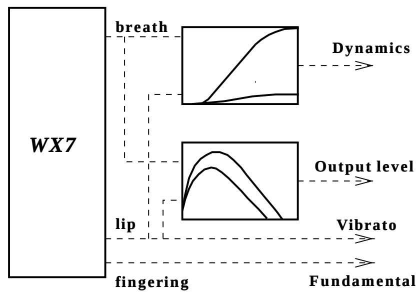

Figure 4: Example of First Mapping Layer for a Composed Clarinet [5]

Additional instrument modeling has been developed to simulate amplitude variations in each partial of a clarinet spectrum due to variations in distance between the instrument and a fixed microphone. A module in ESCHER simulated the effect of the first reflection (from the floor, in this case) in the sound such as that which occurs when an instrument moves during performance. This module has been added to the basic ESCHER environment and was controlled by a position sensor placed on the controller. The "virtual" microphone position could be chosen in the model, one or two meters away from the clarinet.

Granular Speech Processing
A second application implementing the paradigm of composed instruments using ESCHER is a granular synthesizer processing segments of speech recordings. The granular synthesizer engine is controlled by four parameters -- grain period, pitch, duration, output channel -- plus a statistical variation for each one, making a total of eight parameters for the grain being played.

The interface for the instrument are the continuous abstract parameters *mass* and *density*, which are mapped to the eight synthesis parameters via parameter spaces. Note-like events trigger a syllable. Each syllable is treated like a musical note, with attack, sustain and release sections.

Finally, a parameter space consists of a speech segment and an individual mapping for the interpolation, governed by the abstract parameters.

4. CONCLUSIONS AND FUTURE WORK

In this paper we presented ESCHER, a modular sound synthesis environment designed for an intuitive control of different sound synthesis methods by various gestural controllers. It has been designed in order to allow the control of user-defined composed instruments, where the user is able to select the synthesis method, the controller device and the level of control he desires. The system has been devised with the goal of providing a flexible tool for human-computer interaction in a real-time musical context. In its current state, ESCHER runs natively on SGI workstations, and a port is being prepared for Linux PCs.

Future additions to the system will include the integration of analysis and re-synthesis of noise components in additive synthesis, the use of spectral envelopes for partial transformation and manipulation, and different synthesis methods, such as PSOLA. [11]

5. ACKNOWLEDGMENTS

The authors would like to thank Shlomo Dubnov, Xavier Rodet and Diemo Schwarz for their comments and suggestions.

6. REFERENCES

[1] J. Paradiso, "New Ways to Play: Electronic Music Interfaces", *IEEE Spectrum*, Dec. 1997.

[2] A. Mulder, "Getting a Grip on Alternate Controllers: Addressing the Variability of Gestural Expression in Musical Instrument Design," *Leonardo Music Journal*, vol. 6, 1996, pp. 33-40.

[3] C. Roads, *Computer Music Tutorial*, chapter 14, MIT Press, 1996, pp. 617-658.

[4] T. Winkler, "Making Motion Musical: Gestural Mapping Strategies for Interactive Computer Music," *Proceedings of the International Conference on Computer Music*, ICMC'95, pp. 261-264.

[5] J. Rovan, M. M. Wanderley, S. Dubnov, and P. Depalle, "Instrumental Gesture Mapping Strategies as Expressivity Determinants in Computer Music Performance," *Proceedings of KANSEI - The Technology of Emotion Workshop*, 1997, pp. 68-73.

[6] A. Mulder, S. Fels, and K. Mase, "Empty-Handed Gesture Analysis in MAX/FTS," *Proceedings of KANSEI - The Technology of Emotion Workshop*, 1997, pp. 87-91.

[7] N. Schnell, *GRAINY - Granularsynthese in Echtzeit*, Beiträge zur Elektronischen Musik, No. 4, Institut für Elektronische Musik, Graz, 1995.

[8] D. Wessel, "Timbre Space as a Musical Control Structure," in C. Roads and J. Strawn, *Foundations of Computer Music*, MIT Press, 1987, pp. 640-657.

[9] R. Vertegaal and E. Bonis, "ISEE: An Intuitive Sound Editing Environment," *Computer Music Journal*, Vol. 18, No. 2, Summer/1994, pp.21-29.

[10] L. Haken, E. Tellman, and P. Wolfe, "An Indiscrete Music Keyboard, " *Computer Music Journal*, Vol. 22, No. 1, Spring/1998, pp.30-48.

[11] G. Peeters, "Analyse et Synthèse des Sons Musicaux par la Méthode PSOLA", *Proceedings of the Journées d'Informatique Musicale*, 1998, pp. G4-1 - G4-6.