

# Audio-Material Reconstruction for Virtualized Reality using a Probabilistic Damping Model

Category: Research

Paper Type: Algorithm/Technique



Fig. 1. A real-time interactive virtual environment where striking objects produces dynamic sounds using our method (left); a ball striking plates of various sizes plays a melody (middle); and a set of wind chimes blowing in a virtual forest (right).

**Abstract**— Modal sound synthesis has been used to create realistic sounds from rigid-body objects, but requires accurate real-world material parameters. These material parameters can be estimated from recorded sounds of an impacted object, but external factors can interfere with accurate parameter estimation. We present a novel technique for estimating the damping parameters of materials that probabilistically models many of these external factors. We use a generative model to represent the combined effects of material damping, support damping, and sampling inaccuracies, then use maximum likelihood estimation to fit a damping model to recorded data. This technique simplifies the recording process, only requires audio as input without knowing the precise object geometry or the exact hit location, and uses multiple recorded impact sounds to improve accuracy. We validate the effectiveness of this technique with a comprehensive analysis of a synthetic dataset and a perceptual study on material identification.

**Index Terms**—Sound synthesis, modal analysis, damping modeling, maximum likelihood estimation

## 1 INTRODUCTION

Maintaining the sense of immersion is one primary goal in creating interactive virtual environments. An important aspect of immersion is the feeling of cohesiveness between senses, e.g. an object that looks like wood should also sound like wood. Furthermore, objects colliding with one another or being impacted by a user should produce different sounds depending on the location, direction, and magnitude of impact. For rigid objects such as tables, dishes, and dice, a physically-based, real-time technique, *modal sound synthesis*, can be used to analyze the vibrations of the objects and produce dynamic impact sounds [21]. Modal sound synthesis can improve a user’s sensory cohesion when dealing with rigid objects, but it requires accurate real-world material parameters. Damping, which determines the rate at which vibrations and sound decay over time, is crucial in differentiating between different materials, such as a reverberative metal pan and a muted wood plate [25], giving the characteristic sound for each audio material. Some parameters, e.g. density and Young’s modulus, can be looked up for known materials, but damping properties can be difficult to identify and parameterize.

One way to obtain material parameters is by estimating them from recorded impact sounds. Given an object made of a particular material, we can strike the object and record the resulting sound. Existing methods use the sound, along with some mandatory knowledge about the shape and properties of the struck object, to estimate a number of material parameters [26]. The material parameters can be applied to any virtual object, effectively “virtualizing” the audio characteristics of a given material. While recent techniques have been able to estimate material damping properties, they assume that there is very little effect on damping from external factors.

For example, 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* will introduce additional damping, as energy is transferred from the vibrating object

to the more stationary support. To account for the support damping, recordings must be made with supports that introduce minimal damping, requiring a carefully controlled recording environment using special support [22], e.g. strings or rubber bands, to suspend the object [25]. Other factors that affect estimated damping values, such as complex modes of vibration, background noise, and accumulated error during estimation are assumed by prior work to be minimized.

In this paper, we present a practical and efficient technique to estimate material damping properties from recorded impact sounds that accounts for the additional factors affecting damping. By automatically acquiring the material parameters directly from a given audio sample, our technique is more general and applicable to real-world recordings in less controlled environments. Unlike previous work [25], this method is fast and requires no prior knowledge about the recorded object’s geometry, size, or hit location. As a result, we are able to create virtualized versions of real-world materials that can be applied to any sounding objects. Figure 2 shows the full pipeline for estimating material parameters and using them to synthesize sound. The only input needed to estimate damping parameters is a set of recorded sounds. The key contributions of this work include:

- A new model of the probability distribution of damping values that separately considers each source of damping; and
- Application of this probabilistic model to recorded audio for estimating material damping parameters.

