Interactive Modal Sound Synthesis Using Generalized Proportional Damping*

Auston Sterling† Ming C. Lin‡

Abstract

We present a modal sound synthesis technique using a *generalized proportional damping* (GPD) model capable of capturing nonlinear frequency-dependent damping functions. We extend a prior method for automatic extraction of audio material parameters directly from recorded audio clips to determine material parameters for alternative damping models. We demonstrate the results with example-guided synthesized sounds, accompanied by a preliminary, perceptual study comparing the audio quality of the commonly used, linear Rayleigh damping model against a collection of alternative models.

Keywords: Material Damping Models, Sound Rendering

Concepts: •**Applied computing** → **Sound and music computing;**

1 Introduction

Engaging multiple senses in a virtual environment (VE) or an interactive 3D application is critical for an immersive user experience. Realistic sound corresponding to the visuals of a scene can considerably improve the quality of interaction and enhance the sense of presence in a virtual world. Ideally, physical behavior in a VE would dynamically create the corresponding auditory feedback. One area of particular interest is the simulation of sounds created by vibrating rigid bodies. These audio cues may include the types of sounds created by knocking on a door, rolling dice, ringing a bell, or dropping a spatula. A common technique for reproducing these sounds is to analyze the vibrations of the sounding object using modal analysis, then dynamically create new sounds at runtime with modal synthesis. Modal analysis requires as parameters a model of the rigid object and a set of material properties. These material properties are tedious to set by hand, but determine whether the object sounds like glass, metal, or another material.

One aspect of sound synthesis is the damping model, which characterizes how the amplitude of the sound decays over time. Damping is a complex phenomenon, and it can be difficult to determine exactly how the vibrations of a modeled object will decay. Additionally, the presence of damping may give rise to *complex* modes of vibration, which are more difficult to model than *normal* modes [Caughey and Okelly 1965]. The most common approach is to assume all damping is viscous and to approximate the decay rate of one part of an object as a linear combination of its mass and stiffness. This model is referred to as Rayleigh damping or linearly proportional damping, and only produces normal modes. It is the de-facto technique for modeling damping using modal sound synthesis. Rayleigh damping uses a simple linear model, but there are few known limitations about the damping properties of synthesized sound using properly set Rayleigh damping coefficients.

The limitations are: (1) Rayleigh damping is only a first-order approximation and (2) it was originally chosen for its ease of computation, not its physical accuracy. Other damping models are common in material and structural analysis, but have not been thoroughly examined in computer graphics for interactive 3D sound synthesis. The most general damping model to date that limits vibrations to normal modes is *generalized proportional damping* (GPD) [Adhikari 2006], of which Rayleigh damping is a special case. These alternative damping models may be able to improve sound quality by providing a better fit to the real-world damping behavior. By improving the quality of synthesized sound, we can enhance the immersion in virtual environments to create more effective 3D games, telepresence applications, and training simulations.

In this paper, we explore the use of generalized proportional damping for interactive modal sound synthesis. We first describe how GPD can be integrated into current methods for modal sound synthesis. We also propose specific damping models within the larger space of GPD functions that may be of interest for modal sound synthesis. We further extend an optimization framework originally designed to compute Rayleigh damping parameters given audio samples to compute material parameters for the GPD model. Finally, we conduct a preliminary user study to evaluate the perceptual differences between multiple damping models in modal sound synthesis.

To sum up, the main results of this paper include:

- Investigation of higher-order generalized damping models for modal sound synthesis (Section 3);
- Estimation of material parameters for Generalized Proportional Damping in sound rendering (Section 4); and
- Evaluation, comparison, and analysis of percevied audio quality using these GPD models (Section 5).

We'll briefly survey some of the related work in modal sound synthesis and audio parameter identification.

2 Previous Work

Modal analysis has historically been used for mostly engineering applications, but has also been applied for synthesizing sound from shapes with analytically-computed eigen-systems [van den Doel and Pai 1996]. This approach was later extended to use numerical methods for computing eigen-systems of objects with arbitrary shapes and setting the precedence of using Rayleigh damping for sound synthesis [O'Brien et al. 2002]. Real-time synthesis for many objects simultaneously is enabled by optimization based on psychoacoustic principles [Raghuvanshi and Lin 2006] or by performing synthesis in frequency space [Bonneel et al. 2008]. While this paper focuses purely on modal sound synthesis, it is worthwhile to mention work done on coupling synthesis and propagation [James et al. 2006] and work incorporating sound from sources other than modal free-vibration sounds [Chadwick et al. 2012],

*http://gamma.cs.unc.edu/gpdsynth

†e-mail: austonst@cs.unc.edu

‡e-mail: lin@cs.unc.edu

I3D 2016, February 27–28, 2016, Redmond, WA

ISBN: 978-1-4503-4043-4/16/03

DOI: http://dx.doi.org/10.1145/2856400.2856419

Damping models are widely used in creating virtual sound, but they were originally proposed to determine the damping properties of materials for structural analysis [Adhikari and Woodhouse 2001]. Rayleigh damping is the oldest and still the most popular damping model used in damping analysis and in sound synthesis [Rayleigh 1896], due to its simplicity and ease of implementation. Necessary and sufficient conditions for normal modes have been discovered [Caughey 1960] and damping models have been formed around these concepts [Caughey and Okelly 1965].

