Dissertation Proposal: Audio-Material Modeling and Reconstruction for Multimodal Interaction

Auston Sterling
University of North Carolina at Chapel Hill
Chapel Hill, NC 27599
austonst@cs.unc.edu

Figure 1: Virtual environments with interactive sounding objects: a museum of objects of different shapes and materials (left); windchimes blowing in the wind (right). The objects in these scenes should produce realistic impact sounds consistent with their visual appearances.

1 Introduction

User interaction scenarios use virtual environments to achieve diverse goals. Interactive training simulations enable low-risk practice of high-risk tasks, such as performing surgery or piloting an airplane. Immersive story-driven video games let users interact with another environment or become involved in an engaging narrative. Emerging social applications unite multiple users from around the world in one virtual location such that the users actually feel as through they are in the same space.

In these scenarios, a user should be able to forget their presence in the real world and temporarily experience a *sense of presence* in the virtual environment. If the user remembers the virtual environment is fake, they experience a *break in presence*. Breaks in presence reduce the emotional weight of the virtual environment, making the environment less effective at its intended goal. Better maintaining a user's sense of presence can improve the quality of their experiences. Training simulations feel more lifelike, video games convey more powerful emotions, and social interactions with other users flow more naturally.

Virtual environments most commonly recreate the senses of sight and hearing. The visual appearances and audio of the real world are relatively easily replaceable with those of a virtual world. A virtual reality (VR) headset such as the Oculus Rift or the HTC Vive replaces the sense of sight, while headphones (sometimes built into the VR headset) replace the sense of hearing. Examples of VR-enabled environments are shown in Figure 1. Humans also rely heavily on the sense of touch, but virtual environments are limited by current hardware, which cannot effectively replace our sense of touch.

Figure 2: Examples of objects which can be reconstructed virtually. The top row shows pictures of the real objects, while the bottom row shows manually-constructed meshes and textures modeling the objects. We seek to create realistic multimodal interactions with these objects.

Undesirable breaks in presence have many causes; a common cause is violation of a user's expectations about their interactions. This does not mean that virtual environments must always be realistic to maintain presence, but they must be consistent. *Sensory conflict* occurs when two senses provide conflicting expectations about one another. Each sense creates expectations for the other senses, and avoiding conflict between senses helps a user retain their sense of presence. For example, if a table looks like wood when visually inspected, but sounds like ringing metal when struck, the user's expectations have been violated and a break in presence is likely. Similarly, a rough surface will visually have a diffuse scattering of light instead of a sharp reflection, and if the user feels that surface with a finger, the roughness they feel should match the roughness they see.

Maintaining a user's expectations about their interactions does not require perfectly realistic virtual environments. As long as the user can establish consistent expectations about the environment, the environment is effective, regardless of how accurate the recreations are compared to the real world. Realism is often desirable for reasons other than presence, such as having training simulations closely match the real thing. Even in these cases, it is not necessary to perfectly recreate reality as long as a user perceives the environment as consistent with their expectations of reality. As long as *perceptual realism* is maintained, the user maintains their sense of presence.

For this proposal, I seek to improve the perceptual realism for the sense of hearing. Specifically for objects in virtual environments, *impact sounds* are common as objects are moved or struck. To create realistic impact sounds, my work uses methods for physically-based impact sound synthesis, which model vibrations in struck objects. When using physically-based methods for sound synthesis, perceptual realism depends on the *material parameters* used as input. Since users have expectations about how virtual objects should sound from their other senses, it is important to use accurate material parameters.

An impact sound's rate of decay largely depends on the material of the struck object: a plastic object has a short-lasting sound while a metal object has a long-lasting sound. Different materials cause different amounts of *damping*, producing different decay rates. *Damping models* are a common approximation that simplify computations by modeling the damping as a function of an object's mass and stiffness. When performing modal sound synthesis, selecting realistic damping rates is important for recreating the sound of the appropriate material.

A common way of creating virtual environments and objects is modeling existing real-world objects. Figure 2 shows examples of real-world objects that have been modeled by hand, but this process can be automated to a limited degree. The shape of an object can be acquired through a 3D scan, and the object's material parameters can be acquired through vision or its impact sounds. These acquired object properties can be used to virtually reconstruct the original object. However, few methods have attempted to combine these two input modalities (vision and impact sounds) in a single coupled process. An object reconstruction method that ensures cohesion between input modalities could better recreate virtual objects with high sensory cohesion.

Evaluation of the methods in this proposal can be performed through comparisons to ground truth or through perceptual evaluation. With comparisons to ground truth, an error metric determines the difference between our results, prior results, and the ground truth. Perceptual evaluation is more frequently used in my work due to the emphasis on ensuring perceptual realism. To evaluate perceptual realism, I rely on user studies with human subjects. In some user studies, subjects report their opinions on the realism, effectiveness, or similarity of real or synthetic virtual objects. In other user studies, subjects must complete tasks using either our methods or those from previous work. User studies directly evaluate the methods with respect to user expectations, providing insight into the methods' performance in an immersive setting.

In order to improve the quality of interaction with virtual objects, the main goals of my research are to:

1. Create a multimodal representation of object surface detail that can be used for multiple modalities of interaction to reduce sensory conflict;

2. Design expressive material damping models and evaluate the perceptual differences between the models when used for impact sound synthesis;

3. Perform audiovisual reconstruction of objects, using both the visual appearance and material properties, and use each modality to validate and improve the other in a coupled multimodal pipeline; and

4. Model damping rates in recorded sounds probabilistically, accounting for and minimizing error in estimates due to stochastic environmental factors.

In the rest of this proposal, I will address the challenges and approaches for each of these goals.

1.1 Thesis Statement

Interaction with objects in virtual environments can made more perceptually realistic by using expressive object material models that account for real-world phenomena and by improving multimodal sensory cohesion.

2 Related Work

In this section, I review work related to the main aspects of this research.

2.1 Sound Synthesis

Sound synthesis techniques recreate natural sounds for virtual environments. Sounds are dynamic and can be created by a variety of sound sources. Different types of sound sources produce different types of sounds, so different models are needed. Examples of sound sources that have been modeled are liquids [19, 24], paper [40], and fire [9].

