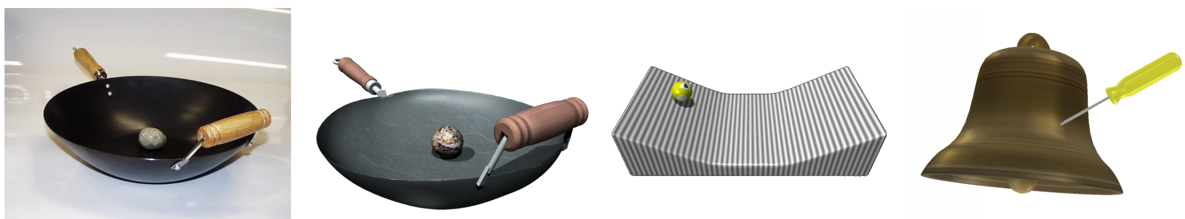
FoleyAutomatic: Physically-based Sound Effects for Interactive Simulation and Animation

Kees van den Doel, Paul G. Kry, and Dinesh K. Pai
University of British Columbia
{kvdoel | pgkry | pai}@cs.ubc.ca



(a) Real rock in wok (b) Virtual rock in wok (c) Rolling and Sliding (d) Interaction

Figure 1: Animations for which sound effects were automatically added by our system, demonstrated in the accompanying video. (a) A real wok in which a pebble is thrown; the pebble rattles around the wok and comes to rest after wobbling. (b) A simulation of a pebble thrown in wok, with all sound effects automatically generated. (c) A ball rolling back and forth on a ribbed surface. (d) Interaction with a sonified object.

Abstract

We describe algorithms for real-time synthesis of realistic sound effects for interactive simulations (e.g., games) and animation. These sound effects are produced automatically, from 3D models using dynamic simulation and user interaction. We develop algorithms that are efficient, physically-based, and can be controlled by users in natural ways. We develop effective techniques for producing high quality continuous contact sounds from dynamic simulations running at video rates which are slow relative to audio synthesis. We accomplish this using modal models driven by contact forces modeled at audio rates, which are much higher than the graphics frame rate. The contact forces can be computed from simulations or can be custom designed. We demonstrate the effectiveness with complex realistic simulations.

CR Categories and Subject Descriptors: I.3.7 [Computer Graphics]: Three-Dimensional Graphics and Realism - Animation, Virtual reality; I.3.5 [Computer Graphics]: Computational Geometry and Object Modeling - Physically based modeling; I.6.5 [Simulation and Modeling]: Model Development - Modeling methodologies; I.6.8 [Simulation and Modeling]: Types of Simulation - Animation, Combined, Gaming; H.5.5 [Information Interfaces and Presentation (e.g., HCI)]: Sound and Music Computing - Methodologies and techniques, Modeling; H.5.2 [Information Interfaces and Presentation (e.g., HCI)]: User Interfaces - Auditory (non-speech) feedback.

Additional Key Words: Animation Systems, Computer Games, Multimedia, Physically Based Animation, Physically Based Modeling, Sound Visualization, Virtual Reality, Head Mounted Displays.

1 Introduction

The importance of sound in computer graphics and interaction has been recognized for a long time. Sounds are known to be useful for human-computer interfaces in general [3, 4, 15, 16]; Buxton [4] points out that sound can be used for alarms and warnings, status and monitoring indicators, and encoded messages. Sounds play a similar but more subtle role in animation and interaction, conveying — in addition to quantitative information — a sense of presence, realism, and quality. Sound effects, sometimes called Foley sounds, are therefore widely used in animation, film and games industries.

However, the creation of sound effects remains a slow, labor intensive process; sound effects need to be added by hand by talented sound designers. With the system described in this paper, many types of sound effects due to contact interaction can be synthesized automatically. Fig. 1 shows images of some examples, which are presented with audio on the accompanying video tape.

The desire to remedy this situation, by automatically synthesizing sound effects based on the physics of interaction, also has a long history. For instance, Gaver, in his pioneering work [14, 15] discussed how contact sounds are used in "everyday listening" and can be synthesized for simple objects like bars. In another pioneering step, Takala and Hahn [39] described a general methodology for producing sound effects for animation. A sound could be attached to each object, triggered by events and synchronized with the animation, and rendered using a sound pipeline analogous to the usual image rendering pipeline. They also describe synthesis of collision sounds using a version of modal synthesis that we describe below. We discuss other relevant work in §1.1.

Despite these early successes nearly a decade ago, automatic sound synthesis is still not an integral part of animation or interactive simulations such as games. Why is this? We speculate that there are several reasons.

First, the models of the physical interaction used previously correspond to simple impacts which triggered sounds. This works well for animations of bouncing objects but fails to capture the subtleties of *continuous contact* sounds produced by sliding and rolling which inevitably accompany the interaction. It is precisely these kinds of continuous contact sounds which depend on the physical behavior of the animation that are most difficult for humans to create by hand and would benefit from automatic synthesis.