Real-world audio recordings have been used to guide sound synthesis. Older techniques are able to reproduce the sound of a specific object given many recordings representing a complete *audio sampling* of a given object [Pai et al. 2001]. More recent results involve obtaining an object's modal information from sound recordings [Lloyd et al. 2011] and automatic extraction of Rayleigh damping parameters from a single audio sample [Ren et al. 2013b]. In all these cases, Rayleigh damping is used to model losses, where the viscous damping is represented as a linear combination of the mass and stiffness.

3 Modal Sound Synthesis

To generate realistic, physically-based sound, we use modal analysis for sound synthesis. In this section, we describe these methods and associated damping models. We also describe how generalized proportional damping can be integrated into such a system.

3.1 Modal Analysis

When a rigid object is struck, it vibrates in response, though these vibrations may be imperceptible to the eye. As the surface of the object vibrates and deforms, the surrounding air is rapidly compressed and expanded, creating pressure waves which propagate through the environment. Our ears perceive the variation in air pressure as sound. In modal analysis, the shape and material properties of the object are analyzed to decompose the vibrations into a set of modes of vibration. Each normal mode of vibration describes one way in which the object could deform sinusoidally over time. Vibrations from an impact can roughly be represented as a linear combination of normal modes with different amplitudes, frequencies, and phases.

Nowadays, modal analysis is performed numerically, where the object is represented using a discretized model such as a FEM mesh or spring-mass thin-shell system. Regardless of the choice of discretization, we can consider the dynamics of the system as it vibrates using a system of equations:

$$\mathbf{M}\ddot{\mathbf{r}} + \mathbf{C}\dot{\mathbf{r}} + \mathbf{K}\mathbf{r} = \mathbf{f} \tag{1}$$

Here, $\mathbf{r}$ is a vector of the displacement of each element, where a vector of all zeros represents the object at rest. Since we usually work with three-dimensional objects, an object with n discrete elements would have a $\mathbf{r}$ vector of size $3n$. $\mathbf{f}$ is the vector of forces applied to each element, inducing vibrations. $\mathbf{M}$ is the mass matrix, which describes where mass is located on the object. $\mathbf{C}$ is the viscous damping matrix, which describes how the velocity of the elements $\dot{\mathbf{r}}$ decays over time. $\mathbf{K}$ is the stiffness matrix, in which the connectivity of the elements is defined. Given these matrices, we can properly simulate the vibration of the object in response to an impulse. $\mathbf{M}$ and $\mathbf{K}$ can be constructed through knowledge of the shape of the object and its material properties, notably its density, Poisson's ratio, and Young's modulus. The damping matrix $\mathbf{C}$, is not as simple to construct, and is the focus of this paper.

Modal analysis examines the eigenvalues and eigenvectors of the system in free vibration, that is, with $f = 0$ after some initial impulse has been applied. Temporarily ignoring damping, we can set up a generalized eigenvalue problem of the form:

$$\mathbf{K}\mathbf{v} = \lambda\mathbf{M}\mathbf{v} \tag{2}$$

Finding this eigendecomposition and combining the eigenvectors into a matrix $\mathbf{\Phi}$ allows the matrices $\mathbf{M}$ and $\mathbf{K}$ to be diagonalized. Specifically, the eigenvectors are mass-normalized such that:

$$\mathbf{\Phi}^T\mathbf{M}\mathbf{\Phi} = \mathbf{I} \quad \text{and} \quad \mathbf{\Phi}^T\mathbf{K}\mathbf{\Phi} = \mathbf{\Omega^2} \tag{3}$$

The matrix $\mathbf{\Phi}$ can be intuitively described as a matrix which transforms between object space and mode space: each column of $\mathbf{\Phi}$ contains the shape of a normal mode, while $\mathbf{\Phi}^T\mathbf{f}$ converts forces on elements to normal mode amplitudes. The natural undamped frequencies of the system are contained in the diagonal matrix $\mathbf{\Omega}$, while their squares in $\mathbf{\Omega}^2$ are the eigenvalues of the system. We can continue the decoupling by considering the system in mode space $\mathbf{z} = \mathbf{\Phi}^T\mathbf{M}\mathbf{r}$:

$$\mathbf{\Phi^T}\mathbf{M}\mathbf{\Phi}\ddot{\mathbf{z}} + \mathbf{\Phi^T}\mathbf{C}\mathbf{\Phi}\dot{\mathbf{z}} + \mathbf{\Phi^T}\mathbf{K}\mathbf{\Phi}\mathbf{z} = \mathbf{\Phi^T}\mathbf{f} \tag{4}$$

$$\ddot{\mathbf{z}} + \mathbf{\Phi^T}\mathbf{C}\mathbf{\Phi}\dot{\mathbf{z}} + \mathbf{\Omega^2}\mathbf{z} = \mathbf{\Phi^T}\mathbf{f} \tag{5}$$

We now run into problems with the damping matrix $\mathbf{C}$. While $\mathbf{\Phi}$ diagonalizes $\mathbf{M}$ and $\mathbf{K}$, if it does not diagonalize $\mathbf{C}$ then the system does not properly decouple and the resulting modes are not linearly independent. The linearly *dependent* modes are called complex modes, and accurately modeling them is much more difficult compared to the linearly *independent* normal modes. To get around this issue, for sound synthesis purposes we restrict ourselves to *classical* damping with only normal modes, which means all of our damping matrices must be diagonalizable by $\mathbf{\Phi}$.