For this proposal, the focus is on sounds created by vibrating objects. One and two-dimensional objects such as strings and drums create sound which can be simulated using a wavetable and filters [18]. Simple objects with known analytical vibration patterns can be simulated through additive synthesis, where individual sine waves are added together to create more complex sounds [11]. Arbitrary rigid objects use the same additive synthesis method, but to determine their frequencies of vibration, or *modes of vibration*, discretized models of the objects need to be analyzed first [28]. This is referred to as *modal sound synthesis*.

Damping has long been a concern in analysis of vibrations of buildings and other structures [25]. There are a number of ways to model material-based damping to varying degrees of accuracy [50, 43]. Previous sound synthesis work used a simple linear model to determine the damping rate per mode, but existing alternative models may be more accurate for real-world materials [2].

For interactive applications, as a user performs actions to create sounds, sound synthesis algorithms must run fast enough to generate sound in real time. The computation requirements at runtime are proportional to the complexity of the analyzed input shape, making some objects' sounds too slow for our purposes without optimizations. Vibration modes can be culled based on psychoacoustic principles, for example, humans cannot tell the difference between two frequencies very close to one another, so those modes can be combined into one [34]. If an object has any geometric symmetries, these can also be exploited to reduce memory usage and caching requirements [20].

Modal sound synthesis roughly simulates the sounds produced by rigid, vibrating objects, but in the real world more factors influence the final sound we hear, such as acoustic radiance, sound propagation effects, and contacts with other objects. Acoustic radiance is the efficiency of propagation for each mode: depending on the shape of an object some modes radiate in different directions with different strengths [16, 21]. Once the vibrations transfer to the surrounding air, sound waves bounce around the environment before reaching a listener's ears. Sound propagation can be simulated most realistically with wave-based simulation [23], but can be simulated more quickly with geometric methods [10, 39] and can be coupled with modal sound synthesis [37]. Objects also rarely float in midair; usually they rest on a surface or are supported by a hand. These contacts with other objects modify the produced sound and can be accounted for with contact models [52]. Contact modeling can be exploited to create real objects that vibrate only at desired frequencies. An object can be placed on foam blocks, specifically positioned to damp out the undesired frequencies while leaving the desired frequencies alone [4].

2.2 Object Understanding Through Sound

The inverse of the modal sound synthesis problem is to use impact sounds as input to understand the original objects. A common application is to learn properties of a real-world object in order to *resynthesize* similar sounds in a virtual environment. Ideally, sounds from struck real-world objects could be used to recreate the shape and material properties of the objects. While the ideal case of using one sound to reconstruct an entire object is known to be underconstrainted [17], prior research has explored what information can be estimated under different constraints.

Some methods use a single recorded sound, then apply modifications to create realistic variety in resynthesized sounds. Deterministic features of a sound can be extracted, then stochastic noise can be added to those features to model slight variations [42]. Alternatively, the modal content of a sound can be extracted, then resynthesized, slightly modifying mode amplitudes to create variations [22].

Other methods use multiple input sounds for a single object, generated by striking the object in known locations. The sounds' spectral content can be interpolated spatially to approximate hit points at new locations [31]. The Young's modulus for small parts of the object can be individually optimized to best match the input sounds, estimating more fundamental material parameters [51].

By estimating material parameters, such as the Young's modulus, those parameters can be applied to synthesis of sounds for any object with that material. If the exact shape of the struck object is known, material parameters can be estimated from a single recorded impact sound [36].

These methods are often limited in their robustness by using constrained models that do not account for environmental factors or multimodal input. Methods that are capable of estimating material damping parameters only support one material damping model. All methods assume that properties of the recording environment are known or are assumed to be minimal. If both video and audio of the object are available, there are not currently methods available to improve the visual reconstruction of the object using the audio. These limitations are addressed as part of this proposal.

2.3 Multimodal Interaction with Virtual Objects

Multimodal interaction, in the context of this proposal, refers to interaction using multiple senses simultaneously. The senses of sight, hearing, and touch are each different *interaction modalities*. Methods for rendering content for each sense has been independently researched. Since this proposal focuses on audio, rendering for the sense of hearing has been discussed in the earlier Section 2.1.

Realistic visual rendering has been the focus of the computer graphics field for many decades, and photorealistic visual appearances are possible given talented artists and sufficient computational resources. Many books provide an introduction to the field [3, 12]. Creating realistic visual appearances in interactive environments in real time is more challenging, but can be accomplished using optimizations. For example, *texture mapping* uses low-resolution 3D triangle meshes with higher-resolution 2D textures to model detailed objects. Using normal mapping [5], depth mapping [29, 33], and other methods [45], performance can be improved. Examples of texture maps can be seen in Figure 3.

Haptics refers to interaction using the sense of touch, and focuses on the textures of surfaces. Haptic rendering lets a user feel an object's shape and textures by applying forces based on point-contacts with the object [14, 15], given appropriate hardware. Normal maps, though originally designed for visual rendering, can be used for haptic rendering [47].

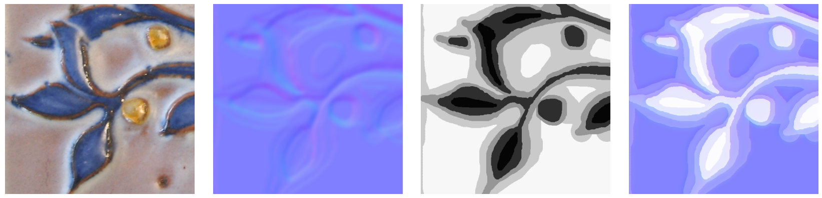

Figure 3: Examples of texture maps used to provide surface detail. RGBA values encode normal vectors and depth information. Normal and relief maps are used as unified representations of surface detail for the senses of sight, hearing, and touch.

For realistic multimodal interaction, it is important that content is not only rendered well for individual senses, but that each sense is consistent with one another. depth maps can modify contacts between objects, coupling the visual appearances of the objects with their physical movements [27]. Modal sound synthesis can be coupled with physics simulations to couple the movements of objects and their resulting sounds [28, 52]. When multiple objects are in contact, the long-lasting contacts produce continuous sounds which depend heavily on the objects' textures, further coupling motion and sound [35]. These methods only involve a few interaction modalities each, and do not use a single representation of surface detail to inform all modalities.