Second, the physical simulation used to drive the sound synthesis had the wrong temporal and spatial scales, having been originally designed for visual simulation. High quality audio synthesis occurs at a sampling rate of 44.1 KHz, about 1000 times faster than the graphics. It is necessary to bridge this gap since many surface properties such as roughness produce force (and perceptible sound) variations that can not be captured at the visual simulation rate. Running a detailed simulation using FEM at audio rates is an interesting possibility [28], but it is expensive, requiring extremely small time-steps. It is also difficult to integrate such an approach with rigid body simulators, which are widely used in animation. On the spatial scale, physical simulations were also designed for polyhedral models, which produce annoying auditory artifacts due to discontinuities in the contact force. For rolling, methods for dynamic simulation of contact between smooth surfaces are needed.

Third, the models associated single sounds with objects, and failed to account for the variation in sound timbre and surface properties over the surface. The importance of timbral variation was described by [45] but they only described simple impact interactions, and did not integrate the surface variation with 3D geometric models in a dynamics simulation. While it has been believed that it should be possible to associate timbre and roughness textures with the 3D geometry, just as one associates color and material information for visual rendering, this hasn't been effectively demonstrated previously in part because it was not clear exactly what values must be stored in the texture.

Finally, previous work attempted to synthesize too large a category of sounds from a single class of models, attempting to synthesize sounds ranging from contact sounds to animal cries. This often resulted in poor quality sounds, because the underlying model does not sufficiently constrain the sounds to be realistic. They also force the user to think about sound simulation at too low a level, in terms of waveforms and filter graphs. While experienced sound designers can use these tools to produce good sounds, most users find it extremely difficult.

All these factors add up to the fact that automatically generated sounds could not compete with the realism and subtlety of recorded sounds massaged by an experienced sound designer. While we believe that there will always be a need for hand construction of some sound effects, we believe it is not necessary for many animations, and impossible for interactive simulations where the interaction is not known in advance. We will show in this paper how the problems listed above can be effectively addressed, making automatic synthesis of contact sounds practical and attractive for animation and interactive simulation.

Specifically, we (1) show why the modal synthesis technique is an appropriate model for interaction sounds including continuous contact sounds; (2) describe a dynamic simulation algorithm for computing contact forces suitable for sound synthesis; (3) develop micro-simulation techniques for producing high resolution "audio force" from nominal contact forces produced by dynamics simulation at video rates; (4) demonstrate an implemented system which combines all these techniques, and utilizes sound maps that include variations in timbre and roughness over the surface of an object.

By focusing on the specific, but large, category of interaction sounds due to contact, we are able to incorporate more of the relevant physical phenomena to constrain the sound synthesis; the resulting sounds are not only synchronized to subtle contact variations, but are also of high quality. We believe that the work described in this paper has the potential to finally realize the dream of automatic synthesis of interaction sounds in animation and interactive simulations such as games.

1.1 Related Work

Apart from the pioneering work mentioned in the introduction, a number of studies dealing with sound-effects in animations have appeared.

Sound-effects for virtual musical instrument animation were investigated by Cook [7]. In [17] a number of synthesis algorithms for contact sound are introduced which include a scraping algorithm. Synthesizing sounds from FEM simulations is described in [28]. Synthesizing sound effects tightly synchronized with haptic force display is described in [10]. Automated measurement of sound synthesis parameters is described in [30].

Many studies of room acoustics and three-dimensional sound in graphical environments have appeared. These studies are orthogonal to the work presented here and provide algorithms to further process the contact sounds described in this paper to add additional perceptual cues about the environment and sound location. See, for instance, [42, 13, 32], the book [2], and references therein.

Numerous publications on the synthesis of musical instrument sounds have appeared. Most musical instrument synthesis algorithms use the unit-generator paradigm introduced by Mathews [24]. In [21] audio synthesis is discussed from a theoretical point of view and synthesis techniques are analyzed using various criteria. Additive synthesis algorithms based on the FFT were described in [11]. Synthesis of xylophone sounds, which are close to the sounds we are interested in, is described in [5]. A number of algorithms for percussion sounds were described in [8]. In [34, 31] an analysis-synthesis approach was given using a decomposition into sinusoidal and residual components.

Tribology research on noise generated by sliding and rolling contacts has focused mainly on machine noises. In [1] the role of surface irregularities in the production of noise in rolling and sliding contacts was investigated. In [19, 26] the nonlinear vibrations at a Hertzian contact were studied.

The perception of object properties from their sounds is studied, for instance, in [14, 16, 23].

1.2 Overview

The new contributions of this paper are the theoretical models for contact interactions for impact, rolling, and sliding, and their close integration with the dynamics of simulations, and the extension of modal synthesis to incorporate multiple simultaneous continuous interactions at different locations with realistic timbre shifts depending on the contact point.

We have implemented a software system called FoleyAutomatic, composed of a dynamics simulator, a graphics renderer, and an audio modeler to create interactive simulations with high-quality synthetic sound.

The remainder of this paper is organized as follows. In Section 2 we review modal synthesis, describe how we use it to obtain contact location dependent sounds using the modal gains mapped to the geometry (the "audio-texture") and describe how we obtain parameters for physical objects. Section 3 discusses the use of a dynamics simulator in animation and simulation and how it can be used to provide control parameters which can drive realistic responsive audio synthesis. In Section 4 we introduce novel algorithms for physically parameterized impact forces, scraping and sliding forces, and rolling forces, which are used to excite the modal resonance models. Section 5 describes the implementations we made to demonstrate the effectiveness of the algorithms and we present our conclusions in Section 6.