3.2 Damping Models

3.2.1 Conventional Damping Models

Various damping models have been developed which guarantee only normal modes.

The most popular model is Rayleigh damping [Rayleigh 1896], in which the damping is a linear combination of mass and stiffness:

$$\mathbf{C} = \alpha_1\mathbf{M} + \alpha_2\mathbf{K} \tag{6}$$

α_1 and α_2 are real-valued parameters to Rayleigh damping models. Rayleigh damping has been, to the best of our knowledge, the only damping model used for sound synthesis in computer graphics. An analysis of Rayleigh damping has shown that for perception of synthesized sound, the Rayleigh damping parameters α_1 and α_2 are geometry-invariant and are best seen as material properties [Ren et al. 2013a]. While other damping models should share this property, they have not undergone the same rigorous evaluation for sound synthesis.

Caughey and O'Kelly proposed a more general model, now known as *Caughey damping* or a *Caughey series* [Caughey and Okelly 1965], which they proved to be a necessary and sufficient condition for normal modes:

$$\mathbf{C} = \mathbf{M}\sum_{i=0}^{n-1}\eta_i(\mathbf{M}^{-1}\mathbf{K})^i \tag{7}$$

All η_i are real-valued parameters for Caughey damping models. In practice, the series could truncated after a few terms.

3.2.2 Generalized Proportional Damping

Generalized proportional damping (GPD) formulates the damping matrix even more generally [Adhikari 2006]:

$$\mathbf{C} = \mathbf{M}\beta_1(\mathbf{M}^{-1}\mathbf{K}) + \mathbf{K}\beta_2(\mathbf{K}^{-1}\mathbf{M}) \tag{8}$$

β_1 and β_2 are matrix valued functions whose only restrictions are that they be analytic near the eigenvalues of their arguments. For example, using $\beta_1(\mathbf{A}) = \alpha_1\mathbf{A}$ and $\beta_2(\mathbf{A}) = \alpha_2\mathbf{A}$ replicates Rayleigh damping. This representation is much more convenient to work with than a Caughey series, as arbitrary functions can be easily plugged in to the β functions. GPD still satisfies the necessary condition of the Caughey series since any continuous function used as a β can be expanded as a power series.

3.3 Modal Sound Synthesis

With these damping models, we have a damping matrix guaranteed to be diagonalizable by $\mathbf{\Phi}$. With the system diagonalized, the free-vibration form is now decoupled into independent second order differential equations:

$$\ddot{z} + c\dot{z} + \omega_n^2 z = 0 \tag{9}$$

c is an entry in the diagonalized damping matrix, and is discussed in more detail in Section 3.4. These equations each have known analytical solutions as damped sinusoids:

$$z(t) = ae^{-dt}\cos(\omega_d t) \tag{10}$$

a is the amplitude of the sinusoid, while the damping coefficient $d = c/2$ defines the rate at which the amplitude decreases. ω_n in Equation 9 is the natural undamped frequency of oscillation, but in the presence of damping we use the *damped* frequency ω_d:

$$\omega_d = \sqrt{\omega_n^2 - d^2} \tag{11}$$

For real-time synthesis, a preprocessing step is first performed for a given object and material. In this step, the eigendecomposition is performed and the resulting $\mathbf{\Phi^T}$ and each mode's d and ω_d are saved. At runtime, an applied force f is transformed to mode space by $\mathbf{\Phi^T}$, and the resulting vector contains the amplitudes with which to excite each mode. The resulting damped sinusoids can be combined and sampled at 44 kHz to produce the sound itself.

3.4 Obtaining Damping Coefficients

In practice, we do not actually want to perform the matrix operations in the damping models. Through heavy use of Equation 3, we can find analytical solutions for how $\mathbf{C}$ is diagonalized and compute c in terms of the corresponding eigenvalue ω_n^2. The solution for Rayleigh damping is common in modal sound synthesis work:

$$\begin{aligned}\mathbf{\Phi}^T\mathbf{C}\mathbf{\Phi} &= \mathbf{\Phi}^T\alpha_1\mathbf{M}\mathbf{\Phi} + \mathbf{\Phi}^T\alpha_2\mathbf{K}\mathbf{\Phi} \\ &= \alpha_1\mathbf{I} + \alpha_2\mathbf{\Omega^2} \\ c_i &= \alpha_1 + \alpha_2\omega_{in}^2\end{aligned} \tag{12}$$

Caughey damping is slightly more involved, but leads to a fairly intuitive solution:

$$\begin{aligned}\mathbf{\Phi}^T\mathbf{C}\mathbf{\Phi} &= \mathbf{\Phi}^T\mathbf{M}\sum_{j=0}^{n-1}\eta_j(\mathbf{M}^{-1}\mathbf{K})^j\mathbf{\Phi} \\ &= \mathbf{\Phi}^{-1}\sum_{j=0}^{n-1}\eta_j(\mathbf{\Phi\Omega}^2\mathbf{\Phi}^{-1})^j\mathbf{\Phi} \\ &= \sum_{j=0}^{n-1}\eta_j\mathbf{\Omega}^{2^j} \\ c_i &= \sum_{j=0}^{n-1}\eta_j\omega_{in}^{2\ j}\end{aligned} \tag{13}$$

For GPD, the equation for c was provided along with a lengthier proof [Adhikari 2006], which we will omit here:

$$\begin{aligned}\mathbf{\Phi}^T\mathbf{C}\mathbf{\Phi} &= \mathbf{M}\beta_1(\mathbf{M}^{-1}\mathbf{K}) + \mathbf{K}\beta_2(\mathbf{K}^{-1}\mathbf{M}) \\ c_i &= \beta_1(\omega_{in}^2) + \omega_{in}^2\beta_2(\omega_{in}^{-2}) \\ c_i &= \beta(\omega_{in}^2)\end{aligned} \tag{14}$$

The final form of the equation can be reached without loss of generality (the second term could be embedded in β_1) and is an even more convenient form to work with.