2.4 Visual Object Reconstruction

It is useful to obtain the 3D shape and visual surface texture of a real-world object. For this proposal, I am interested in using this information to create a virtual recreation of the object. This information can be obtained by reconstructing the object from a series of images looking at the object from multiple angles.

Structure from Motion (SFM) [49, 44] and Multi-View Stereo (MVS) [13, 41] are two related classes of techniques for obtaining 3D shape information from a set of 2D images. Bundle adjustment is used to jointly optimize the poses when many images are used as input [48]. Other methods use additional depth information to improve reconstruction [26].

Some objects and materials are difficult to accurately reconstruct. Reflective objects have glare which change in location with the movement of the viewer, while transparent objects make it difficult to determine depth. Additional input modalities may improve results for these difficult objects, for example a time-of-flight camera can correct estimated depth of transparent objects [46].

Impact sound provides an additional input modality, containing cues about the internal structure of an object. Environmental scene classification is a related task approached through spectral analysis [6] or convolutional neural networks [32, 38], but produces broad classifications of an entire environment. However, no current methods use impact sounds in particular to aid in complete shape reconstruction. One goal of the research in this proposal is to use impact sounds to help determine the object shape and material in cases where visual methods struggle.

3 Expected Contributions

In this proposal, I describe projects which improve interaction with virtual objects by addressing limitations of previous work and simplifying virtual replication of real-world objects. The expected contributions of my thesis include:

Multimodal Surface Interaction: We propose a method for using a single texture map as a unified representation of detail for visual rendering, audio rendering, tactile rendering, and physical simulation. Our method runs in real time and allows for multimodal interaction with textured surfaces, while the unified representation ensures cohesion between senses. In task-based user evaluation, our method improves results over alternative, possibly conflicting representations of detail.

Damping Modeling for Modal Sound Synthesis: We present alternative damping models for modal sound synthesis, expanding upon the traditional Rayleigh damping model. These models are able to express

Figure 4: A selection of applications based on our method for multimodal interaction: a virtual environment with a normal mapped surface (left) and a virtual environment with a relief map (right). In both environments, the texture map informs all interaction modalities.

a wider range of damping behaviors. We propose a material parameter estimation method that uses a single input sound to accurately estimate damping parameters for any damping model. Perceptual evaluation demonstrates that no single damping model best represents the damping properties of every material, and multiple damping models should be considered.

Robust Material Parameter Estimation: We propose a novel method for estimating material damping parameters from recorded impact sounds. We use a probabilistic model for observed damping values in recorded sounds, which expressly models multiple external factors affecting the damping values. This method requires no information about the shape of the object or the locations of the impacts, and is able to reduce the effect of external factors to produce more accurate estimates of the material damping parameters in suboptimal recording environments. User evaluation shows that sounds synthesized using our estimated parameters are comparable in realism to those of previous work. Given that our method has significantly fewer requirements about the inputs, our method significantly reduces manual effort needed to obtain high-quality results.

Multimodal Object Reconstruction: We plan to propose a new method for coupled estimation of the visual and audio components of an object. We will use a combination of visual input (2D images or 3D volumetric data) and impact sounds to estimate an object's material and shape. Each of the two multimodal inputs will be able to build on each other for overall higher accuracy. We believe this coupled approach will lead to more accurate reconstructions for both senses, while minimizing sensory conflict during interactions with the object in a virtual environment. Evaluation will be performed through comparisons to ground-truth materials and shape, or by perceptual comparison against reconstructions made by previous work.

4 Research Summary

In this section, I discuss the research projects relevant to my thesis and how they have addressed limitations of previous work to move towards the dissertation goals.

4.1 Integrated Multimodal Interaction Using Texture Representations

As noted in Section 2.3, there have been few efforts to unify interaction in virtual environments across senses. However, they do not clearly consider sensory conflict, nor have any brought together all of sight, touch, hearing, and physical simulation. Sensory conflict is particularly important when considering textured objects, which are often modeled through approximations. In this project, we (myself and Dr. Ming C. Lin) use texture representations of detail, particularly normal and relief maps, as a unified source of information for all interaction modes.

Haptic rendering, sound rendering, and rigid-body simulation with textured surfaces have been independently explored [30, 35], but have not been integrated together cohesively. For example, a previous method for sound rendering of contacts with textured surfaces [35] displays a pen sliding smoothly across highly bumpy surfaces. While the generated sound from this interaction is dynamic and realistic, the smooth *visual* movement of the pen does not match the texture implied by the sound. In order to minimize sensory conflict, it is critical to present a unified and seamlessly integrated multimodal display to users, ensuring rendering is consistent across the senses of sight, hearing, and touch.

In a virtual environment, a user may see a rough, bumpy surface represented by its texture equivalent mapped to a flat surface. In the real world, objects behave differently when bouncing, sliding, or rolling on bumpy or rough surfaces than they do on flat surfaces. The visually complex detail would contrast with simple physical behavior due to the flat surface, causing sensory conflict and breaking the sense of immersion. In order to model such physical behavior, the shape geometry used in a physics simulator would require a fine triangle mesh with sufficient surface detail, but in most cases a sufficiently fine mesh is unavailable or would require prohibitive amounts of memory. Since texture maps contain information about the fine detail of a mapped surface, it is possible to use that information to recreate the physical behavior of the fine triangle mesh.

To accomplish this, we propose a new effective method for simulation of physical behaviors for rigid objects textured with normal maps. We also propose seamlessly integrated multisensory interaction using normal and relief maps to improve sensory cohesion. By using a single representation of surface detail for each interaction modality, we ensure sensory cohesion for users. See Figure 4 for examples of interaction with textured surfaces. A virtual pen is controlled through a haptic device, allowing the user to interact with the environment while feeling forces in response. A simulated ball rolls on the surface, its motions affected by both the surface texture and the pen. Contacts between the pen, ball, and surface create physically-based sound, bringing together sight, hearing, touch, and physical simulation.