We test our approach using impact sounds from several real-world objects with varying hit points and environmental factors. First, we experimentally verify the effectiveness of our algorithm via an auditory perceptual study on material identification. Next, we validate this approach by estimating parameters of a synthetic audio dataset with

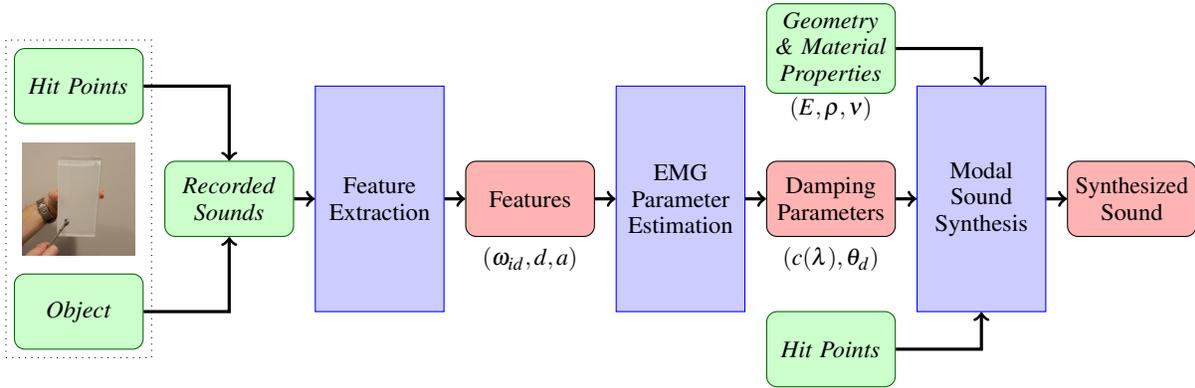


Fig. 2. Our pipeline for estimating material parameters from recorded audio and using the parameters to synthesize sound for objects of the same material. Inputs are in green with italic text. If the object and hit points are unknown, the pipeline can begin with recorded sounds instead. Feature extraction is described in Section 4.1, parameter estimation with an exponentially modified Gaussian (EMG) is described in Section 4.3, and modal sound synthesis with contacts is described in Sections 3.1 and 4.5.

known parameters. Finally, the results on the synthetic audio dataset are compared against two alternative techniques. Fig. 1 shows the demonstration of our system on several complex virtual environments – both indoor and outdoor – consisting of virtual objects of different materials. These virtual scenes are designed and implemented in the Unreal Engine and integrated with an HTC Vive headset, allowing users to interact with objects in real time.

The rest of the paper is organized as follows. Section 2 reviews related work and Section 3 describes the mathematical formulations for modal sound synthesis and damping models. We introduce the algorithm for parameter estimation in Section 4 and present the results and analysis in Section 5. We conclude by discussing limitations and possible future work.

## 2 PREVIOUS WORK

Parameter estimation has been extensively studied across a diverse engineering and scientific domains, as well as in computer graphics. We focus our discussion only on works related to sound synthesis.

**Sound synthesis** techniques attempt to recreate realistic audio, while providing variance between sounds so that each is distinct and natural. Strings and drums are relatively easy to simulate through the use of a wavetable and a series of filters [16]. For simple objects with known analytical modes of vibration, the frequencies of their modes can be used to synthesize impact sounds [32]. For more complicated rigid objects, a discretized model of the object can be used to approximate the vibrations of the modes and synthesize sound for any conceivable object [21]. Simulation of acoustic radiance can help spatialize synthesized modal sound [14], and can be approximated for real-time applications [18].

Existing methods can provide real-time sound synthesis of relatively simple objects. However, as the complexity of object geometry increases, optimizations must be made. For example, modes can be efficiently culled based on psychoacoustic principles [23] to improve runtime performance or significantly compressed based on symmetries in the object to reduce memory usage [17].