2 Modal Resonance Models

Interacting solid objects can make a variety of different sounds depending on how and where they are struck, scraped, or rolled. During such interactions, the rapidly varying forces at the contact points cause deformations to propagate through the solid, causing its outer surfaces to vibrate and emit sound waves.

A good physically motivated synthesis model for solid objects is modal synthesis [46, 15, 25, 8, 44, 45], which models a vibrating object by a bank of damped harmonic oscillators which are excited by an external stimulus.

The modal model $\mathcal{M} = \{\boldsymbol{f}, \boldsymbol{d}, \boldsymbol{A}\}$, consists of a vector $\boldsymbol{f}$ of length N whose components are the modal frequencies, a vector $\boldsymbol{d}$ of length N whose components are the decay rates, and an $N \times K$ matrix $\boldsymbol{A}$, whose elements a_{nk} are the gains for each mode at different locations. The modeled response for an impulse at location k is given by

$$y_k(t) = \sum_{n=1}^{N} a_{nk} e^{-d_n t} \sin(2\pi f_n t), \qquad (1)$$

for $t \geq 0$ (and is zero for $t < 0$). The frequencies and dampings of the oscillators are determined by the geometry and material properties (such as elasticity) of the object, and the coupling gains of the modes are related to the mode shapes and are dependent on the contact location on the object.

This model is physically well motivated, as the linear partial differential equation for a vibrating system, with appropriate boundary conditions, has these solutions. Another attractive feature of modal models is that the number of modes can be changed dynamically, depending on available computational resources, with a graceful degradation in audio quality. A modal resonator bank can be computed very efficiently with an $O(N)$ algorithm [15, 45, 43] for a model of N modes.

The modal model parameters can be obtained from first principles for some very simple geometries [45], or they could be created by "hand and hear", for example using a modal model editor such as described in [6], but for realistic complex objects we obtain them by fitting the parameters to recorded sounds of real objects (see [30]).

Points on the surface of the object where we have sampled impulse responses are mapped to points in the "sound-space" of the object, which we define in this context as the space spanned by the gains of the modes. At intermediate points we use a barycentric interpolation scheme. Audible artifacts in the form of too abrupt changes in timbre can result from too coarse a surface sampling. We set the density of the surface sampling by trial and error. This map of gains on the surface samples the absolute value of the mode-shapes [45], and with sufficiently dense sampling the mode-shapes could be recovered. The timbre of the sound can also be affected by changing the spectral content of the excitation force, as is customary in waveguide models [35] for bowed and plucked strings for example, however this is an expensive operation to perform in real-time on the audio-force, whereas changing the gain vector has negligible computational cost.

3 Dynamic Simulation for Sound

The audio synthesis techniques described in this paper can be incorporated with most multi-body methods since the parameters needed for audio synthesis are often directly available or easily computable from the simulation. For example, multi-body methods commonly compute constraint forces during periods of continuous contact and impulses to resolve collision when transient contacts occur. In addition, the speed of the contact on each surface and the slip velocity necessary for rolling and sliding sounds, though not readily available as by-products in most simulations, are easily computable from the relative velocity and the contact location.

We observe nevertheless there are features which make some multi-body methods more desirable than others. Because continuous contact audio synthesis is parameterized by contact forces and velocities, multi-body methods that accurately simulate rolling and sliding contacts are preferable. More importantly, smooth surface models should be used because the discontinuities that arise in dealing with polyhedral approximations do not lead to sufficiently smooth rolling forces.

We have developed a simulation method for continuous contact which is particularly suited to sound generation for exactly these reasons. Our technique uses rigid bodies defined by piecewise parametric surfaces. In particular, we use Loop subdivision surfaces with parametric evaluation as described in [36].

3.1 Contact Evolution

Our method evolves a system of contacting bodies in a reduced set of coordinates. Suppose body 1 and body 2 are in contact. Let the shape of the contacting patch on body 1 be described by the function $\boldsymbol{c} : (s,t) \to \mathbb{R}^3$, and on body 2 by $\boldsymbol{d} : (u,v) \to \mathbb{R}^3$. We can describe any contact configuration between patches $\boldsymbol{c}$ and $\boldsymbol{d}$ of bodies 1 and 2 by the 2-dimensional location of the contact in the domain of each patch along with the angle of rotation ψ shown in Fig. 2. Assembled in a column vector, we call these 5 independent variables the *contact coordinates* and we denote it q; i.e., $q = (s\ t\ u\ v\ \psi)^T$.

As described in [29], we use contact kinematics equations which relate the relative motion between two smooth contacting bodies to a change in the contact coordinates. For a given contact configuration, q, these equations can be written as $\phi = H\dot{q}$, where ϕ is the relative spatial velocity, H is a linear transformation from 5 dimensional space of contact coordinate velocities into a 5 dimensional subspace of the 6 dimensional space of spatial velocities.