3.5 Modal Sound Synthesis with GPD

The technical change needed to use GPD for modal sound synthesis is conceptually simple: during precomputation of damping coefficients use Equation 14 instead of Equation 12. GPD's increased flexibility has its downsides: with Rayleigh damping it is tedious, but possible, to select the parameters α_1 and α_2 by hand and fine tune until the resulting sound is acceptable. The challenge now lies in selecting an appropriate β function for the sounding object in question, which covers a much broader space of functions.

β defines a curve in eigenvalue-damping space, which should match as closely as possible to the real-world damping values. Considering damping modeling as a curve fitting problem, Rayleigh damping's linear model is only accurate as long as the true damping curve remains approximately linear.

3.5.1 Power Law Model

One possible solution we propose is to pick functions with real-valued coefficients known to provide good fits to damping curves. Rayleigh and Caughey damping use real-valued coefficients and stay in the toolkit, but it opens up the possibility of other models. As one alternative model, in the study of sound attenuation during propagation there is a well-known power law relation between frequency and attenuation [Szabo 1994]. As sound propagates through a material, the pressure of the sound P attenuates depending on the distance traveled Δx and frequency ω according to:

$$P(x + \Delta x) = P(x)e^{-\mu_1(\omega)^y\Delta x} \tag{15}$$

μ_1 and y are real-valued coefficients which vary depending on material. If we assume the physical phenomenon causing attenuation over distance and damping over time are similar, we can use Equation 14 to derive a similar damping model based on this power law:

$$c_i = \beta(\omega_{in}^2) = \mu_1\omega_{in}^{2\mu_2} \tag{16}$$

μ_1 and μ_2 are now the real-valued parameters to this power law model. While the 2 in the exponent could be incorporated into μ_2, it

allows the function to be written in terms of the eigenvalue for clarity. Because this model is a continuous function of the eigenvalue, GPD guarantees that there is a damping matrix $\mathbf{C}$ which diagonalizes to produce these c values and therefore creates only normal modes of vibration.

Empirical findings for μ_2 in the context of attenuation place it in a range between 0 and 1, with 1 being a common finding for many materials. If damping can be said to be similar, this may provide some physical justification for Rayleigh damping, whose second term fits this model. However, Rayleigh damping could not handle any materials with an $\mu_2 \neq 1$ while a power law damping model could adapt for each material. We use this power law model in later evaluation, but GPD allows for a wide range of models, and we would encourage trying out different models to find optimal fits.

4 Material Parameter Extraction

Instead of fine-tuning damping model parameters values by hand, we can instead automatically estimate them from recorded audio. Rayleigh damping has been studied to determine that α_1 and α_2 are geometry-invariant and can be considered as high level material properties [Ren et al. 2013a]. Other damping models have not undergone the same level of rigorous testing, but we hypothesize that for any damping model with real-valued parameters, the parameters will be similar across objects with different shapes and the same material. Ideally, we would like to use the recorded audio to extract all the material properties needed to synthesize sound of an object in any shape but made of the same material. The relevant material properties are Young's modulus, Poisson's ratio, density, and damping model parameters for the chosen damping model.

Ren et al. have suggested an optimization-based framework for extracting these parameters using a Rayleigh damping model [Ren et al. 2013b]. We extend the optimization framework to automatically identify material parameters for any damping model, including both Caughey series and GPD models. In this section, we first review the parameter identification pipeline for turning an audio recording into example-guided, physically-based synthesized sound. We then describe how we generalize such a system to extract parameters for alternative damping models.

4.1 Extraction of Rayleigh Coefficients

4.1.1 Feature Extraction

First, the input audio file is processed to extract audio *features*, where each feature represents a single damped mode and consists of a frequency, damping coefficients, and initial amplitude. Peaks are automatically selected by searching over multiple power spectrograms of the same audio, where the spectrograms vary in temporal and spectral resolution. The spectrograms with high temporal resolution have low spectral resolution and will be useful in different situations than the spectrograms with low temporal resolution and high spectral resolution. Once a peak is identified, an optimizer searches the local variations in frequency, damping, and amplitude to produce the best fit. The power spectrogram of the new peak is subtracted from the current spectrograms and the process repeats until a large enough percentage of the power is accounted for in the extracted features.

4.1.2 Parameter Estimation

In order to estimate the material parameters of the recorded object, some additional information is needed. Poisson's ratio is not optimized as part of this system, so it must be predetermined before starting. Additionally, eigenvalues scale proportionally with Young's modulus (E) and inversely with density, so only the ratio $\gamma = E/\rho$ can be optimized and a density value needs to be predetermined in order to get an absolute Young's modulus. Finally, modal analysis is performed on a discretized model of the object with assumed material parameters and it is struck with a unit impulse at the same hit point as the real-world object. The resulting $\mathbf{\Phi}^T f$ contains the initial mode amplitudes of the assumed object, and since the same hit point was used for the recorded object, its amplitudes should be a scaled version of the same. The final set of parameters used in optimization is γ, scale, and real-value coefficients. Scale is not a *material* parameter, but without it the optimizer would be unable to properly match the volumes of the recorded and reconstructed audio.