We evaluate the method through perceptual user studies. In these studies, subjects are asked to identify the surface displayed to them, but in some trials certain interaction modalities are removed. When all modalities are present using our method, performance on the task is at its highest, demonstrating that the senses are not in conflict with one another. More results are available on the project website: `http://gamma.cs.unc.edu/MultiDispTexture/`.

4.2 Interactive Modal Sound Synthesis Using Generalized Proportional Damping

In order to create higher quality modal sound, the next project evaluates improved methods for damping modeling. For this project, we (myself and Dr. Ming C. Lin) consider more expressive models for the damping properties of objects. Since the damping properties of a sound tell us about an object's material, more realistic damping modeling in synthesized sound should be better able to recreate the sounds of real-world materials.

Damping is a complex phenomenon to model, and it can be difficult to determine exactly how vibrations in an object decay. Additionally, the presence of damping may give rise to *complex* modes of vibration, which are more difficult to model than *normal* modes [8]. In practice, approximations are used to produce simpler models. The most common approach is to assume all damping is viscous and to approximate the decay rate of a material as a linear combination of its density and stiffness. This model is referred to as Rayleigh (or linearly proportional) damping, and only produces normal modes. It is the de-facto technique for modeling damping for modal sound synthesis, but has always been understood to be an approximation for convenience.

Other damping models are common in material and structural analysis, but have not been thoroughly examined for interactive sound synthesis. Caughey damping is a polynomial extension of the linear Rayleigh damping model [7, 8]. The most general damping model to date that limits vibrations to normal modes is *generalized proportional damping* (GPD) [1]. These alternative damping models may be able to improve sound quality by providing a better fit to observed real-world damping.

In this project, we explore the use of alternative damping models for sound synthesis. These damping models are more expressive, and are able to model damping behavior that would be only coarsely approximated by the Rayleigh damping model. We describe how more expressive damping models can be integrated

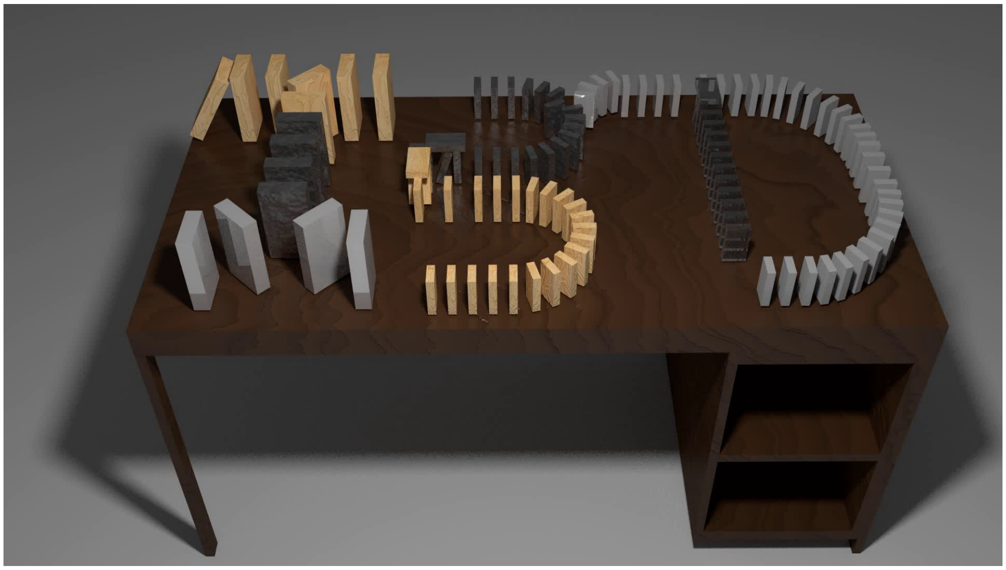

Figure 5: A scenario with dominoes made of different materials. Each material uses a set of material parameters estimated from recorded impact sounds, including parameters for a damping model.

into current methods for modal sound synthesis, and propose specific damping models within the larger space of GPD functions. We generalize a prior optimization method previously for damping parameter estimation from recorded impact sounds to estimate parameters for any arbitrary damping model. We also conduct a user study to evaluate the perceptual differences between multiple damping models. Figure 5 shows one possible scenario with objects creating sound based on their materials estimated from recorded impact sounds on real objects. Results, including audio samples and videos, can be found on the project website: `http://gamma.cs.unc.edu/gpdsynth/`.

4.3 Audio-Material Reconstruction for Virtualized Reality using a Probabilistic Damping Mode

In order to create realistic virtual objects, recorded impact sounds can be used to estimate damping parameters, even if the sounds are recorded in noisy and uncontrolled settings. This project explores a novel probabilistic damping model for estimating the damping parameters of an object's material from a set of impact sounds. The researchers on this project are myself, Nicholas Rewkowski, and Ming C. Lin.

While recent methods have been able to estimate material damping properties [36], they assume all observed vibrational decay is due to the material damping and not any other source. However, there are multiple external factors that would produce effects similar to damping.

In the real world, an object struck for the purposes of recording either needs to be held by hand or left to rest on another surface. In either case, the interface between the object and its *support* introduces additional damping, as energy is transferred from the vibrating object to the more stationary support. See Figure 6 for an example: depending on how the bowl is held, it produces dramatically different sound.

Other factors include complex modes of vibration, room acoustics, and error in the feature extraction step. Standard damping models capture only normal modes of vibration, not complex modes of vibration. These models are approximations, as real objects always have some amount of complex vibrations, causing error in damping estimates. If impact sound recording is performed in an enclosed room, the room acoustics—reflections off walls—extend the length of sounds. Feature extraction steps are common in most damping parameter estimation methods, but even with clean input these steps often introduce their own error.

In realistic and uncontrolled environments with significant impact from external factors, the parameters estimated by current methods are not truly material parameters. Instead, they are parameters modeling *both* the material and the environment used for the recording and thus do not generalize to arbitrary environments.