The Netwon-Euler equations for a rigid body can be combined with the derivative of the contact kinematics equations

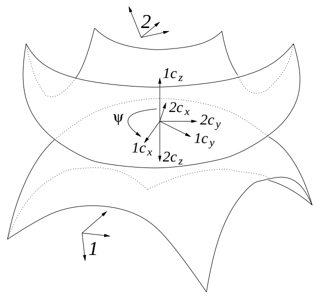


Figure 2: Contact coordinates are defined by the parameter space location of the contact on each patch along with the angle ψ.

to form an ordinary differential equation which we solve for contact coordinate accelerations. The reduced coordinate dynamics equations, though slightly more complex, can be integrated easily using explicit integrators without the need for stabilization as truncation errors do not cause interpenetration. Since the constraint in this formulation is bilateral, we check for when the constraint force goes negative and allow the objects to separate. Once separate, we allow the bodies evolve freely until we observe transient collisions occurring in close succession and proximity, at which time we switch back to reduced coordinates. We use sphere trees built from a polyhedral approximations of our models for collision detection. Approximate contact coordinates are computed from the minimum distance reported by the sphere tree. Before using these coordinates we apply a few Newton iterations, as described in [27], to find a better approximation of the contact point on the smooth surfaces.

4 Micro-simulation of Contact Interactions

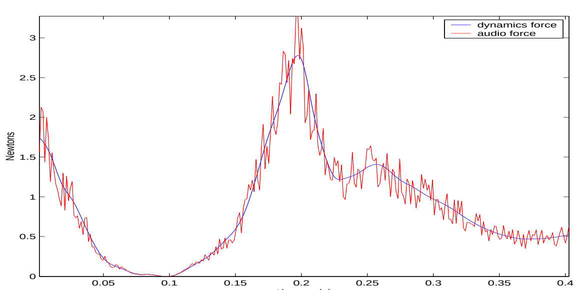


Figure 3: The dynamical force (blue) and the audio force (red) change in time at different rates. Resonance models are excited by the difference between the audio-force and the dynamics force, i.e., the rapidly varying part. Scales are exaggerated in this picture.

Even though modal models by themselves are a useful class of models, that by itself is not sufficient. For realistic contact sounds we need good, physically-based models of both the resonators and the contact interactions. The simulation of the contact interactions needs to be time stepped at least at the audio sampling rate, which is generally much higher than the simulation rate or the graphics frame-rate. Therefore the models need to be fast and simple. Often stochastic models are appropriate, since most contact interactions involve some random element. We will refer to the contact force sampled at audio rate as the "audio-force", to distinguish it from the "dynamics-force" which is sampled much coarser. See Fig. 3.

The simulation of the audio-force is similar to audio synthesis, but it does not need to be of high auditory quality, since it will not be heard directly, but is used to excite resonance models.

4.1 Impact

When two solid bodies collide, large forces are applied for a short period of time. The precise details of the contact force will depend on the shape of the contact areas as well as on the elastic properties of the involved materials.

The two most important distinguishing characteristics of an impact on an object are the energy transfer in the strike and the "hardness" of the contact [12]. The hardness affects the duration of the force, and the energy transfer relates directly to the magnitude of the force profile.

A generic model of contact forces based on the Hertz model and the radii of curvatures of the surfaces in contact was considered in [22]. A Hertzian model was also used to create a detailed model of the interaction forces between the mallet and the bars of a xylophone [5].

A simple function which has the qualitative correct form for an impact force is, for example, $1 - \cos(2\pi t/\tau)$ for $0 \leq t \leq \tau$, with τ the total duration of the contact. The force increases slowly in the beginning, representing a gradual increase in contact area, and then rises rapidly, representing the elastic compression of the materials.

We have experimented with a number of force profiles and found that the exact details of the shape is relatively unimportant and the hardness is conveyed well by the duration.

For "very hard" collisions such as a marble on a stone floor, the impact events are very fast (times of $\sim 50\mu s$ for a single contact break were measured in [38]). However, such contacts sound too "clean" in practice when represented by a single non-zero sample. Experimental data [38] shows that there are sequences of very fast contact separations and collisions during hard impacts. These micro collisions are caused by modal vibrations of the objects involved, and we have simulated this by a short burst of impulse trains at the dominant modal frequencies. Informal experiments showed that the microcollisions take place within $15ms$ for objects we studied. An impact audio-force composed of impulses at the first 4 modal resonance frequencies sounded very convincing and would be appropriate for hitting a large resonating object with a small hard object.

4.2 Scraping and Sliding

During scraping and sliding the audio-force is generated by combining an effective surface roughness model and an interaction model. The audio-force will depend on the combined surface properties of the objects in contact.

Interaction Models

A simple and effective model of scraping a surface is the phonograph needle model, whereby a needle exactly follows a given surface track. In order to generate the audio-force at a sampling rate of f_s, we generate an effective surface profile at a spatial resolution of v_{max}/f_s, where v_{max} is the maximum contact velocity for which the model is applicable. This ensures that at the maximum contact speed the surface profile can get sampled at the audio sampling rate. For a homogeneous surface this is easy to implement with a wave-table of a surface profile which will be “played back” at the rate v/v_{max} for a contact moving at speed v, where a rate of one represents the audio-force at maximum speed. This wave-table can be constructed from a mathematical roughness model as explained below or obtained experimentally. The wave-table should be oversampled in order to decrease the reduction of quality on slowdown.

For resampling we use a linear interpolation algorithm. Spurious high frequency components resulting from the discontinuities of the first derivative can be reduced by using a computationally more expensive quadratic interpolation scheme (or even more sophisticated algorithms to filter out artifacts) but this offered no audible improvement in practice. Such a model is appropriate for scraping a rough surface with an edge, or for sliding interaction as depicted in Fig. 4.

