Models and Deformations in Procedural Synchronous Sound for Animation

Robin Bargar (rbargar@ncsa.uiuc.edu), Alex Betts, Kelly Fitz, Insoo Choi
NCSA, University of Illinois at Urbana-Champaign, 405 N. Mathews, Urbana, IL 61801

Abstract

Numerical simulations used in computer animation provide computational environments for algorithmic sound production. We discuss the parallel production of images and sounds from a common model, and the data generated in such a model. Two cases are presented, one based upon classical differential equations, the other based upon parameterized gestures.

1. Introduction

Advanced computer graphics are moving away from manual modeling and animation, where each object and each frame are created by hand, towards automated spatial and temporal image production. Spatial: modeling objects based upon photographic data, and temporal: animated movements of objects based upon simulations of dynamical systems. These practices have enabled more efficient production of complex scenes, and they foreground the algorithmic relevance of aesthetic results from numerical simulations. The emergence of tools for data-driven and procedural animation provides a platform for sound synthesis production in those systems. Automated production of synchronous sound has potential applications in computer graphics-based production. At the same time, parallel image and sound rendering systems indicate new possibilities for the aesthetic integration of images and sounds.

The 20th century marks the large-scale pursuit of mechanical and electronic synchronization of image and sound. Some experimental approaches provide alternatives to the dominant realism in film and video production. These alternatives may be thought of as algorithms for situating a recording apparatus in an environment, and staging the artist's choices as a function of the transformations generated by the apparatus in the resulting representation. These experiments remind us that 'realism' is a fictional style, far removed from the documentary material of the real. In present-day animation production we see the return of mechanized algorithms, often applied towards achieving an automated cinematic realism. The rise of an "algorithmic realism" provides a new site for musical composition using automation. It also provides a site for aesthetic inquiries concerning representation and realism in music.

Case Studies

There are numerous compositional approaches to the integration of sounds and images. Results reported by Florens (1998), Cook (1995a) and Takala (1993) are related to the present work. We introduce experimental data-driven methods into a mature animation production environment, creating a series of compositions in computer animation where sounds are generated in correspondence to the data and dynamics used to control the animation. The correspondences can be tuned to be audible in the computed sounds. The dynamics of the sounds are parameterized in synthesis algorithms, and these algorithms are controlled by the application of techniques for the visual display of data. Where a graphical display is dynamically computed to visualize data or a numerical model, sound can be computed to respond to the same dynamics, to achieve a coupling of sound and image that is informative about the states of the underlying simulation.

In this paper we discuss results from two approaches to automated synchronization: (1) sound synthesis controlled by a physically-based particle system, and (2) deformation of sounds using parameterized gestures for deformation of polygonal objects. In the first approach a particle system generates detailed graphical motion that would be impractical to draw by hand. Collisions and accelerations are applied to the control of the animation, and these same events are applied to create sounds, varying resonance and spectral characteristics according to the forces and materials in each collision. In the second approach deformation of 3D graphical objects is coupled to the deformation of sounds using a bandlimited spectral peaks method of analysis and re-synthesis, and a recently-reported method for sound morphing. The deformation gestures are parameterized in animation control channels that provide time-varying functions applied to transformations of the 3D model. These functions and their channel organization for animation provide a hierarchical interface for the control of sound synthesis.

Data-Driven Synchronization

Data-driven synchronization of sound and image is based upon the instruction sets that generate graphics, rather than relying upon the graphics themselves to provide the control data. Figure 1 shows how this approach differs from film and video production. Classical post production relies upon the linear recording medium as a common denominator, with the frame count as the only data shared between image and sound. This approach requires manual interpretation to fit sounds to images. It is a labor-intensive process where the sounds rarely have a material relationship to the images. The realism fabricated between image and sound can be considered an idealist mode of production. We explore a materialist mode where a composer’s vocabulary articulates a position with respect to an automaton. By not assuming the mechanics of cinematic reproduction we shift away from the 19th and 20th century dependence upon the linearity of the storage medium as an organizing principle for composition, towards a non-linearity of alternatives generated during interactive presentation. This represents a relocation of perceptual models from the recording apparatus to the computational apparatus. The relocation suggests a mode of production where the composition is articulated as a relationship between perceptual models and analyzed and re-synthesized media.
Figure 1 and results were first reported in 1993 as part of CAVE research (Bargr 1993), an interactive real-time display. To apply this practice to traditional animation we used a hybrid approach, converting the results of the Figure 1B architecture into a linear movie format of Figure 1A.