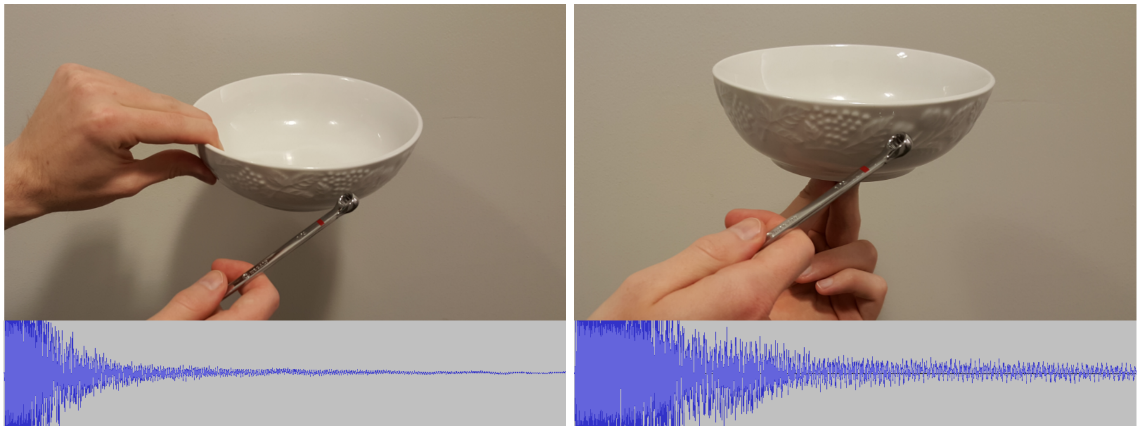

Figure 6: A porcelain bowl struck in the same location produces different sound when supported with a tight grip (left) or supported by resting on a single point (right). Methods for material damping parameter estimation should be robust to these external damping factors.

For this project, we propose a practical and efficient method to estimate material damping properties from recorded impact sounds, while accounting for the additional damping due to the recording environment. We specifically model the external damping factors in a probabilistic damping model. For a given mode of vibration with an estimated frequency, this model provides a probability distribution for possible observable damping values. When multiple impact sounds are provided as input, the damping model's parameters can be optimized to fit the real-world sounds, providing an estimate of the material's damping parameters. As a result, our model can express stochastic properties of the recording environment, reducing their effect on the estimated parameters.

Our method is more applicable to real-world recordings in less controlled environments. The method is fast, requires no prior knowledge about the recorded object, and can use multiple recordings to improve accuracy. Figure 7 shows a visual representation of the material damping models as estimated from the real-world sound features shown as points. In the absence of external damping factors, all of these points would fall along one line representing the damping from the material alone. However, the external factors cause the points to vary, throwing off a more traditional least squares (LSQ) approach. This project is nearing completion, but a complete project website is not yet available.

4.4 Audiovisual Object/Scene Reconstruction (In Progress)

This project aims to combine aspects of material estimation with aspects of multimodal interaction. The researchers working on this project are myself, Justin Wilson, Sam Lowe, and Ming C. Lin. Given a video showing an object from many angles, a 3D model of that object can be reconstructed. Given impact sounds from an object, material parameters for that object can be estimated. While these have been accomplished individually, better results may be obtainable if these two processes are coupled.

Reconstructed shape can aid in material estimation. Using the current state of the art, density and Young's modulus can only be estimated if the shape is known. Without knowledge of the scale of an object or the sizes of its geometric features, there is ambiguity between material and size.

Similarly, recorded impact sounds can aid the visual reconstruction. Objects that are overly reflective, transparent, or seen under variable lighting conditions may have poor reconstructions from visual data alone. Impact sounds provide an additional modality that is invariant to visual appearance, providing information about interior or occluded parts of an object.

We seek to use these two modalities together to create a system for multimodal object reconstruction. Used together, these two input modalities could cover each others' weaknesses. The primary goal of this project is to use sound to improve object reconstruction and segmentation for highly-reflective or transparent objects. This should lead to a greater semantic understanding of a scene containing rigid objects.

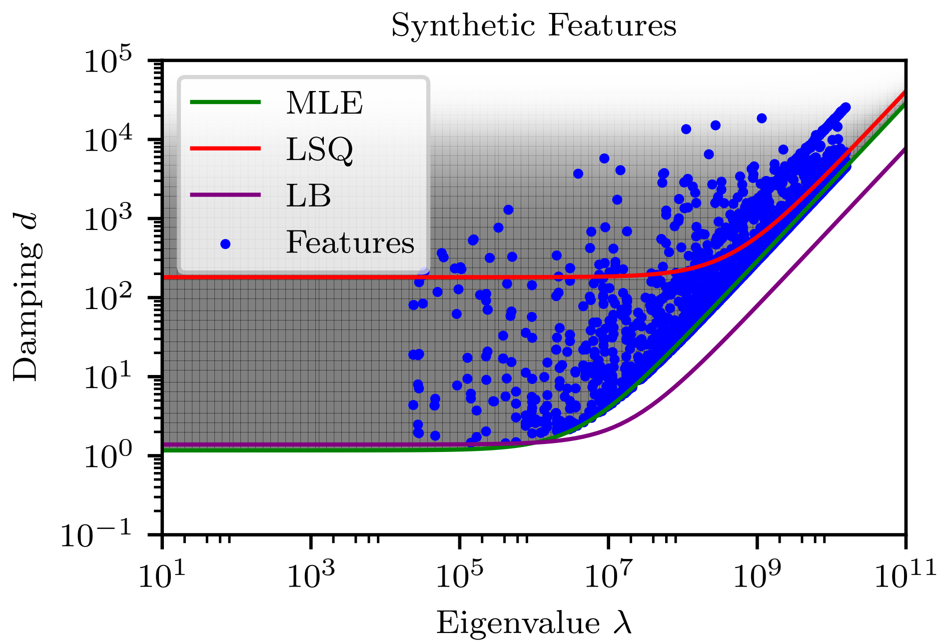

Figure 7: Parameter estimation on sound features. Each feature is one extracted mode of vibration, consisting of an eigenvalue λ_i (approximately the square of its frequency) and its corresponding damping coefficient d_i. Estimated Rayleigh damping curves are plotted, with the variation from the curve caused by external factors. Our method is labeled MLE, and provides the ideal fit to the sound features

5 Dissertation Schedule

5.1 Current Progress

- Fall 2014
  - Submitted "Integrated Multimodal Interaction Using Normal Maps" to Graphics Interface 2015, accepted.
- Spring 2015
  - Prepared submissions for Fall.
- Fall 2015
  - Submitted "Integrated Multimodal Interaction Using Texture Representations" to Computers and Graphics, accepted.
  - Submitted "Interactive Modal Sound Synthesis Using Generalized Proportional Damping" to Symposium on Interactive 3D Graphics and Games (I3D) 2016, accepted.
  - Presented for PRP, passed.
- Spring 2016
  - Researched support damping and modeling.
- Summer 2016
  - Internship at Aurora Flight Sciences. Worked on machine learning methods for estimating and predicting stages of aircraft flight from sensor data.