The parameters are determined through optimization looking to minimize a similarity metric by varying the parameters. The chosen metric combines both evaluation of differences between power spectrograms and differences between features. Power spectrograms are compared after being transformed based on psychoacoustic principles. Since humans cannot easily distinguish between similar frequencies of sound, and since this effect varies in strength across the range of hearing, the frequency dimension is transformed to the Bark scale which properly accounts for this effect. Perception of loudness also varies based on frequency, so the intensities are converted to the sone scale, in which the loudness is scaled depending on the frequency of the sample. With the spectrograms converted to perceptually-based scales, they can now be compared to one another by finding the squared difference between them:

$$\Pi_{psycho}(\mathbf{I}, \bar{\mathbf{I}}) = \sum_{m,z} \left(\mathbf{T}(\mathbf{I})[m,z] - \mathbf{T}(\bar{\mathbf{I}})[m,z]\right)^2 \tag{17}$$

The other part of the metric operates on (frequency, damping, amplitude) features extracted from the recorded audio or taken from an assumed mode of vibration and its corresponding entry in the $\mathbf{\Phi}^T f$ amplitude vector. Once again, the frequency is converted to the Bark scale for psychoacoustic purposes. The damping is also inverted to become duration, which is less sensitive to differences between very short bursts of sound. The sets of features are then matched with one another using the *Match Product Ratio* metric. A single feature f_1 can be compared to the set of possible matches in the other set of features $\bar{\mathbf{f}}$ using the *point-to-set* match ratio:

$$R(f_i, \bar{\mathbf{f}}) = \frac{\sum_j u_{i,j} k(f_i, \bar{f}_j)}{\sum_j u_{i,j}} \tag{18}$$

u is a matrix of weights to give higher priority to prominent features, while k is a measure of distance between the two points on $[0, 1]$ such that a 1 means an exact fit. A full set of features $\mathbf{f}$ can be compared to another set of features $\bar{\mathbf{f}}$ using the *set-to-set* match ratio:

$$R(\mathbf{f}, \bar{\mathbf{f}}) = \frac{\sum_i w_i R(f_i, \bar{\mathbf{f}})}{\sum_i w_i} \tag{19}$$

w is a vector of weights similar in purpose to u. With these match ratios defined, the Match Ratio Product metric for the extracted audio features $\mathbf{f}_{extract}$ and the assumed audio features $\mathbf{f}_{assume}$ is:

$$\Pi_{MRP} = -R(\mathbf{f}_{extract}, \mathbf{f}_{assume}) R(\mathbf{f}_{assume}, \mathbf{f}_{extract}) \tag{20}$$

The final metric to optimize is:

$$\Pi_{hybrid} = \frac{\Pi_{psycho}}{\Pi_{MRP}} \tag{21}$$

This metric takes into account both the power spectrograms and features, using psychoacoustic scales where it can better match human hearing.

The starting points are generated by choosing multiple pairs of two dominant features extracted from the recorded audio and fitting a line to them to generate starting material parameters. For each mode from the modal analysis on assumed parameters, the eigenvalue and the corresponding amplitude in the $\mathbf{\Phi}^T f$ vector are used to generate the starting γ and scale values. The starting γ is selected as the value which would cause the selected mode to have the same frequency as one of the dominant features. Similarly, the starting scale is the one that would scale the amplitude of the selected mode to the amplitude of the dominant feature. Together, these define a starting point for the optimizer.

By running a non-gradient based optimizer on this metric from each starting point and selecting the best final point, material parameters which best recreate the original sound are selected. The resulting γ can be used to find the Young's modulus, while the material parameters, such as α_1 and α_2 in Rayleigh model, define the damping curve. These parameters can then be transferred to other geometries, effectively applying the material parameters of the original recorded object to different virtual models.

Ren et al. also presented a method for taking the residual sound (anything not captured by the modal feature extraction) and transferring it to alternative shapes, making even the residual somewhat geometry-invariant [Ren et al. 2013b]. For this paper, we focus on the extraction of damping parameters and we do not adopt residuals for sound synthesis.

4.2 Extraction of GPD Parameters

In order to extract damping parameters from an arbitrary damping model, we reformulate the set of optimized parameters to include γ, scale, and all of the damping model parameters (of which there could be many). Any instances of Rayleigh damping computation are replaced with a general β function (instead of real values). Feature extraction and metric evaluation are still applicable, as the damping model plays no role there.

The most significant difference lies in generation of the starting points. With Rayleigh damping's linear fit, any two points define a new line and new starting α_1 and α_2, but arbitrary β functions may require many points to define a curve. Instead, we repeatedly sample a customizable percentage of the dominant features—weighted by dominance. On each sample, we perform least-squares nonlinear regression on the damping model to create the starting damping model parameter values. The sampling percentage is ideally set such that there are enough features to get a useful fit, but not so many features that the starting points are tightly clustered.