2. Procedural Sound from Equations of Motion

The short animation Music for Unprepared Piano (Bargr 1998) uses particle system dynamics to determine graphics and sound events. The graphics are generated by a visualization system for rendering particles in a scene made of rigid-body polygonal objects. The sounds are generated in VSS (Bargr 1994) using physically-based models from the STK toolkit (Cook 1995b) to simulate blown, struck and plucked instruments. Techniques for spatial and temporal composition are applied. Differential equations describe a numerical simulation of 3D space, while 3D geometry describes a corresponding graphical space. The space is designed to provide a variety of material conditions for generating events, comprising a meta-instrument for graphics and sound production. Compositional decisions determine the dimensions of the space, the locations of objects that may be struck by particles, and the material properties of those objects.

The particles introduced into the space are the only objects with sound-producing motion. The graphics depicting immobile objects are not simulated from physically-based equations. Instead their positions and materials are recorded and used to index sound production. Sound synthesis models represent the variety of physical properties of the objects.

Animation Software Interface

Graphics in Music for Unprepared Piano were produced using Alias/Wavefront’s Maya. Audio control messages were generated by analyzing the animation. Maya is a high-end commercial application used for producing three-dimensional computer animation and visual effects. Users create geometric models, set up motion paths, and render a series of frames that are viewed on videotape or film. Typically, animation is achieved by posing the objects at important points called keyframes; frames in-between the keyframes are interpolated by the software. A great many keyframes must be specified to create realistic-looking motion. Maya provides the ability to generate animation procedurally with physically-based simulations. Using this method, animators specify the physical properties and initial locations and velocities of the objects, and the software generates the movement by solving a system of differential equations. Procedural animation can produce fairly realistic motion with much less effort on the part of the animator, and is most useful for particle effects such as liquids, smoke, and hair.

Music for Unprepared Piano uses Maya’s particle dynamics to simulate a barrage of tennis balls and marbles striking the strings and case of a piano. Visually we provide a geometric depiction of a grand piano and a fire hose on a concert stage - the hose designating the initial position for particles. Figure 2 shows the wire frame model of this space. Events are composed solely by scheduling a sequence of parameter changes to the particle system. Parameters include the number of particles, their initial positions, their initial velocity (force and direction of motion), and the times at which new particles are introduced. Material properties of the particles include diameter, mass, friction, and restitution. Each blast of the fire hose schedules a new group of particles, and produces an extended sequence of events caused by the motion path and collisions of each particle. The acoustic dramaturgy derives directly from the simulated causality in the equations of motion. Figure 3 represents the “score” and resulting sound file for one realization of the animation.
3. Sound Morphing from Geometric Deformations

Toy Angst is a short animation featuring sound morphing to dramatize narrative events. The sounds are related to a visual object that is understood to produce sounds: an inflated toy ball that squeaks when pressure is applied. The degree of pressure and the resulting deformation of the ball exhibit a corresponding transformation of the squeaking sounds. When excessive pressure is applied to the ball it deforms into the shapes of other toys. When the ball morphs the squeaking sound morphs into a corresponding new sound. The repertoire includes an elephant, a car, a cat, and their combinations.

As with Music for Unprepared Piano the sound control in Toy Angst is derived from the instruction sets and data that control the graphics. The sound is not derived directly from the graphical image. In this case the underlying model is spatial and gestural: deformations of geometric objects are encoded as parameterized gestures. The parameters are represented as animation control channels, and each channel provides a time-varying function applied to the 3D transformation of the vertices and edges of one or more polygons that make up the object. Some control channels affect groups of other channels. Select control functions are applied in parallel to the parameters of a sound morphing synthesis algorithm. The synthesis algorithm is based upon spectral analysis and does not operate in real-time. The resulting sounds are manually situated in the animation soundtrack aligned to the graphical events.