- Fall 2016
  - Submitted "Audio-Material Reconstruction for Virtualized Reality using a Probabilistic Damping Model" to Eurographics 2017.
  - Presented "Material Damping Parameter Estimation From Audio" at the meeting of the North Carolina chapter of the Acoustical Society of America.
- Spring 2017
  - Resubmitted "Audio-Material Reconstruction for Virtualized Reality using a Probabilistic Damping Model" to SIGGRAPH 2017.
  - Resubmitted "Audio-Material Reconstruction for Virtualized Reality using a Probabilistic Damping Model" to SIGGRAPH Asia 2017.
- Summer 2017
  - Taught COMP 116: Introduction to Scientific Programming.
  - Researching audiovisual reconstruction and coupled optimization.
- Fall 2017
  - Resubmitted "Audio-Material Reconstruction for Virtualized Reality using a Probabilistic Damping Model" to IEEE VR 2018.
  - Submitted proposal.
  - Submitted "ISNN: Impact Sound Neural Network for Audio-Visual Object Classification" to CVPR 2018.

5.2 Future Plans

- Spring 2018 (current)
  - Resubmitted "ISNN: Impact Sound Neural Network for Audio-Visual Object Classification" to ECCV 2018.
  - Resubmit "Audio-Material Reconstruction for Virtualized Reality using a Probabilistic Damping Model" to SIGGRAPH Asia 2018.
  - Work on dissertation.
- Fall 2018
  - Oral exam.
  - Finalize dissertation.
  - Defend dissertation.

References

[1] S. Adhikari. "Damping modelling using generalized proportional damping". In: *Journal of Sound and Vibration* 293.1-2 (2006), pp. 156–170. ISSN: 0022-460X. DOI: http://dx.doi.org/10.1016/j.jsv.2005.09.034. URL: http://www.sciencedirect.com/science/article/pii/S0022460X05006942.

[2] Sondipon Adhikari. "Damping models for structural vibration". PhD thesis. University of Cambridge, 2001.

[3] Tomas Akenine-Moller, Tomas Moller, and Eric Haines. *Real-Time Rendering*. 2nd. Natick, MA, USA: A. K. Peters, Ltd., 2002. ISBN: 1568811829.

[4] Gaurav Bharaj et al. "Computational Design of Metallophone Contact Sounds". In: *ACM Trans. Graph.* 34.6 (Oct. 2015), 223:1–223:13. ISSN: 0730-0301. DOI: 10.1145/2816795.2818108. URL: http://doi.acm.org/10.1145/2816795.2818108.

[5] James F. Blinn. "Simulation of Wrinkled Surfaces". In: *SIGGRAPH Comput. Graph.* 12.3 (Aug. 1978), pp. 286–292. ISSN: 0097-8930. DOI: 10.1145/965139.507101. URL: http://doi.acm.org/10.1145/965139.507101.

[6] Michael Büchler et al. "Sound Classification in Hearing Aids Inspired by Auditory Scene Analysis". In: *EURASIP Journal on Advances in Signal Processing* 2005.18 (Nov. 2005), p. 387845. ISSN: 1687-6180. DOI: 10.1155/ASP.2005.2991. URL: https://doi.org/10.1155/ASP.2005.2991.

[7] TK Caughey. "Classical normal modes in damped linear dynamic systems". In: *Journal of Applied Mechanics* 27.2 (1960), pp. 269–271.

[8] TK Caughey and MEJ O'Kelly. "Classical normal modes in damped linear dynamic systems". In: *Journal of Applied Mechanics* 32.3 (1965), pp. 583–588.

[9] Jeffrey N. Chadwick and Doug L. James. "Animating Fire with Sound". In: *ACM SIGGRAPH 2011 Papers*. SIGGRAPH '11. Vancouver, British Columbia, Canada: ACM, 2011, 84:1–84:8. ISBN: 978-1-4503-0943-1. DOI: 10.1145/1964921.1964979. URL: http://doi.acm.org/10.1145/1964921.1964979.

[10] Anish Chandak et al. "AD-Frustum: Adaptive Frustum Tracing for Interactive Sound Propagation". In: *IEEE Transactions on Visualization and Computer Graphics* 14.6 (Nov. 2008), pp. 1707–1722. ISSN: 1077-2626. DOI: 10.1109/TVCG.2008.111. URL: http://dx.doi.org/10.1109/TVCG.2008.111.

[11] Kees van den Doel and Dinesh K. Pai. "The Sounds of Physical Shapes". In: *Presence* 7 (1996), pp. 382–395.

[12] James D. Foley et al. *Computer Graphics: Principles and Practice (2Nd Ed.)* Boston, MA, USA: Addison-Wesley Longman Publishing Co., Inc., 1990. ISBN: 0-201-12110-7.

[13] M. Goesele et al. "Multi-View Stereo for Community Photo Collections". In: *2007 IEEE 11th International Conference on Computer Vision*. Oct. 2007, pp. 1–8. DOI: 10.1109/ICCV.2007.4408933.

[14] Chih-Hao Ho, C Basdogan, and MA Srinivasan. "A ray-based haptic rendering technique for displaying shape and texture of 3D objects in virtual environments". In: *ASME Winter Annual Meeting*. 1997.