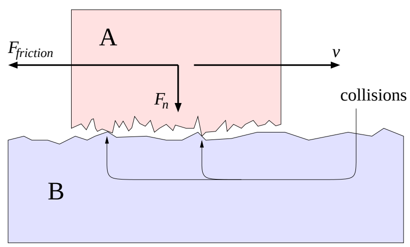


Figure 4: Sliding involves multiple micro-collisions at the contact area

For Coulomb friction, $F_{friction} = \mu F_{normal}$, the scraping audio-force volume is proportional to $\sqrt{v F_{normal}}$, assuming the force is proportional to the frictional power loss.

Effective Profile Models

We want to create a simple synthetic model of a scraping force, which has the perceptual dimensions of roughness, contact speed, and average contact force.

We characterize the profile by noise with an overall spectral shape, on which one or more peaks are superimposed. The spectral peak frequencies are moved with the contact velocity and provide the perception of changing pitch at different speeds.

We implemented this by generating fractal noise, which is noise with a power spectrum proportional to ω^{β}, passed through a reson filter [37]. The parameter β provides a surface roughness parameter and is related to the fractal dimension D by $D = \beta/2 + 2$. Empirical data shows [33, 40] than many surfaces are fractal over a wide range of scales. Values of D between 1.17 and 1.39 were reported for various machined surfaces at a length scale of 10^{-6}m. At CD quality audio sampling rate of $44100Hz$ this would correspond to a sliding speed of about $5cm/s$. The frequency of the reson is scaled with the contact velocity to produce the illusion of scraping at different speeds. The width of the resonance influences the simulated randomness of the surface, narrow peaks sounding more pitched. Informal experiments suggest that the fractal dimension correlates well with the perception of auditory roughness.

We can obtain the fractal dimension from a recording of a scrape interaction with a contact microphone and fitting the power spectrum. See Fig. 5. The reson parameters were obtained by measuring the surface structure of objects at a resolution of 0.03 mm and making an autoregressive fit using the covariance method.

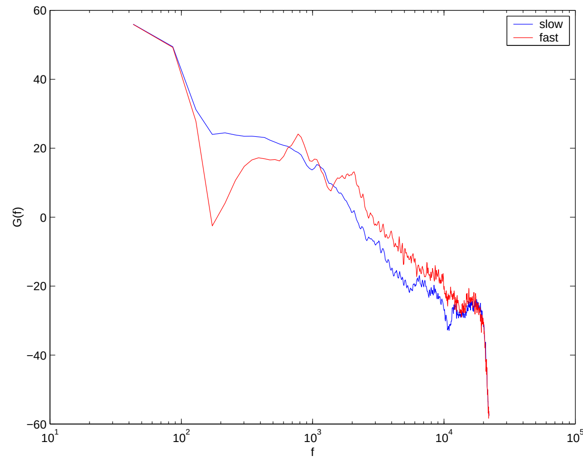


Figure 5: Power spectrum scraping smooth plastic. The fractal dimension extracted from a linear fit is roughly D=.88. The fractal model appears valid up to about 1000 Hz.

4.3 Rolling

Like scraping forces, rolling forces are produced by irregularities of the surfaces in contact, but because the surfaces have no relative speed at the contact point, the contact interactions are different and this is reflected in a difference in sound.

Rolling seems to be a less understood interaction as far as audio is concerned. Experimental studies of rolling sounds have focused mainly on very specific machine sounds like train wheels [41] and do not suggest a universal rolling model.

What exactly causes the perception of rolling versus sliding is also not known. Some studies [20] suggest that the periodicity of rolling sounds plays an important role in distinguishing it from sliding. However, we believe rolling sounds are different in other respects too.

Why do we hear anything when a smooth ball rolls on a rough table? A possible reason is that the surface of the ball rides on the asperities of the surface, and constantly collides with them but because of the large (w.r.t. the length scale of the roughness) radius the ball only “sees” the large scale surface structure. The smaller the ball, the more small details are “felt” by the rolling. Because the collisions in rolling occur just in front of the contact area (very small area) the downward motion of the ball is very small so the collisions will be very soft, i.e., drawn out in time. See Fig. 6. This suggests rolling interactions are similar to scraping interactions but because of this size effect only the low frequency content of the effective profile plays a role.

This suggests a similar model as for scraping but with an additional low-pass filter with adjustable cutoff frequency capturing the rolling quality.

Though this simple model provides reasonably convincing rolling sounds, they are not as convincing as the scraping and sliding sounds. Analysis of recorded rolling sounds suggested that the rolling force couples stronger to the modes than the sliding force. This would mean the linear model of an independent audio-force being applied to a linearly vibrating ob-

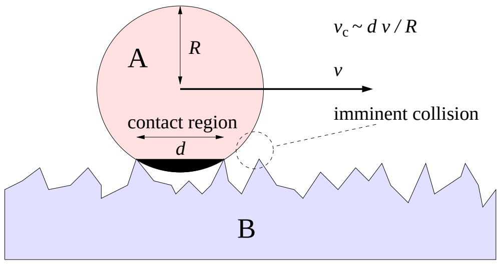


Figure 6: Rolling audio-force. The collision velocity v_c is related to the contact region d as indicated.

ject is no longer applicable, as the rolling audio-force seems to "know" the modes of the object. We speculate this is because the surfaces are not moving relative to each other and are therefore in stronger contact than during sliding, which could generate a kind of feedback mechanism leading to this effect.