Some recent papers consider the effect of contacts with a sounding object. In order to create an object that vibrates at user-selected frequencies, the unwanted frequencies can be reduced by resting the object on foam blocks [4]. The foam blocks are positioned such that they inhibit vibration at locations where the unwanted frequencies are strong and the desired frequencies are weak. A contact model can be used to modify the damping matrix for sound synthesis based on forces applied by other objects [37]. With this contact model, object contacts increases the damping of frequencies based on the force of the contact, the coefficient of friction, and a material-dependent value.

**Damping** has long been a concern in the construction of buildings and other structures [20], but also plays a significant role in modal sound synthesis. The physical phenomena behind damping in vibrating structures are complex. There are a number of ways to model their

effects to varying degrees of accuracy [29, 34]. Under the common assumption of viscous damping, various *damping models* describe the damping of an object as a function of its mass and stiffness [1].

**Parameter recovery from recorded sounds** is an ideal [15] that can be powerful, but is currently impossible to achieve for general objects without sufficient constraints on the problem. One of the objectives of this work is to minimize the number of assumptions on the required input for accurate estimation of material parameters that give each sounding object its characteristic audio quality. Previous analysis/synthesis techniques extract deterministic features of an input sound while treating the rest of the sound as stochastic noise [28]. Another option is to extract the modal content of the sound, and then apply random variation to the gains to produce plausible variations on the initial sound [19].

With many audio samples at known locations on the object’s surface, the spectral content can be interpolated to approximate the sound at an arbitrary point [22]. Alternatively, the Young’s modulus for small parts of the object can be individually optimized to best recreate the input sounds [35]. These techniques produce results specific for a single object that cannot be easily transferred to another shape. A more recent technique focuses on estimating material parameters from a single audio recording, where the resulting material parameters are not specific to any one geometry and have more versatile applications [26]. This technique has been extended to support optimization over arbitrary damping models [30]. However, both methods assume that the object geometry, its exact dimensions, and the precise hit point are all known, which is not always the case with pre-existing audio recordings.

Material parameters for *visual* simulation of object motion can also be estimated from video sources. Existing techniques can estimate similar parameters for various rigid [11] and deformable [33, 36] objects and materials. In contrast, this paper focuses solely on reconstruction of audio-material parameters.

## 3 MODAL SOUND SYNTHESIS

As an object vibrates, its surface deforms and oscillates. These oscillations produce pressure waves in the surrounding medium which propagate through the environment. Upon reaching the ear, the variation in pressure over time is perceived as sound. The standard range of human hearing covers sound waves between 20 Hz and 20 kHz. In this section, we describe a popular sound synthesis technique and explain the need for accurate damping parameters.

### 3.1 Modal Analysis

Linear modal sound synthesis is a common technique to produce the modal components of a sound. We give a brief overview on modal sound synthesis here and refer to [21] for more comprehensive tutorials and [30] on damping models for sound synthesis.

Modal sound synthesis assumes that an object’s vibrations can be broken down as a linear combination of its *modes of vibration*. Modal

sound synthesis traditionally operates under the assumption of viscous damping within each object [3]. As described in Section 3.3, it also assumes that parameters defining the rate of damping for each mode are simply a function of that mode’s frequency and not a function of the object’s geometry.

There are two key steps to modal sound synthesis: preprocessing *modal analysis* and runtime *modal synthesis*. Modal analysis uses the geometry of the object along with real-world material parameters, such as the Young’s Modulus and Poisson’s Ratio, to perform the decomposition of vibrations. This step can be performed numerically using a discretized tetrahedral or spring-mass model of the object and representing its vibrations with a system of equations:

$$M\ddot{r} + C\dot{r} + Kr = f. \quad (1)$$

This system models the acceleration, velocity, and displacement of each discrete element  $r$  in response to a force vector  $f$ . With  $n$  discrete elements,  $r$  and  $f$  are of dimension  $3n$  to account for impulses and movement in three dimensions. The mass matrix  $M$  and the stiffness matrix  $K$  can be constructed based on the geometry of the mesh and the connectivity of its elements. The matrix  $C$  is the viscous damping matrix, which describes how and where vibrations decay within the object.  $C$  has specific limitations on how it can be constructed, which will be described in Section 3.3.