To generate a starting γ, we pair up each mode's eigenvalue from the modal analysis on assumed parameters with each damping value from extracted features. γ is computed through root-finding as the value that maps the mode's eigenvalue to the feature's damping value through our sampled damping model. We are effectively asking, "If this eigenvalue happened to be damped at this rate, what would γ have to be?" See Figure 1 for a visual example. Similarly, the scale is chosen to match the mode's amplitude to the extracted feature's amplitude. This is a fairly exhaustive search and the search space has many local minima, so quite a few starting points are needed to find a nearly-global minima. Once these starting points are generated, the optimizer can proceed to minimizing the metric.

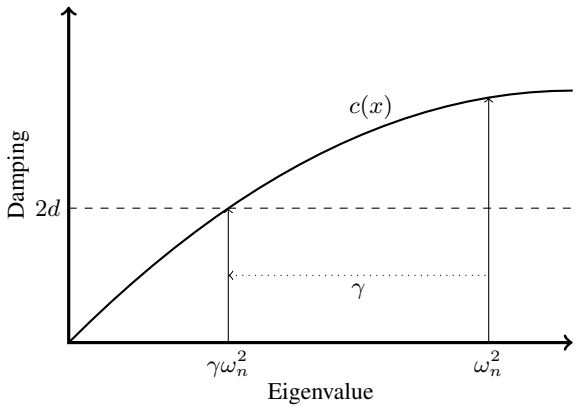

Figure 1: *Generation of γ (ratio of Young's modulus to density) starting value using a damping value d from an extracted feature, an eigenvalue ω_n^2 from modal analysis on assumed parameters, and a sampled damping model $c(x)$. γ is chosen such that $c(\gamma\omega_n^2) = 2d$.*

5 Results

5.1 Sound Synthesis

We implemented Rayleigh, Caughey, and GPD-based sound synthesis using FEM meshes as our discretization. Audio was played using the STK library [Cook and Scavone 1999; Scavone and Cook 2005], and videos were created using Blender with Bullet Physics for rigid-body simulation [Coumans 2015].

Our meshes contain around 10,000–20,000 tetrahedra, resulting in up to 30 minutes of precomputation time on modal analysis using a desktop workstation computer. Run time for material parameter estimation is most dependent on the length of the sound: one starting point for a short impact converges in a few seconds, while a reverberative object requires up to ten minutes. At runtime, sound is synthesized at 44 kHz: the highest frequency we can perceive is around 22 kHz, so there is no benefit in synthesizing sound more often than twice that frequency. The synthesis steps are fast enough that the 44kHz update rate can be easily maintained for a number of sounding objects even on a laptop.

5.2 Parameter Extraction

We implemented our extended version of the material parameter optimization process, and have been able to extract parameters using different damping models. See Figure 2 for the full set of objects used, comparing the real objects in the top row to the meshes in the bottom row.

Table 1 presents some results from performing extraction of material damping parameters given recorded audio, using Rayleigh damping, second-order Caughey damping, and power law damping. These parameters may not be the most globally optimal as reported by the optimizer, but they are all parameters that have reasonably low metric values and are at least locally optimal. Plastic 1 comes from the rigid, clear plastic bowl, while Plastic 2 comes from the thin and much more flexible dog food scoop. We can assign some physical meaning to these parameters; for example Porcelain has smaller values for Rayleigh's α_2 and Power's μ_2, indicating relatively less damping at higher frequencies. Also note that while most materials are best fit by a Caughey series whose coefficients alternate signs with each term, Plastic 2 was better fit by a set of

Figure 2: *The objects used to extract material parameters. The top row shows pictures of the real objects, while the bottom row shows meshes modeling the objects.*

		Plastic 1	Plastic 2	Porcelain	Wood	Aluminum
Shared	E	8	2.4	20	9.9	0.88
Rayleigh	α_1	125	58	189	35	.225
	α_2	8e-6	1e-6	1.5e-8	4.6e-7	1.45e-6
Caughey	η_1	280	85	420	277	9.7
	η_2	-3.6e-7	6.6e-7	-2.4e-6	-2.0e-6	-3.4e-6
	η_3	2.0e-15	5.6e-14	3.8e-15	4.8e-15	5.8e-13
Power	μ_1	1.13	.19	163	6.7	.02
	μ_2	.3	.37	.01	.18	.445

Table 1: *Extracted material parameters for a selection of materials. Young's modulus is given in GPa. See Section 3.4 for the usage of each damping parameter. These are not necessarily the parameters that minimize the MPR metric, but they are locally optimal and agree on a somewhat physically plausible Young's modulus E across damping models.*

only positive coefficients. The other damping models are unable to capture this unusual damping behavior as well as the higher-order Caughey series.

5.3 User Study

One hypothesis with this work is that alternative damping models can recreate a wider range of more realistic audio with more complex non-linear damping characteristics. In order to evaluate the perceptual realism of the damping models, we conducted a preliminary user study where subjects were asked to compare sound generated with different damping models. This study is a first exploration of the differences between damping models. The study evaluates if subjects can tell the difference between them, and if so, which they find more realistic.

5.3.1 User Study Setup

This study was conducted entirely online through the subject's web browser. Subjects were informed about the procedure of the study and instructed to use headphones or earbuds in order to better control the audio environment.

Subjects were presented with a series of pairs of videos of an object being dropped on a flat surface. Refer to Figure 2 for images of the objects used in the study. Each pair of videos showed the same visual imagery, but had different audio generated using either Rayleigh damping, a second-order Caughey series, or a power law model. Subjects were asked to rate, on a scale from 1 to 11, which video they perceived as more realistic, with a 1 indicating a strong preference for the video on the left, an 11 indicating a strong preference for the video on the right, and 6 being in the middle. Subjects were also asked to rate the similarity of the sound in the videos, where a 1 is very different and an 11 is indistinguishable. The videos could be watched repeatedly and subjects could return to previously-answered questions in case their opinions change.