[15] Chih-Hao Ho, Cagatay Basdogan, and Mandayam A. Srinivasan. "Efficient Point-Based Rendering Techniques for Haptic Display of Virtual Objects". In: *Presence: Teleoper. Virtual Environ.* 8.5 (Oct. 1999), pp. 477–491. ISSN: 1054-7460. DOI: 10.1162/105474699566413. URL: http://dx.doi.org/10.1162/105474699566413.

[16] Doug L James, Jernej Barbič, and Dinesh K Pai. "Precomputed acoustic transfer: output-sensitive, accurate sound generation for geometrically complex vibration sources". In: *ACM Transactions on Graphics (TOG)*. Vol. 25. 3. ACM. 2006, pp. 987–995.

[17] Mark Kac. "Can One Hear the Shape of a Drum?" In: *The American Mathematical Monthly* 73.4 (1966), pp. 1–23. ISSN: 00029890, 19300972. URL: http://www.jstor.org/stable/2313748.

[18] Kevin Karplus and Alex Strong. "Digital Synthesis of Plucked-String and Drum Timbres". In: *Computer Music Journal* 7.2 (1983), pp. 43–55. ISSN: 01489267, 15315169. URL: http://www.jstor.org/stable/3680062.

[19] Timothy R. Langlois, Changxi Zheng, and Doug L. James. "Toward Animating Water with Complex Acoustic Bubbles". In: *ACM Trans. Graph.* 35.4 (July 2016), 95:1–95:13. ISSN: 0730-0301. DOI: 10.1145/2897824.2925904. URL: http://doi.acm.org/10.1145/2897824.2925904.

[20] Timothy R. Langlois et al. "Eigenmode Compression for Modal Sound Models". In: *ACM Transactions on Graphics (Proceedings of SIGGRAPH 2014)* 33.4 (Aug. 2014). DOI: 10.1145/2601097.2601177. URL: http://www.cs.cornell.edu/projects/Sound/modec.

[21] Dingzeyu Li, Yun Fei, and Changxi Zheng. "Interactive Acoustic Transfer Approximation for Modal Sound". In: *ACM Trans. Graph.* 35.1 (2015). DOI: 10.1145/2820612.

[22] D. Brandon Lloyd, Nikunj Raghuvanshi, and Naga K. Govindaraju. "Sound Synthesis for Impact Sounds in Video Games". In: *Proceedings of the Symposium on Interactive 3D Graphics and Games 2011*. ACM, 2011. URL: http://research.microsoft.com/apps/pubs/default.aspx?id=146525.

[23] R. Mehra et al. "WAVE: Interactive Wave-based Sound Propagation for Virtual Environments". In: *IEEE Transactions on Visualization and Computer Graphics* 21.4 (Apr. 2015), pp. 434–442. ISSN: 1077-2626. DOI: 10.1109/TVCG.2015.2391858.

[24] William Moss et al. "Sounding Liquids: Automatic Sound Synthesis from Fluid Simulation". In: *ACM Trans. Graph.* 29.3 (July 2010), 21:1–21:13. ISSN: 0730-0301. DOI: 10.1145/1805964.1805965. URL: http://doi.acm.org/10.1145/1805964.1805965.

[25] Ahid D Nashif, David IG Jones, and John P Henderson. *Vibration damping*. John Wiley & Sons, 1985.

[26] Richard A. Newcombe et al. "KinectFusion: Real-Time Dense Surface Mapping and Tracking". In: *International Symposium on Mixed and Augmented Reality (ISMAR)* (2011).

[27] Scott Nykl, Chad Mourning, and David Chelberg. "Interactive Mesostructures". In: *Proceedings of the ACM SIGGRAPH Symposium on Interactive 3D Graphics and Games*. I3D '13. Orlando, Florida: ACM, 2013, pp. 37–44. ISBN: 978-1-4503-1956-0. DOI: 10.1145/2448196.2448202. URL: http://doi.acm.org/10.1145/2448196.2448202.

[28] James F. O'Brien, Chen Shen, and Christine M. Gatchalian. "Synthesizing Sounds from Rigid-body Simulations". In: *Proceedings of the 2002 ACM SIGGRAPH/Eurographics Symposium on Computer Animation*. SCA '02. San Antonio, Texas: ACM, 2002, pp. 175–181. ISBN: 1-58113-573-4. DOI: 10.1145/545261.545290. URL: http://doi.acm.org/10.1145/545261.545290.

[29] Manuel M. Oliveira, Gary Bishop, and David McAllister. "Relief Texture Mapping". In: *Proceedings of the 27th Annual Conference on Computer Graphics and Interactive Techniques*. SIGGRAPH '00. New York, NY, USA: ACM Press/Addison-Wesley Publishing Co., 2000, pp. 359–368. ISBN: 1-58113-208-5. DOI: 10.1145/344779.344947. URL: http://dx.doi.org/10.1145/344779.344947.

[30] M.A. Otaduy et al. "Haptic display of interaction between textured models". In: *IEEE Visualization Conference*. Oct. 2004, pp. 297–304. DOI: 10.1109/VISUAL.2004.37.

[31] Dinesh K Pai et al. "Scanning physical interaction behavior of 3D objects". In: *Proceedings of the 28th annual conference on Computer graphics and interactive techniques*. ACM. 2001, pp. 87–96.

[32] K. J. Piczak. "Environmental sound classification with convolutional neural networks". In: *2015 IEEE 25th International Workshop on Machine Learning for Signal Processing (MLSP)*. Sept. 2015, pp. 1–6. DOI: 10.1109/MLSP.2015.7324337.

[33] Fábio Policarpo, Manuel M. Oliveira, and João L. D. Comba. "Real-time Relief Mapping on Arbitrary Polygonal Surfaces". In: *Proceedings of the 2005 Symposium on Interactive 3D Graphics and Games*. I3D '05. Washington, District of Columbia: ACM, 2005, pp. 155–162. ISBN: 1-59593-013-2. DOI: 10.1145/1053427.1053453. URL: http://doi.acm.org/10.1145/1053427.1053453.

