Workshop in Virtual Reality Interactions and Physical Simulation VRIPHYS (2008)
F. Faure, M. Teschner (Editors)

Audio texture synthesis for complex contact interactions

C. Picard†[1,2] and N. Tsingos‡[1] and F. Faure§[2]

[1]REVES/INRIA Sophia-Antipolis, France
[2]INRIA Rhône-Alpes, Laboratoire Jean Kuntzmann, Université de Grenoble and CNRS, France

Abstract

This paper presents a new synthesis approach for generating contact sounds for interactive simulations. To address complex contact sounds, surface texturing is introduced. Visual textures of objects in the environment are reused as a discontinuity map *to create audible position-dependent variations during continuous contacts. The resulting synthetic profiles are then used in real time to provide an excitation force to a modal resonance model of the sounding objects. Compared to previous sound synthesis for virtual environments, our approach has three major advantages: (1) complex contact interactions are addressed and a large variety of sounding events can be rendered, (2) it is fast due to the compact form of the solution which allows for synthesizing at interactive rates, (3) it provides several levels of detail which can be used depending on the desired precision.*

Categories and Subject Descriptors (according to ACM CCS): I.3.5 [Computer Graphics]: Computational Geometry and Object Modeling Physically based modeling; I.6.8 [Simulation and Modeling]: Types of Simulation Animation

1. Introduction

Compelling audio rendering is becoming a key challenge for interactive simulation and gaming applications. Matching sound samples to interactive animation is difficult and often leads to discrepancies between the simulated visuals and their soundtrack. Furthermore, sounds of complex contact interactions should be consistent with visuals which is even more difficult when a small number of pre-recorded samples is used. Increasing the number of samples is not always possible due to the cost of recording samples. Alternatively, contact sounds can be automatically generated using sound synthesis approaches. Convincing continuous contact such as rolling or sliding requires an appropriate contact interaction model. To date, the proposed models appear quite involved and can make authoring and control challenging for a sound designer.

Our approach proposes a technique for rendering continuous contact sounds that automatically derive excitation profiles from the analysis of a surface image. The image textured onto the surface of interacting object is considered to model the force causing the sound. Contact features are modelled as a discontinuity map computed in a pre-process, making data available on the fly for real-time audio rendering. An implementation of a modal model allows for separating the material properties from the force characteristics. A sound material database is then processed with the excitation profiles resulting in the subtle audio sensation of interaction with textured or rough surfaces. Our flexible audio pipeline further proposes different levels of detail which can be chosen according to the desired granularity of the sound interactions. Our contributions are:

- a contact interaction model suitable for audio rendering,
- a solution for generating on-line audio of subtle sounding events,
- a control mechanism for the resolution of sound interactions.

† cecile.picard@sophia.inria.fr
‡ nicolas.tsingos@sophia.inria.fr
§ francois.faure@imag.fr

2. Previous Work

Interactive generation of audio to picture: Most sound rendering in current audio-visual animations consists of directly retargetting pre-recorded samples to contacts reported from a rigid-body simulation. Looping and pitch shifting audio recordings are the methods of choice for more realistic continuous contact, where the velocity and the intensity of the normal force influence the parameters.
The main problem with this method lies in the fact that each specific contact interaction requires a corresponding pre-recorded contact sound in the database of samples. Due to memory constraints, the number of samples is limited, leading to repetitive audio. Moreover matching sampled sounds to interactive animation is difficult and often leads to discrepancies between the simulated visuals and their companion soundtrack. Finally, little flexibility is provided for authoring.

Physically-based sound synthesis: Physically based synthesis of sound sources has been explored in computer graphics [OSG, RL06, vdDP96, vdDKP01] and computer music [Coo02]. Most approaches target sounds emitted by vibrating solids. For interactive simulations, a widely used solution is to apply vibrational parameters, i.e frequencies, corresponding gains and decay rates, obtained through modal analysis. Modal data can be obtained from simulations [OSG,RL06] or extracted from recorded sounds of real objects [vdDKP01]. In order to speed-up mode-based computations, a fast sound synthesis approach that exploits the inherent sparsity of modal sounds in the frequency domain has recently been introduced in [BDT*08].