5.3.2 User Study Results

40 subjects participated in the study, and while little demographic information was collected, the recruitment methods used were likely to attract many subjects with little experience in evaluating sound quality. We can begin by combining data from objects together to get a general sense of the perceived realism ratings as a whole. Recall that perceived realism was rated on a scale from 1 to 11, with 6 being in the middle. In comparisons between Rayleigh damped and Caughey damped audio, a 1 indicates preference for Rayleigh and an 11 indicates preference for Caughey. Across all objects, when subjects compared Rayleigh and Caughey damping, the realism rating was 6.5 ± 3.3, and there was not a significant preference in realism between the two ($p > .05$). When comparing Rayleigh to Power damping, where a 1 again indicates a preference for Rayleigh, the realism rating was 4.78 ± 2.57 and there was a preference for Rayleigh ($p < .0001$). Finally, when comparing Caughey to Power damping, where a 1 indicates a preference for Caughey, the average realism rating was 3.95 ± 2.92 and there was a preference for Caughey ($p < .0001$).

We can also look at the subject-reported similarity values to determine if the subjects could notice a perceptual difference between the models. Similarity was rated on a scale from 1 (very different) to 11 (very similar). In comparisons between Rayleigh and Caughey damping, the similarity was 5.7 ± 3.0. Between Rayleigh and Power damping, the similarity was 6.9 ± 3.4. Finally, Caughey and Power damping had a similarity of 4.4 ± 2.7.

Object	Models	$\bar{x}$	σ	p
Metal Plate	R/C	6.5	3.2	.35
	R/P	*	*	*
	C/P	*	*	*
Plastic Bowl	R/C	6	3.3	1
	R/P	2.9	1.9	.0002
	C/P	3.9	2.0	.01
Plastic Scoop	R/C	3.6	2.2	.0035
	R/P	4.3	2.3	.095
	C/P	5.2	3.4	.294
Porcelain Bowl	R/C	6.1	3	.93
	R/P	6	0	1
	C/P	7.1	3.0	.32
Porcelain Plate	R/C	6.6	3.5	.53
	R/P	4.8	2.9	.16
	C/P	4.2	2.8	.04
Small Floor Tile	R/C	8.9	2.4	.001
	R/P	4.4	3.0	.13
	C/P	2.1	1.5	$<$.0001
Short Wood Block	R/C	6.6	4.2	.63
	R/P	5.2	3.6	.5
	C/P	4.0	2.9	.014
Long Wood Block	R/C	7.8	1.8	.005
	R/P	5.2	2.9	.34
	C/P	2.2	1.6	$<$.0001

Table 2: *Realism values from the user study. For each object and each pair of damping models (R for Rayleigh, C for Caughey, P for power), the range of realism ratings is shown as a mean $\bar{x}$ and a standard deviation σ. Ratings lower than 6 are a preference for the damping model on the left side of the slash. The p-value evaluates whether there is a significant difference in realism preference from the "no preference" realism rating of 6. *The metal plate power model was not included in the study.*

For more detailed results, realism values for each of the objects individually are laid out in Table 2. For simplicity, comparisons between two damping models, say Rayleigh and Power, are abbreviated as R/P in the table. The results for the small floor tile and the long wood block contain some of the most significant results, with Caughey damping greatly preferred over the other two. The plastic scoop was the only object for which Rayleigh was preferred over Caughey, but for most of the objects the difference was not statistically significant. The porcelain bowl is an interesting case where Rayleigh and Caughey are nearly identical in realism, but for once the power model is considered to be nearly as (possibly more) realistic.

5.3.3 Discussion

When compared to either of the alternatives, Power was perceptually considered to be less realistic by .47 standard deviations in the case of Rayleigh damping and by .7 standard deviations in the case of Caughey damping. This is only a moderate preference, but enough to be statistically significant. One simple explanation for this result is that a power law may not provide a good curve fit to the data. Despite this, the two most similar sounding damping models were reported to be Rayleigh and Power. The power law model often seems to be perceptually similar to Rayleigh damping at higher frequencies, while having less damping on the lower frequencies. In some cases the amplified lower frequencies sound more realistic, but in most of the cases in this user study it comes across as too strong and unrealistic.

In theory, Caughey damping can only improve upon Rayleigh damping since the higher order terms can simply be set to 0 if the linear model would be optimal. The result that the difference in realism between the two of them was not statistically significant could imply that the benefit gained from the second-order term is not be large enough to be perceptibly noticeable. However, the similarity rating between them is not particularly high, so a better interpretation might be that there *is* a perceptually noticeable difference between the sounds, but subjects had difficulty determining which of the two different sounds was more realistic.

Subjects did not have access to any ground truth sound recordings, which made the task more difficult. However, this is reasonable given that the primary application we are considering is using extracted parameters to synthesize contact sounds in interactive virtual environments. The study focuses on subjects' *perception* of the sounds presented in an entirely virtual environment to understand how they would react to these sounds in a game, virtual teleconference, or training simulation. The subjects only need to perceive the sounds to be realistic. This perceptual approach also reduces the need for participants to be skilled at evaluating sound; the "perceptual ground truth" is not consistent between subjects.