This observation is consistent with the observation in [18] that a gamma-tone model with $\gamma = 2$ driven by noise generates better rolling sounds than a pure modal model. A $\gamma = 2$ model driven by noise has the same spectrum as a reson filter driven by noise with spectral envelope $S(\omega) = 1/\sqrt{(\omega-\rho)^2+d^2}$, with ρ and d the frequency and damping of the reson. This spectral envelope is enhanced near the objects resonance modes, as we observed in data.

We have experimented with this real-time spectral modification of the rolling-force to obtain what appear to be better rolling sounds, at the price of the extra computation required for the filtering of the audio-force. Further study is needed to carefully evaluate the improvement.

5 Results

We have implemented the ideas presented above in our FoleyAutomatic system. An audio synthesis toolkit was developed in pure Java. The toolkit consists of three layers of software objects, a *filter-graph* layer which provides objects to build unit-generator graphs [24], a *generator* layer with basic audio processing blocks such as wave-tables and filters, and a *model* layer which contains implementations of the physical models described above. For efficiency, the audio processing elements are all vectorized, i.e., they process audio in frames (typically 20 ms blocks) of samples rather than on a sample per sample basis as in for example [9]. On a $900Mhz$ Pentium III we found that we can synthesize about 800 modes at a sampling rate of $44100Hz$, good enough for about 10 reasonably complex objects, with enough cycles to spare for the graphics and the dynamics.

Our dynamics simulation is implemented in Java and uses Java3D for the graphics rendering. We use Loop subdivision surfaces to describe the boundaries of objects in our simulations. Sphere trees detect collisions between the surfaces. The simulator runs in real time for the relatively simple examples shown on the accompanying video, and can also record simulations along with audio parameters to be played back later in real time for sound design and higher quality graphics rendering.

Fig. 1(d) shows real time interaction with a model of bell. Using a PHANToM haptic interface we can tap and scrape the sides of the bell with a virtual screwdriver. Audio for the bell is generated with the 20 largest modes and sounds very convincing.

We created a detailed dynamic simulation of a pebble in a metal wok, that can bounce around, roll, and slide. A modal model of the wok was created from measurements of a real wok. It would be nice to use a scientific method to objectively select parameters like the density of the surface sampling or the number of modes rendered, but this is not possible due to the lack of objective criteria to determine how well a given synthetic sound is perceived as approximating a target sound. Lacking such a perceptual metric, we had to resort to "earballing" the approximations, with user input. The construction of such a measure would greatly advance the state of the art of our models.

The impulse response was recorded at 5 locations at equidistant radial points, and a modal model of 50 modes was extracted (of which only 10 contribute significantly). A set of 5 gain vectors $\boldsymbol{a}$ (of dimension 50) was obtained. We display the "brightness" of the sound-map, as measured by the frequency weighted average of the gains in Fig. 7. The

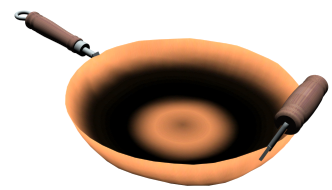


Figure 7: Sound map on wok. Visual brightness is mapped onto audio brightness.

surface roughness was measured with a contact mike and was used to construct audio surface profiles. Our dynamics simulator was set up to simulate a small, irregularly shaped rock being thrown into the wok. No more than one contact could occur at a time. This simulation was then used to drive the audio synthesis and graphics display simultaneously. In Fig. 8 we show dynamics variables driving the audio. Various other examples are shown in the accompanying video, and demonstrate impacts, sliding and rolling.

5.1 Discussion

Although our audio algorithms are physically well motivated, they are not derived using first principles from known physical laws. The physical phenomena responsible for the generation of contact sounds, such as the detailed micro-collision dynamics at the interface of sliding surfaces, are so complex that such an approach to interactive simulation seems infeasible.

We evaluated the quality of our generated audio informally, by comparing the animation with reality. See, for example, the video of the real versus the animated pebble thrown in a wok. As there are currently no existing real-time audio synthesis systems to compare our results with, our evaluation remains qualitative.

Ultimately, one would like to assign an audio quality metric to the synthesized sound-effects, which measures how good the audio model is. There are two obstacles preventing us from constructing such a measure. First, phenomena such as an object falling and rolling have sensitive dependency on initial conditions, and it is impossible to compute the exact trajectories using Newton's laws. In order to compare the real sound with the synthesized sound one therefore has to somehow extract the perceptually relevant statistical properties of complex sounds. This is an unsolved and important