Contact modeling: Contacts between bodies have been extensively investigated for sound rendering [ARR02, ARS02, PvdDJ*01]. According to [PvdDJ*01], realistic contact sounds need good, physically-based models of both the resonators and the contact interactions. Contacts are divided into two parts: impacts and continuous contacts, i.e. scraping, sliding and rolling. In [vdDP03] surface profiles were created by simply scraping a real object with a contact microphone. This technique implies that the considered objects/surfaces are available which is not necessary the case for all objects of a virtual scene. In addition, extracted profiles are dependent on experiment conditions and the main features of the surface may not be modelled. Another solution is to directly synthesize the contact force from a stochastic noise model [vdDKP01], which is quite involved. Scraping and sliding are modeled as a combination of an effective surface roughness model and an interaction model. Sliding involves multiple micro-collisions at the contact area. Scraping is characterized by noise with an overall spectral shape on which one or more peak are superimposed. The frequency of the reson filter is scaled with the contact velocity to produce the illusion of scraping at different speeds. For rolling, the surfaces have no relative speed at the contact point leading to a difference in the sound rendering. In [vdDKP04], an *impact map* is introduced to alleviate the load on the computer's CPU: the computation of impact events which drive audio synthesis is mapped onto a problem the GPU can handle. The pixels are read from the rendering surface and searched for impacts. If an impact is found, the relevant modes of the model associated with the collided object are then excited. This allows the association of multiple modal models with a single collision object.

Texture modeling: *Texture* for graphics and haptics can inspire sound rendering. In computer graphics, Perlin introduced a type of coherent noise that is the sum of several coherent-noise functions of increasing frequencies and decreasing amplitudes [Per85]. Based on a fractal observation of natural phenomena, Perlin noise is ideal for generating natural, self-similar textures such as granite, wood, marble, and clouds. In [OJSL05] a novel force model for haptic texture rendering based on the gradient of directional penetration depth is presented. The approach proposes the description of objects with high combinatorial complexity by coarse representations with their fine geometric details stored in texture images refered to as *haptic textures*.

3. Overview

Our approach borrows from physically-based sound synthesis and textured-based modeling. Contact sounds result from the combination of the material property of the objects in contact and the characteristics of the interaction force .

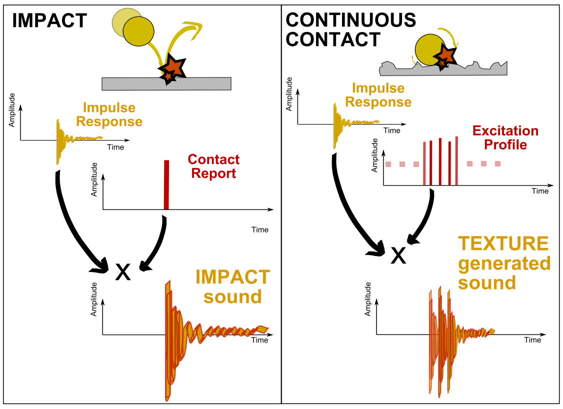

Figure 1: *Sound rendering method for impacts and continuous contacts.*

The proposed solution consists in deriving the force transmitted between interacting objects from the textured surfaces used for visual rendering. Thus, the method allows to render the sound emitted from real and virtual objects. The force interaction causing the sound is extracted as a height map. Modal parameters, i.e frequencies, gains and decays, are then processed with the excitation profiles. In addition, the excitation profiles can be of different levels of detail to achieve the desired granularity of the rendered sound. Figure 1 illustrates the modal model used for sound rendering.

In Section 4 we present our method for extracting excitation profiles for complex sounding interactions. For real-time audio-visual animations, we introduce the extraction of a *discontinuity map* and several texture encoding schemes. Section 5 presents how our technique is used in the audio pipeline and details our procedure for sound rendering. Finally, Section 6 discusses current limitations of our approach and some possible extensions.