6 Conclusion

We have presented the integration of generalized proportional damping with modal sound synthesis techniques. We explained how to derive GPD-based damping models, using a power law model as an example. We extended an existing method for extracting material parameters to extract real-valued parameters from an arbitrary damping model. We conducted a preliminary user study comparing Rayleigh, Caughey, and power law damping models.

While the user study did not find an improvement in perceived realism when using our example of a GPD-based damping model, this result provides other benefits. From a somewhat different perspective, this study provides additional validation of the popular Rayleigh damping model in that second order Caughey damping models were not always perceptually more realistic than Rayleigh damping and that GPD-based models which provide a perceptual improvement may not be easy to find. In light of this result, future research may find success in using models which encapsulate a larger function space. For example, genetic programming and neural nets can both approximate continuous functions without needing to specify a damping model in advance.

Future work in the area of GPD should likely focus on using the extended systems described in this paper to explore alternative GPD-based damping models. Additionally, it would be an improvement to incorporate the residual audio after the material parameter estimation process, transferring it to other geometries. There is some uncertainty about the transferability of arbitrary GPD parameters; an analysis similar to the one done for Rayleigh damping [Ren et al. 2013a] could help determine if the real-valued model parameters can all be considered material properties. Work in these areas would help improve understanding of damping properties and hopefully lead to more immersive sound.

6.1 Acknowledgments

This research is supported in part by the National Science Foundation and the UNC Arts and Sciences Foundation.

References

ADHIKARI, S., AND WOODHOUSE, J. 2001. Identificaion of damping: Part 1, viscous damping. *Journal of Sound and Vibration 243*, 1, 43 – 61.

ADHIKARI, S. 2006. Damping modelling using generalized proportional damping. *Journal of Sound and Vibration 293*, 12, 156 – 170.

BONNEEL, N., DRETTAKIS, G., TSINGOS, N., VIAUD-DELMON, I., AND JAMES, D. 2008. Fast modal sounds with scalable frequency-domain synthesis. *ACM Transactions on Graphics (SIGGRAPH Conference Proceedings) 27*, 3 (August).

CAUGHEY, T., AND OKELLY, M. 1965. Classical normal modes in damped linear dynamic systems. *Journal of Applied Mechanics 32*, 3, 583–588.

CAUGHEY, T. 1960. Classical normal modes in damped linear dynamic systems. *Journal of Applied Mechanics 27*, 2, 269–271.

CHADWICK, J. N., ZHENG, C., AND JAMES, D. L. 2012. Precomputed acceleration noise for improved rigid-body sound. *ACM Transactions on Graphics (Proceedings of SIGGRAPH 2012) 31*, 4 (Aug.).

COOK, P. R., AND SCAVONE, G. P. 1999. The synthesis toolkit (stk). In *In Proceedings of the International Computer Music Conference*.

COUMANS, E. 2015. Bullet physics simulation. In *ACM SIGGRAPH 2015 Courses*, ACM, New York, NY, USA, SIGGRAPH '15.

JAMES, D. L., BARBIČ, J., AND PAI, D. K. 2006. Precomputed acoustic transfer: output-sensitive, accurate sound generation for geometrically complex vibration sources. In *ACM Transactions on Graphics (TOG)*, vol. 25, ACM, 987–995.

LLOYD, D. B., RAGHUVANSHI, N., AND GOVINDARAJU, N. K. 2011. Sound synthesis for impact sounds in video games. In *Proceedings of the Symposium on Interactive 3D Graphics and Games 2011*, ACM.

O'BRIEN, J. F., SHEN, C., AND GATCHALIAN, C. M. 2002. Synthesizing sounds from rigid-body simulations. In *Proceedings of the 2002 ACM SIGGRAPH/Eurographics Symposium on Computer Animation*, ACM, New York, NY, USA, SCA '02, 175–181.

PAI, D. K., DOEL, K. V. D., JAMES, D. L., LANG, J., LLOYD, J. E., RICHMOND, J. L., AND YAU, S. H. 2001. Scanning physical interaction behavior of 3d objects. In *Proceedings of the 28th annual conference on Computer graphics and interactive techniques*, ACM, 87–96.

RAGHUVANSHI, N., AND LIN, M. C. 2006. Interactive sound synthesis for large scale environments. In *Proceedings of the 2006 Symposium on Interactive 3D Graphics and Games*, ACM, New York, NY, USA, I3D '06, 101–108.

RAYLEIGH, J. W. S. B. 1896. *The theory of sound*, vol. 2. Macmillan.

REN, Z., YEH, H., KLATZKY, R., AND LIN, M. C. 2013. Auditory perception of geometry-invariant material properties. *Visualization and Computer Graphics, IEEE Transactions on 19*, 4, 557–566.

REN, Z., YEH, H., AND LIN, M. C. 2013. Example-guided physically based modal sound synthesis. *ACM Trans. Graph. 32*, 1 (Feb.), 1:1–1:16.

SCAVONE, G. P., AND COOK, P. R. 2005. Rtmidi, rtaudio, and a synthesis toolkit (stk) update. In *In Proceedings of the International Computer Music Conference*.

SZABO, T. L. 1994. Time domain wave equations for lossy media obeying a frequency power law. *The Journal of the Acoustical Society of America 96*, 1, 491–500.

VAN DEN DOEL, K., AND PAI, D. K. 1996. The sounds of physical shapes. *Presence 7*, 382–395.