The central step of modal analysis is a generalized eigendecomposition of the system. The eigendecomposition produces an eigenvector matrix  $\Phi$ , that transforms between element space and mode space, and a diagonal matrix  $\Omega^2$  containing the eigenvalues.

### 3.2 Modal Synthesis

Assuming the object is in free vibration after the initial impulse, each equation in the decoupled system is a second order differential equation:

$$\ddot{z}_i + c_i\dot{z}_i + \omega_m^2 z_i = 0. \quad (2)$$

Here,  $z_i$  is the current amplitude of the mode,  $c_i$  is the corresponding entry in the decoupled damping matrix, and  $\omega_m$  is the corresponding undamped natural frequency from  $\Omega$ , the square root of the eigenvalues. These equations have known solutions as damped sinusoids:

$$z_i(t) = a_i e^{-d_i t} \cos(\omega_{id} t). \quad (3)$$

The initial amplitude of the mode is given by  $a_i$ , while  $d_i = c_i/2$  is the rate of decay of vibrations.  $\omega_{id}$  is the *damped* frequency of vibration, which can be calculated as:

$$\omega_{id} = \sqrt{\omega_m^2 - d^2}. \quad (4)$$

When an impact occurs at runtime, the force vector is transformed by  $\Phi^T$  to determine the initial mode amplitudes. Each damped sinusoid can then be sampled at 44 kHz to produce the resulting sound in real time. With various optimizations [17, 23], sound synthesis can be performed on many objects simultaneously in an interactive application.

### 3.3 Material Damping Modeling

There are a number of *material damping models* that provide means to construct appropriate  $C$  matrices as a function of the other two matrices  $M$  and  $K$ . These model damping of vibrations as a result of the material of the object by itself. The damping models are necessary to separate the vibrations into independent *normal* modes of vibration. While other damping matrices are possible, these damping models avoid *complex* modes which are difficult to model and slow to evaluate in real-time.

The most common damping model for sound synthesis is Rayleigh (proportional) damping [24], in which  $C$  is a linear combination of  $M$  and  $K$ :

$$C = \alpha_1 M + \alpha_2 K. \quad (5)$$

Physically speaking, this model assumes that damping occurs in each discrete element proportional to the mass of that element, as well as

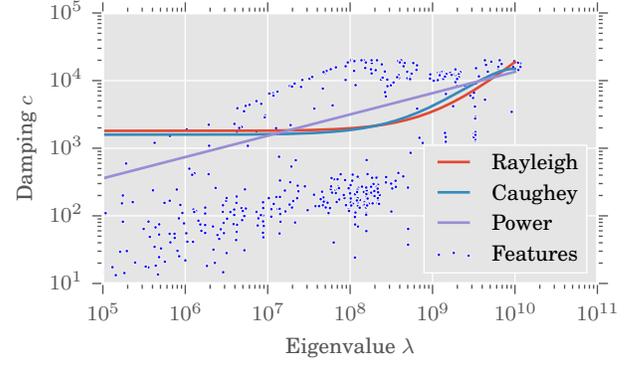


Fig. 3. Three possible damping models fit to real-world  $(\lambda, c)$  pairs using least squares fits. These demonstrate the usual shapes of each damping function on log-log plots. The parameters estimated by least squares here vastly overestimate the material damping and are not ideal, as described in Section 4.2.

in each interface between elements proportional to the stiffness of the connection. Upon decoupling Equation 5, each mode’s entry in  $c$  is:

$$c_i = \alpha_1 + \alpha_2 \omega_m^2. \quad (6)$$

Other material damping models also guarantee diagonalizable  $C$  matrices. In Caughey damping, each  $c_i$  is a polynomial function of  $\omega_m^2$  to an arbitrary degree [6, 7]. Generalized Proportional Damping (GPD) is an even more general model, where each  $c_i$  can be an arbitrary function of  $\omega_m^2$  as long as the function is analytical and continuous near the eigenvalues of the system [2].