4. Synthesis of excitation patterns

Our goal is to model the excitation force causing the object to vibrate and to output a sound comparable to a real-life situation. It has been shown that the precise details of the contact force will depend on the shape of the contact areas [PvdDJ*01]. The visual texture image is used both for visual rendering and to determine the potential excitation profile of the interacting object, in the case of rigid body interaction. This approach can be compared to Shape from Shading approach [ZTCS99] which computes the three-dimensional shape of a surface from one image of that surface. As ambiguities exist when interpreting the surface, the main features of excitation profiles are considered to be independent from the light source, the surface reflectance or the camera position in the image. As a result, synthetic *audio excitation* profiles created directly from an image surface represent the potential audio effect resulting from the interaction between this surface and another smooth one. As an example, if an image of a tiled floor is simply mapped on a groundplane, no contact information about a smooth ball colliding with the gaps is given. Thanks our approach, the synthetic *audio excitation* profiles created from the tiled floor picture provides the missing pieces of information and the sound resulting from the interaction can be rendered.
The simulation of the excitation profiles needs to advance at the audio sampling rate or greater, i.e. 44kHz which is much higher than the physics simulation rate and the graphics frame-rate. The simulation of audio excitation profiles does not need to be of high auditory quality since profiles are used to excite resonance models and will not be heard directly. Thus, profiles are generated by re-sampling a *discontinuity map* extracted in a pre-processing task, along the trajectory of the contact interaction.
Finally, the synthesis of excitation textures can create the audio sensation of interacting with regularly featured or textured surfaces and also rough surfaces. Regularly featured surfaces have been shown as significant for the sensation of rolling objects [vdDKP01], i.e. the resulting sound should be noticeably repetitive. As a consequence, our method of extracting the prominent features from visual textures appears relevant.

4.1. Extracting the discontinuity map

Extracting the discontinuities of a textured surface with an edge detection filter allows us to render the resulting sound from the interaction with this surface. Different methods of edge detection in images were tested for the modeling of the excitation profiles. The requirements for edge detection in vision are not the same as for audio. It is likely that audibly speaking, main/strong features of a surface would be prominent and should be enhanced in comparison with other "noisy" parts of the map. We used a method for discontinuity extraction which depends on the image texture appearance. We distinguished between "simple" and "complex" image textures depending on the feature content. This is analyzed in a pre-processing task using the histogram of the image. "Simple" image textures are characterized by a narrow-band histogram whereas "complex" image textures demonstrate a broad-band histogram, as seen in Figure 2. We used the *CImg* library [CIm].

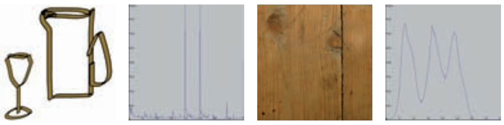

Figure 2: *"Simple" and "complex" image textures and their respective histogram.*

"Simple" image textures have prominent features without pronounced noise. Because of its efficiency in terms of computation, the Sobel filter [Chi74] is used to detect the discontinuities which may produce sound. The Sobel operator is a discrete differentiation operator, computing an approximation of the gradient of the image intensity function. This filter approximates high frequency variations but is adequate for enhancing the main features of a "simple" image texture. The *audio excitation* profile is stored as an image, an example is given in Figure 3. Knowing the positions of the objects in interaction, an excitation force is created based on the interpolated value of the closest pixel values. In order to guarantee audio quality, the discontinuity map has to be of high resolution. However, since the information is binary, the size is moderate.
"Complex" image textures are analyzed in terms of

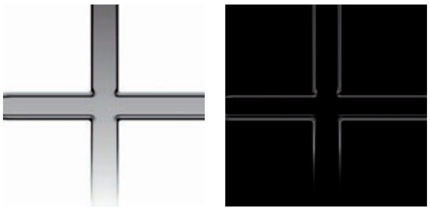

Figure 3: *"Simple" image texture: original image and discontinuity map by Sobel filtering.*

isophotes, i.e. lines drawn through areas of constant brightness. This is computed using a marching squares algorithm. The isocurves are saved as a set of points. Figure 4 shows that this technique preserves the subtle discontinuities of the tiles which are not completly smooth. In comparison, the Sobel filtered image enhances the main features only and the fine parts of the *audio excitation* profiles are not adequately rendered. An excitation force is created according to the proximity of the contact interaction to an isocurve. The difference between the current elevation gradient value and the previous elevation gradient value of the isophotes are used to modulate the amplitude of the excitation.

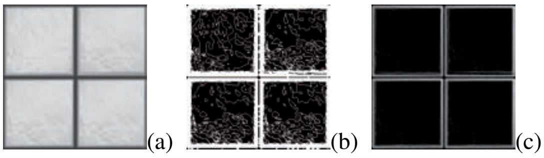

Figure 4: *"Complex" image texture: original image texture (a), corresponding isocurves image texture (b), to be compared with the Sobel filtered image (c).*