Bandwidth-Enhanced Analysis and Re-synthesis

Sounds for deformation and morphing are derived from spectral analyses of natural sounds, using a new analysis and re-synthesis method to apply transformations. The Reassigned Additive Bandwidth-Enhanced Model (RBEAM) shares with traditional sinusoidal methods the notion of temporally connected partial parameter estimates, but by contrast the reassigned estimates are non-uniformly distributed in both time and frequency, yielding greater resolution in time and frequency than is possible using conventional additive techniques (Fitz 2000a, 2000b).

RBEAM expands the notion of a partial to include the representation of both sinusoidal and noise energy by a single component type. RBEAM partials are defined by a trio of synchronized breakpoint envelopes specifying the time-varying amplitude, center frequency, and noise content (or bandwidth) for each component. The bandwidth envelope represents a mixture of sinusoidal and noise energy with a single component. Bandwidth association is the process...
of constructing the bandwidth envelopes for RBEAM partials, determining how much noise energy should be represented by each partial.

RBEAM partials are synthesized using a stochastic modulation technique for spectral line widening. The time-variant parameters of RBEAM partials can manipulate both sinusoidal and noisy components of sound in an intuitive way, using a familiar set of controls. The encoding of noise associated with an RBEAM partial is robust under partial parameter transformations, and is independent of other partials in the representation. RBEAM partials can be modified without destroying the character of noisy sounds or introducing audible artifacts related to the representation of noise.

### Sound Morphing

Sound morphing is achieved by interpolating the time-varying frequencies, amplitudes, and bandwidths of corresponding partials obtained from RBEAM analysis of the source sounds (Haken 1998). Correspondences between partials in the source sounds are established by the process called distillation. The frequency spectrum is partitioned into channels having time-varying widths and center frequencies, and each channel having a unique identifier. All partials falling in a particular channel are distilled into a single partial, which is assigned the corresponding identifier, leaving at most a single partial per frequency channel (channels that contain no partials are not represented in the distilled partial data). Distillation is performed on the initial and the destination sound to be used in the morph, with common channel identifiers assigned to channels that share the same frequency range. When a morph operation is applied each pair of distilled partials having a common identifier is interpolated at their frequency, amplitude, and bandwidth envelopes according to the morphing function. Distilled partials in one source that have no corresponding partial in the other source are faded in and out according to the morphing function. The product of a morph is a set of partials, just like the source data, specifically, there is a single partial for each distillation channel identifier represented in either of the source sounds.

The various morph sources need not be distilled using identical sets of frequency channels. However if the frequencies of a source and a destination channel are dissimilar, then dramatic partial frequency sweeps may dominate other audible effects of the morph, so care must be taken to coordinate the frequency channels used in the distillation process. Significant temporal features of the source sounds can be synchronized to optimize continuity in morphing results. **Time dilation** stretches or compresses the time domain of the analysis data. Dilation can be applied to each source sound prior to morphing in order to align features that the composer wishes to coincide in the resulting morph.

### Synchronization with Graphics

To synchronize the sound and graphical transformations, two processes are applied: (1) **Time dilations** are scheduled so that audible features of the morph coincide with visible features of the graphical deformation. (2) RBEAM control envelopes are modulated by the graphical control functions of the animation deformation channels. Modulation of the RBEAM channels imparts the temporal character of the graphical deformation gesture onto the time domain characteristics of the audio morph. The mapping of graphics deformation channels onto audio morphing channels may be considered crude compared to an integrated physical model that describes the structure and the equations of motion for both light and audio wave propagation. However, considered as a parameterized model of a gesture the animation channels provide a coherent perceptual model that serves as a source for parallel transformation of sounds, and automates a level of detail that is difficult to manually reproduce.

### 4. Conclusions

These projects required significant hands-on conversion to fit an automated sound production system developed for VR, into a linear track-based post production environment. Further work is required to determine the feasibility to generalize this approach for production environments. Key issues include the reliability of animation data and a general method for its interpretation. Adoption of automated sound production in animation systems could provide a benefit for animators, and provide a new class of visual interface for composers to organize and realize works.

### 5. References

- Fitz, K., Haken, L. and Christensen, P. "A New Algorithm for Bandwidth Association in Bandwidth-Enhanced Additive Sound Modeling." In this volume.
- Fitz, K., Haken, L. and Christensen, P. "Transient Preservation Under Transformation In an Additive Sound Model." In this volume.