We consider one additional GPD-derived damping model: a hybrid model incorporating Rayleigh damping and a power law damping model [30]. The damping matrix entries are described according to the function:

$$c_i = \gamma_1 + \gamma_2 \omega_m^{2\gamma_3}. \quad (7)$$

When  $\gamma_1$  is 0, this becomes the power law damping model, and when  $\gamma_3$  is 1, this becomes the Rayleigh damping model. Since previous work has found that the optimal damping model varies depending on the object [30], this hybrid model can model damping best represented by Rayleigh or power law damping.

For a given damping model, the real-valued parameters (e.g.  $\alpha_j$ ,  $\gamma_j$ ) are the *damping parameters* which define the damping of each mode. By varying these values, the same object can be made to sound like a wide range of materials. Damping parameters have been shown to be perceptually geometry-invariant under the Rayleigh damping model [25]; it is reasonable to assume this holds for other damping models as well. Thus, if damping parameters can be estimated for a metal bowl, synthesizing sound for a box with those parameters will produce a metallic sound.

## 4 PROBABILISTIC DAMPING MODELING

In order to perform the synthesis process described previously in Section 3, we need to know the object’s geometry, Young’s Modulus, density, Poisson’s Ratio, and damping parameters for a chosen damping model. A person attempting to add a realistic sounding object to a virtual environment would need to know all of these, so we now consider how this information can be obtained in the first place. The geometry can either be taken from a real-world object or designed for a virtual object. Young’s Modulus, density, and Poisson’s Ratio can be measured from real-world objects, but for many materials these values have been published and approximate values can be selected for synthesis purposes. Damping parameters, on the other hand, are specific to their damping model and even with the popular Rayleigh damping model, parameters are difficult to find for arbitrary materials. In this section, we present our technique for estimating these damping parameters from recorded real-world impact sounds. We also compare

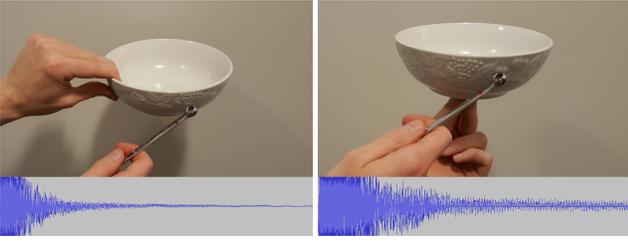


Fig. 4. A porcelain bowl struck in the same location produces different sound when supported with a tight grip (left) or supported by resting on a single point (right). Without accounting for the effect of the support, prior methods would not be able to estimate accurate material parameters from these sounds.

our technique against two alternative parameter estimation methods: a least squares fit and a “lower bound” fit.

#### 4.1 Feature Extraction from Audio

Our technique could use a single recorded sound as input, but in practice multiple sounds are necessary for good accuracy. Our technique requires audio recordings of impact sounds as input. As damping parameters are geometry-invariant [25], we do not need to know the object’s geometry, its size, or its hit location. This is important since the environmental factors may change with each input sound. A mode that is heavily damped by external factors in one sound may be relatively undamped in another, providing additional information about the range of possible damping values.

The first step in our approach is to extract the modal components of each input sound. Assuming the sounds come from rigid objects, the sound produced will be mostly modal and can be decomposed into a set of *features*. Each feature corresponds to one mode of vibration and can be parameterized as a damped sinusoid with a damped frequency  $\omega_{id}$ , an initial amplitude  $a_i$ , and an exponential damping coefficient  $d_i$ .

We adopt a feature extraction process that identifies likely features, then performs local optimization, similar to that of Ren et al. [26]. Multiple power spectrograms of the input audio are constructed, and frequencies with high power are selected. At each of these frequencies, local optimization is performed to find the  $(\omega_{id}, a_i, d_i)$  parameters which best match the local region of the spectrogram. Once a feature is identified, it can be subtracted from the original spectrograms to enable extraction of additional features. Peaks are automatically selected until a target percentage of the original power is accounted for in the extracted modal features. The remaining audio is the *residual* audio, containing background noise and nonlinear effects such as complex modes.