4.2. Coding the discontinuity map

When the rigid-body simulation reports a continuous contact, the discontinuity map is inspected to access the excitation profile for adequate sound rendering. In the case of discontinuity maps with high frequency content, i.e. with a highly noticeable noisy part, the high resolution of the isocurves might not be needed and an approximation would be sufficient. For this reason we propose encoding the texture as two maps corresponding to the main features and a noise map, as seen in Figure 5. In order to extract the main features, the original image texture is filtered with a "Difference of Gaussians" (DOG) filter [You87]. The DOG filter computes two different Gaussian blurs on the image, with a specific blurring radius for each, and subtracts them to yield the result. This filter offers more control parameters than the Sobel filter. The most important parameters are the blurring radii for the two Gaussian blurs. Increasing the smaller radius tends to give thicker edges, and decreasing the larger radius tends to increase the threshold for recognizing something as an edge. The blurring radii were set to 1.0 and $\sqrt[2]{1.6}$. The sensitivity threshold and the sharpness of edge representations were respectively set to 0.998 and 4.0. Our method combines this with a pre-process of bilateral filtering [CM98] in order to smooth out the noise while maintaining edges.

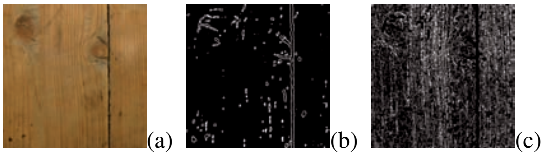

Figure 5: *Coding a texture (a) as the main features (b) and the noise part (c).*

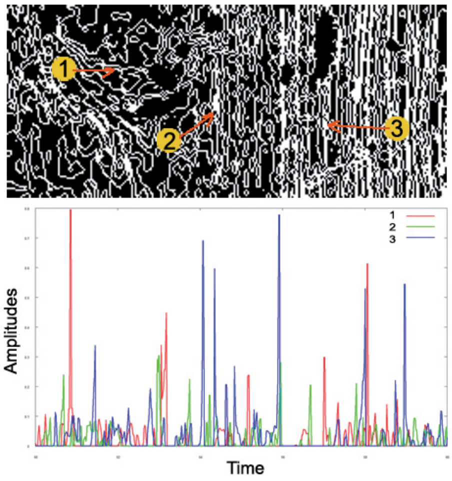

Figure 6: *Three trajectories of a smooth ball interacting with the noise map (top) and the corresponding excitation profiles extracted (bottom).*

The noise map is coded statistically. Figure 6 shows that random excitation profiles have similar behaviors and consequently, the noisy part of the texture can be considered uniformly distributed. The noise map is approximated by one of the excitation profiles, where excitation velocity and scale of the texture are known. During the real-time animation, the noise excitation profile is resampled according to the interaction velocity and the scale of the surface texture. Then, it is combined with the component corresponding to the main features during sound rendering. Table 7 shows that the coding for discontinuity map is efficient. Moreover, vectorization allows trivial scaling of the excitation profiles making them independent of the size of the original image texture. Thus, our method provides two levels of detail which can be used according to the desired precision of the rendered sound. The resolution of the excitation profile can be modulated according to the viewpoint in the scene: as an example, when the interaction is far from the listener, a low resolution is sufficient to provide realistic sound interaction. Moreover,

Type	Original Image	Isophote Vectors	Feat.+Noise Map Coding
Size	786Ko	1.09Mo	544Ko

Figure 7: *Statistics for size of the discontinuity map for an original image of 512x512 pixels (seen in Figure 5).*

complex scenes with multiple objects/surfaces interactions can be synthesized in a more computationally efficient manner using this technique.

5. Real-time audio-visual animations

Our approach proposes a flexible audio pipeline specifically adapted for real-time audio-visual animations focused on quality and variety. Sound rendering can be seen as a flexible *audio shading* (see Figure 8) allowing procedural choice of the parameters of the sound material, i.e modal parameters, and the excitation profiles for synthesis, driven by the contact report of the rigid-body simulation. In [TH92], a modular sonic extension of the image rendering pipeline is introduced where analogies between sound and texture are drawn. Similar to our approach, the sound transformations correspond to shaders in a shade tree during its traversal from a source to a camera. However, the modular architecture aims at providing a general methodology for sound rendering for animations and does not detail methods for audio resulting from complex interactions between objects/surfaces. Our audio pipeline gathers data from sound materials and excitation profiles. In our modal model implementation, modal parameters, i.e frequencies, gains and decays, encode the geometry, dimensions and materials of the interacting actors. The modal description represents any sound as a sum of oscillators characterized by their frequency, amplitude and exponential decay. This recursive representation is efficient for real-time sound synthesis. In our case, modal parameters are extracted from impulse response recordings similar to [vdDP96, vdDKP01]. The modal data is then made available on the fly for real-time sound synthesis. According to [vdDKP01], the main characteristics of the impact forces are the energy transfer and the hardness which affect duration and magnitude of the force respectively. We experimented with a number of force profiles and the exact details of the shape were found to be relatively unimportant, the hardness being well conveyed by the duration. Our method modulates the duration of the force by the kinetic energy of the interaction.