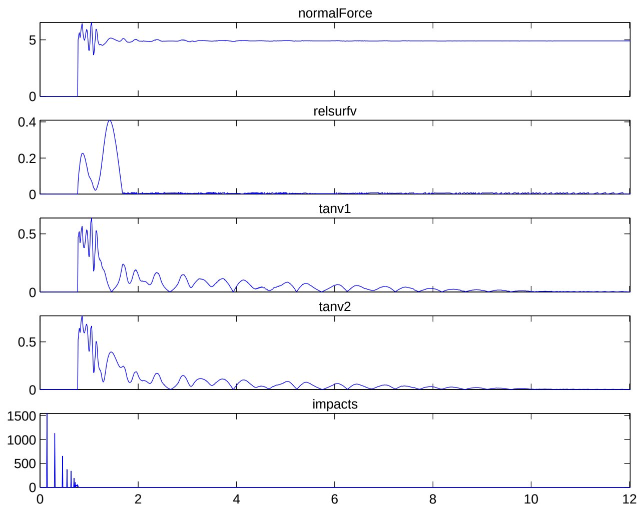


Figure 8: Dynamics variables driving audio for a single object dropped in a wok. Shown are the normal force, the sliding speed, rolling speeds (speed of contact point w.r.t. the surface) on both objects, and the impact force.. The motion (also shown on the video) clearly shows that the object first bounces several times, then performs a roll-slide motion on the curved surface of the wok, and finally wobbles around at the bottom.

problem. Second, even for reproducible events such as single impacts (for example striking a bell) there is no obvious metric to compare synthesized and real sound. Obvious measures such as the least square distance of the sampled audio are not satisfactory because they do not correspond at all to perception. We believe constructing such a measure would require substantial research on audio perception, and would significantly enhance the power of the methods presented here.

6 Conclusions

We have described a collection of methods to automatically generate high-quality realistic contact sounds driven by physical parameters obtained from a dynamics simulation with contacts, using physically motivated sound synthesis algorithms integrated with the simulation. Once the model parameters are defined, the sounds are created automatically. This enables the user of the interactive simulation to experience realistic responsive auditory feedback such as is expected in real-life when touching, sliding, or rolling objects.

Solid objects were modeled by modal resonance banks and realistic timbre changes depending on the location of the interactions were demonstrated. Because of the different rates at which the contact forces which excite the modal resonance models need to be computed, the "dynamics-force" and the "audio-force" have to be modeled with different models and we have presented several algorithms for impact, sliding, and rolling forces, using physically-based models of surface textures and contact interactions which were designed specifically for audio synthesis.

We have implemented the algorithms using a dynamics simulator and an audio synthesis package written in Java. The audio synthesis and simulation can run in real-time on desktop computers without special hardware.

We believe these algorithms have the potential to dramatically increase the feeling of realism and immersion in interactive simulations, by providing high-quality audio effects which provide important perceptual cues which, combined with high quality graphics and dynamics simulation, provide a much more compelling illusion of reality than the sum of their contributions.

References

[1] T. Ananthapadmanaban and V. Radhakrishnan. An investigation on the role of surface irregularities in the noise spectrum of rolling and sliding contacts. *Wear*, 83:399–409, 1982.

[2] D. R. Begault. *3-D Sound for Virtual Reality and Multimedia.* Academic Press, London, 1994.

[3] W. Buxton. Introduction to this special issue on nonspeech audio. *Human Computer Interaction*, 4(1):1–9, 1989.

[4] W. Buxton. Using our ears: an introduction to the use of nonspeech audio cues. In E. Farrell, editor, *Extracting meaning from complex data: processing, display, interaction. Proceedings of the SPIE*, volume Vol 1259, pages 124–127, 1990.

[5] A. Chaigne and V. Doutaut. Numerical simulations of xylophones. I. Time domain modeling of the vibrating bars. *J. Acoust. Soc. Am.*, 101(1):539–557, 1997.

[6] A. Chaudhary, A. Freed, S. Khoury, and D. Wessel. A 3-D Graphical User Interface for Resonance Modeling. In *Proceedings of the International Computer Music Conference*, Ann Arbor, 1998.

[7] P. R. Cook. Integration of physical modeling for synthesis and animation. In *Proceedings of the International Computer Music Conference*, pages 525–528, Banff, 1995.

[8] P. R. Cook. Physically informed sonic modeling (PhISM): Percussive synthesis. In *Proceedings of the International Computer Music Conference*, pages 228–231, Hong Kong, 1996.

[9] P. R. Cook and G. Scavone. The synthesis toolkit (STK), version 2.1. In *Proc. of the International Computer Music Conference, Beijing*, 1999.

[10] D. DiFilippo and D. K. Pai. The AHI: An audio and haptic interface for contact interactions. In *UIST'00 (13th Annual ACM Symposium on User Interface Software and Technology)*, 2000.

[11] A. Freed and X. Depalle. Synthesis of Hundreds of Sinusoidal Partials on a Desktop Computer without Custom Hardware. In *Proceedings of The International Conference on Signal Processing Applications and Technology*, Santa Clara, 1993.

[12] D. J. Fried. Auditory correlates of perceived mallet hardness for a set of recorded percussive sound events. *J. Acoust. Soc. Am.*, 87(1):311–321, 1990.

[13] T. A. Funkhouser, P. Min, and I. Carlbom. Real-time acoustic modeling for distributed virtual environments. *Proc. SIGGRAPH 99, ACM Computer Graphics*, 1999.

[14] W. W. Gaver. *Everyday listening and auditory icons.* PhD thesis, University of California in San Diego, 1988.

[15] W. W. Gaver. Synthesizing auditory icons. In *Proceedings of the ACM INTERCHI 1993*, pages 228–235, 1993.

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

[17] J. K. Hahn, H. Fouad, L. Gritz, and J. W. Lee. Integrating sounds and motions in virtual environments. In *Sound for Animation and Virtual Reality, SIGGRAPH 95 Course 10 Notes*, 1995.