[34] Nikunj Raghuvanshi and Ming C. Lin. "Interactive Sound Synthesis for Large Scale Environments". In: *Proceedings of the 2006 Symposium on Interactive 3D Graphics and Games*. I3D '06. Redwood City, California: ACM, 2006, pp. 101–108. ISBN: 1-59593-295-X. DOI: 10.1145/1111411.1111429. URL: http://doi.acm.org/10.1145/1111411.1111429.

[35] Zhimin Ren, Hengchin Yeh, and M.C. Lin. "Synthesizing contact sounds between textured models". In: *Virtual Reality Conference (VR), 2010 IEEE*. Mar. 2010, pp. 139–146. DOI: 10.1109/VR.2010.5444799.

[36] Zhimin Ren, Hengchin Yeh, and Ming C. Lin. "Example-guided Physically Based Modal Sound Synthesis". In: *ACM Trans. Graph.* 32.1 (Feb. 2013), 1:1–1:16. ISSN: 0730-0301. DOI: 10.1145/2421636.2421637. URL: http://doi.acm.org/10.1145/2421636.2421637.

[37] A. Rungta et al. "SynCoPation: Interactive Synthesis-Coupled Sound Propagation". In: *IEEE Transactions on Visualization and Computer Graphics* 22.4 (Apr. 2016), pp. 1346–1355. ISSN: 1077-2626. DOI: 10.1109/TVCG.2016.2518421.

[38] J. Salamon and J. P. Bello. "Deep Convolutional Neural Networks and Data Augmentation for Environmental Sound Classification". In: *IEEE Signal Processing Letters* 24.3 (Mar. 2017), pp. 279–283. ISSN: 1070-9908. DOI: 10.1109/LSP.2017.2657381.

[39] Carl Schissler and Dinesh Manocha. "Adaptive Impulse Response Modeling for Interactive Sound Propagation". In: *Proceedings of the 20th ACM SIGGRAPH Symposium on Interactive 3D Graphics and Games*. I3D '16. Redmond, Washington: ACM, 2016, pp. 71–78. ISBN: 978-1-4503-4043-4. DOI: 10.1145/2856400.2856414. URL: http://doi.acm.org/10.1145/2856400.2856414.

[40] Camille Schreck et al. "Real-time Sound Synthesis for Paper Material Based on Geometric Analysis". In: *Proceedings of the ACM SIGGRAPH/Eurographics Symposium on Computer Animation*. SCA '16. Zurich, Switzerland: Eurographics Association, 2016, pp. 211–220. ISBN: 978-3-905674-61-3. URL: http://dl.acm.org/citation.cfm?id=2982818.2982847.

[41] S. M. Seitz et al. "A Comparison and Evaluation of Multi-View Stereo Reconstruction Algorithms". In: *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)*. Vol. 1. June 2006, pp. 519–528. DOI: 10.1109/CVPR.2006.19.

[42] Xavier Serra and Julius Smith. "Spectral Modeling Synthesis: A Sound Analysis/Synthesis System Based on a Deterministic Plus Stochastic Decomposition". In: *Computer Music Journal* 14.4 (1990), pp. 12–24. ISSN: 01489267, 15315169. URL: http://www.jstor.org/stable/3680788.

[43] Joseph C Slater, W Keith Belvin, and Daniel J Inman. "A survey of modern methods for modeling frequency dependent damping in finite element models". In: *PROCEEDINGS-SPIE THE INTERNATIONAL SOCIETY FOR OPTICAL ENGINEERING*. SPIE INTERNATIONAL SOCIETY FOR OPTICAL. 1993, pp. 1508–1508.

[44] Noah Snavely, Steven M. Seitz, and Richard Szeliski. "Photo Tourism: Exploring Photo Collections in 3D". In: *ACM Trans. Graph.* 25.3 (July 2006), pp. 835–846. ISSN: 0730-0301. DOI: 10.1145/1141911.1141964. URL: http://doi.acm.org/10.1145/1141911.1141964.

[45] L. Szirmay-Kalos and T. Umenhoffer. "Displacement Mapping on the GPU - State of the Art". In: *Computer Graphics Forum* 27.1 (2008).

[46] Kenichiro Tanaka et al. "Material Classification using Frequency- and Depth-dependent Time-of-Flight Distortion". In: *Computer Vision and Pattern Recognition (CVPR), 2017 IEEE Conference on*. July 2017, pp. 79–88.

[47] Víctor Theoktisto, Marta Fairén González, and Isabel Navazo. "Hybrid Rugosity Mesostructures (HRMs) for fast and accurate rendering of fine haptic detail". In: *CLEI Electron. J.* (2010), pp. -1–1.

[48] Bill Triggs et al. "Bundle Adjustment - A Modern Synthesis". In: *Proceedings of the International Workshop on Vision Algorithms: Theory and Practice*. ICCV '99. London, UK, UK: Springer-Verlag, 2000, pp. 298–372. ISBN: 3-540-67973-1. URL: http://dl.acm.org/citation.cfm?id=646271.685629.

[49] M.J. Westoby et al. "'Structure-from-Motion' photogrammetry: A low-cost, effective tool for geoscience applications". In: *Geomorphology* 179 (2012), pp. 300–314. ISSN: 0169-555X. DOI: http://dx.doi.org/10.1016/j.geomorph.2012.08.021. URL: http://www.sciencedirect.com/science/article/pii/S0169555X12004217.

[50] J Woodhouse. "Linear damping models for structural vibration". In: *Journal of Sound and Vibration* 215.3 (1998), pp. 547–569.

[51] Kazuhiko Yamamoto and Takeo Igarashi. "Interactive Physically-based Sound Design of 3D Model Using Material Optimization". In: *Proceedings of the ACM SIGGRAPH/Eurographics Symposium on Computer Animation*. SCA '16. Zurich, Switzerland: Eurographics Association, 2016, pp. 231–240. ISBN: 978-3-905674-61-3. URL: http://dl.acm.org/citation.cfm?id=2982818.2982849.

[52] Changxi Zheng and Doug L. James. "Toward High-Quality Modal Contact Sound". In: *ACM Transactions on Graphics (Proceedings of SIGGRAPH 2011)* 30.4 (Aug. 2011). URL: http://www.cs.cornell.edu/projects/Sound/mc.