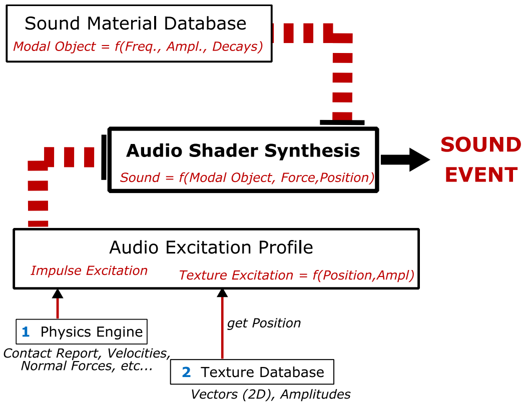

Figure 8: *Data flow into and out of the Audio Shader.*

An arbitrary number of simulated textures, or *patches*, leading to complex contact interactions, (see Section 4), are also maintained for real-time sound rendering. This organization is suggested by object-oriented programming systems, in which objects, i.e classes of texture patches, maintain their state. Each patch is associated with a block of code that implements its particular dynamics model and exports a set of editable parameters to the user interface so that the model may be varied interactively. During real-time processing, data from the rigid-body simulation such as velocity, force and positions, modal parameters and discontinuity maps of available textures are gathered to render the resulting sound. With our method, time performances for sound rendering during impact and continuous contact are 0.01msec and 0.3msec respectively. The time increase is acceptable when compared to the prohibitive cost of generating contact reports from complex geometries.

We implemented an interactive system based on our approach (see Figure 9). The simulation was driven by the physics engine *Ageia's PhysX* [AGE]. Our test application allows the user to interact with an object, (a capsule), making it roll or jump on a floor. The texture of the floor, its scale and the material of the objects interacting can be modified in real-time in order to experience the differences in the resulting sound rendering.

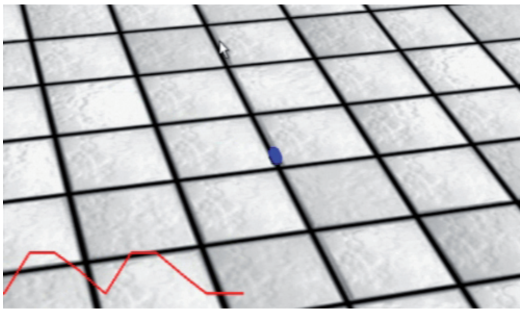

Figure 9: *Audio-visual interface: a user controls a capsule interacting with a tiled floor to experiment with the resulting sound. The amplitude of the excitation force is computed and rendered to the screen in real-time (red curve).*

6. Discussion and limitations

We demonstrated a number of applications of our approach that can be used in the context of interactive sound rendering for animations. Our approach shares a number of similarities with computer graphics and haptic techniques and in particular with the approach of [vdDKP04] where an *impact map* is presented for a two-dimensional parameterization of the impact locations and characteristics. We also introduce a two-scale approach for capturing and storing the geometric details for sound synthesis under contact situations. Our study is analogous to the approach of [OJSL05] where the haptic texture rendering is synthesized based on the gradient of the directional penetration depth. Similarly, our method proposes adding more details to the coarse representation of the texture which are not sufficient for sound rendering. Fine geometric details are coded as discontinuity maps and are related to the excitation profiles.

Our approach presents one main limitation: it does not yet consider the case of two interacting textures, where features intersect creating specific excitation profiles. This will be examined in furture work.

7. Conclusion

We have presented a novel approach for generating contact sounds for interactive simulations. This was achieved by considering the two-dimensional visual textures of objects in interaction as roughness maps to create audible and position-dependent variations during rolling and sliding. This approach allows us to guarantee coherence between visual and sound rendering. The physics engine is used to compute the motion of the objects, and the excitation force can be synthesized based upon the position of the contact in the texture-space. The force is used to excite the modal resonances of the sounding objects. A flexible audio pipeline was introduced, proposing different levels of detail which can be chosen according to the desired granularity of the rendered impacts.