[18] D. J. Hermes. Synthesis of the sounds produced by rolling balls. In *Internal IPO report no. 1226*, IPO, Center for User-System Interaction, Eindhoven, The Netherlands, 2000.

[19] D. P. Hess and A. Soom. Normal vibrations and friction under harmonic loads. II. Rough planar contacts. *Transactions of the ASME. Journal of Tribology*, 113:87–92, 1991.

[20] M. M. J. Houben, D. J. Hermes, and A. Kohlrausch. Auditory perception of the size and velocity of rolling balls. In *IPO Annual Progress Report*, volume 34, 1999.

[21] D. Jaffe. Ten criteria for evaluating synthesis and processing techniques. *Computer Music Journal*, 19(1):76–87, 1995.

[22] K. L. Johnson. *Contact Mechanics.* Cambridge University Press, Cambridge, 1985.

[23] R. L. Klatzky, D. K. Pai, and E. P. Krotkov. Hearing material: Perception of material from contact sounds. *PRESENCE: Teleoperators and Virtual Environments*, 9(4):399–410, 2000.

[24] M. V. Mathews. *The Technology of Computer Music.* MIT Press, Cambridge, 1969.

[25] J. D. Morrison and J.-M. Adrien. Mosaic: A framework for modal synthesis. *Computer Music Journal*, 17(1), 1993.

[26] P. R. Nayak. Contact vibrations. *Journal of Sound and Vibration*, 22:297–322, 1972.

[27] D. D. Nelson, D. E. Johnson, and E. Cohen. Haptic rendering of surface-to-surface sculpted model interaction. In *Proceedings of the ASME Dynamic Systems and Control Division*, volume DSC-Vol. 67, pages 101–108, 1999.

[28] J. F. O'Brien, P. R. Cook, and G. Essl. Synthesizing Sounds from Physically Based Motion. In *SIGGRAPH 01*, 2001.

[29] D. K. Pai, U. M. Ascher, and P. G. Kry. Forward Dynamics Algorithms for Multibody Chains and Contact. In *Proceedings of the 2000 IEEE International Conference on Robotics and Automation*, pages 857–863, 2000.

[30] D. K. Pai, K. van den Doel, D. L. James, J. Lang, J. E. Lloyd, J. L. Richmond, and S. H. Yau. Scanning physical interaction behavior of 3D objects. In *Computer Graphics (ACM SIGGRAPH 2001 Conference Proceedings)*, 2001.

[31] X. Rodet. Musical Sound Signal Analysis/Synthesis: Sinusoidal + Residual and Elementary Waveform Models. In *IEEE Time-Frequency and Time-Scale Workshop 97*, Coventry, Grande Bretagne, 1997.

[32] L. Savioja, J. Huopaniemi, T. Lokki, and R. Vnnen. Virtual environment simulation - Advances in the DIVA project. In *Proc. Int. Conf. Auditory Display*, Palo Alto, USA, 1997.

[33] R. S. Sayles and T. R. Thomas. Surface topography as a non-stationary random process. *Nature*, 271:431–434, 1978.

[34] X. Serra. *A System for Sound Analysis / Transformation / Synthesis Based on a Deterministic Plus Stochastic Decomposition.* PhD thesis, Dept. of Music, Stanford University, 1989.

[35] J. O. Smith. Physical modeling using digital waveguides. *Computer Music Journal*, 16(4):75–87, 1992.

[36] J. Stam. Evaluation of loop subdivision surfaces. In *SIGGRAPH 98*, 1998. Included on course notes CD-ROM.

[37] K. Steiglitz. *A Digital Signal Processing Primer with Applications to Digital Audio and Computer Music.* Addison-Wesley, New York, 1996.

[38] D. Stoianovici and Y. Hurmuzlu. A critical study of the applicability of rigid-body collision theory. *ASME Journal of Applied Mechanics*, 63:307–316, 1996.

[39] T. Takala and J. Hahn. Sound rendering. *Proc. SIGGRAPH 92, ACM Computer Graphics*, 26(2):211–220, 1992.

[40] T. R. Thomas. *Rough Surfaces.* Imperial College Press, London, second edition, 1999.

[41] D. J. Thompson, D. F. Fodiman, and H. Mahe. Experimental validation of the twins prediction program for rolling noise, parts I and II. *Journal of Sound and Vibration*, 193:123–147, 1996.

[42] N. Tsingos, T. Funkhouser, A. Ngan, and I. Carlbom. Modeling acoustics in virtual environments using the uniform theory of diffraction. In *SIGGRAPH 01*, 2001.

[43] K. van den Doel. *Sound Synthesis for Virtual Reality and Computer Games.* PhD thesis, University of British Columbia, 1998.

[44] K. van den Doel and D. K. Pai. Synthesis of shape dependent sounds with physical modeling. In *Proceedings of the International Conference on Auditory Displays 1996, Palo Alto*, 1996.

[45] K. van den Doel and D. K. Pai. The sounds of physical shapes. *Presence*, 7(4):382–395, 1998.

[46] J. Wawrzynek. VLSI models for real-time music synthesis. In M. Mathews and J. Pierce, editors, *Current Directions in Computer Music Research.* MIT Press, 1989.