We believe that improving contact sounds for interactive virtual environments is relevant and has many applications. Our technique is useful since videogames increasingly incorporate procedurally-generated content to increase variety and realism but also to address memory shortcomings, as the complexity of the environments grows.

References

[AGE] AGEIA PhysX SDK 2.7. http://www.ageia.com.

[ARR02] Avanzini F., Rath M., Rocchesso D.: Physically-based audio rendering of contact. In *ICME '02: Proceedings of the IEEE International Conference on Multimedia and Expo* (2002), vol. 2, pp. 445–448.

[ARS02] Avanzini F., Rocchesso D., Serafin S.: Modeling interactions between rubbed dry surfaces using an elasto-plastic friction model. In *Digital Audio Effects (DAFx) Conference* (2002), vol. 2, pp. 445–448.

[BDT*08] Bonneel N., Drettakis G., Tsingos N., Viaud-Delmon I., James D.: Fast modal sounds with scalable frequency-domain synthesis. In *SIGGRAPH '08: ACM SIGGRAPH 2008 papers* (New York, NY, USA, 2008), ACM, pp. 1–9.

[Chi74] Chien Y.: Pattern classification and scene analysis. *Automatic Control, IEEE Transactions on 19*, 4 (Aug 1974), 462–463.

[CIm] CImg Library - C++ toolkit for image processing . http://cimg.sourceforge.net/.

[CM98] C.Tomasi, Manduchi R.: Bilateral filtering for gray and color images. In *Proceedings of the Sixth International Conference on Computer Vision* (New Delhi, India, 1998), pp. 839–46.

[Coo02] Cook P. R.: *Real Sound Synthesis for Interactive Applications*. A. K. Peters, 2002.

[OJSL05] Otaduy M. A., Jain N., Sud A., Lin M. C.: Haptic display of interaction between textured models. In *SIGGRAPH '05: ACM SIGGRAPH 2005 Courses* (New York, NY, USA, 2005), ACM Press, p. 133.

[OSG] O'Brien J. F., Shen C., Gatchalian C. M.: Synthesizing sounds from rigid-body simulations. In *SIGGRAPH '02: ACM SIGGRAPH Symposium on Computer Animation*, ACM Press, pp. 175–181.

[Per85] Perlin K.: An image synthesizer. In *Proc. SIGGRAPH '85* (New York, NY, USA, 1985), ACM Press, pp. 287–296.

[PvdDJ*01] Pai D., van den Doel K., James D., Lang J., Lloyd J. E., Richmond J. L., Yau S. H.: Scanning physical interaction behavior of 3d objects. In *Proc. SIGGRAPH'01* (August 2001), pp. 87 – 96.

[RL06] Raghuvanshi N., Lin M. C.: Interactive sound synthesis for large scale environments. In *SI3D'06: Proceedings of the 2006 Symposium on Interactive 3D Graphics and Games* (2006), ACM Press, pp. 101–108.

[TH92] Takala T., Hahn J.: Sound rendering. *ACM Computer Graphics, SIGGRAPH'92 Proceedings 28*, 2 (July 1992).

[vdDKP01] van den Doel K., Kry P. G., Pai D. K.: Foley automatic: physically-based sound effects for interactive simulation and animation. In *Proc. SIGGRAPH '01* (New York, NY, USA, 2001), ACM Press, pp. 537–544.

[vdDKP04] van den Doel K., Knott D., Pai D. K.: Interactive simulation of complex audiovisual scenes. *Presence: Teleoper. Virtual Environ. 13*, 1 (2004), 99–111.

[vdDP96] van den Doel K., Pai D. K.: Synthesis of shape dependent sounds with physical modeling. In *Proceedings of the International Conference on Auditory Display (ICAD)* (1996).

[vdDP03] van den Doel K., Pai D. K.: Modal synthesis for vibrating objects. *Audio Anecdotes: Tools, Tips, and Techniques for Digital Audio* (2003).

[You87] Young R.: The gaussian derivative model for spatial vision: I. retinal mechanisms. *Spatial Vision 2* (1987), 273–293.

[ZTCS99] Zhang R., Tsai P.-S., Cryer J. E., Shah M.: Shape from shading: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence 21*, 8 (1999), 690–706.