As an additional step, we remove features with  $d_i$  under a threshold. These low-damping features are likely to be a constant pitched background noise unrelated to the impact sound. We also remove features below an amplitude threshold, as they are more susceptible to noise. The extracted  $(\omega_{id}, a_i, d_i)$  features can be converted into pairs of  $(\lambda_i, d_i)$  values for later use. The eigenvalue can be computed by inverting the process in Equation 4:

$$\lambda_i = \omega_{in}^2 = \omega_{id}^2 + d_i^2. \quad (8)$$

As a result of this feature extraction process, we have a set of features roughly corresponding to the modes of vibration of the object.

#### 4.2 External Damping Factors

With the features extracted from the input sounds, we now consider factors that may influence the estimated frequencies or damping values.

Unless an object is flying through the air, there must be something supporting it; the object could be sitting on a desk, held in a hand, or dangling from a ceiling. In this paper, we define a *support* broadly as any long-lasting contact with the sounding object of interest, with enough friction to maintain its contact with the object even when the object is struck. Regardless of the form of support, some energy

from the object’s vibrations will be transferred to the support, causing additional damping. In carefully controlled environments with proper supports, this damping is minimal and can be ignored. In real-world situations where the object is unlikely to be minimally supported, the additional damping significantly affects the sound.

Refer to Figure 4 for an example of the effect of the support on the resulting sound. When the porcelain bowl is gripped tightly with fingers absorbing vibrations of both the upper rim and the lower bowl, the resulting sound is short and heavily damped. If the bowl is instead supported by balancing on fingertips, the same impact produces a much longer-lasting sound.

A real-world object’s modes of vibration may be somewhat *complex* modes of vibration. Intuitively, *normal* modes of vibration are standing waves with consistent nodal points, while *complex* modes of vibration are traveling waves. Modes of vibration for real objects will usually be mostly normal, though a small amount of complex motion is common, and the addition of a support may further increase mode complexity. The damping models described in Section 3.3 cannot fully capture the damping behavior of complex modes. Since most systems only have slightly complex modes, normal modes are a close approximation [13].

Background noise in recorded sounds is too variable to realistically model. The feature extraction step of the method is designed to specifically extract *modal* features from the sound. This mostly eliminates persistent “hums” and “buzzes”, which do not match the exponential-decay model associated with damped modes of vibration. However, background noise may also affect estimates of damping rates for each mode. Acoustic reflections and reverberations from room acoustics are of particular concern. Without knowing the properties of the room acoustics, we cannot separate the effect of a damping material from the effect of the acoustics. For our model, we still assume minimal room reverberations, and we have recorded our sounds in a padded sound booth to minimize these effects. However, some small sources of background noise may be appropriately modeled by a normally distributed effect on damping values. A short sound coincidentally aligned with the start of the impact sound will make the modes with matching frequencies appear to have a higher rate of damping. Conversely, a background “hum” aligned with a particular mode’s frequency may appear to cause a lower rate of damping.

The feature extraction step itself is not perfect; some error is introduced in the process. For example, spectrograms have limited spectral and temporal resolution, and the Fourier transform’s assumption of periodicity in each window is an approximation. The resulting side-lobes may appear as separate peaks, or more likely, affect the estimated damping rate of nearby modes.

Modes do not radiate equally in all directions: the *acoustic radiance* of the object may mean that different microphone placements will result in different initial mode amplitudes. This can be accounted for by keeping the microphone stationary during an impact sound. However, the relative positions of the microphone and object do *not* need to be fixed across all input sounds. Moving the microphone between sounds will change the perceived initial mode amplitudes, but not their frequencies or exponential rate of decay, and thus does not need to be accounted for in our model.

Current damping parameter estimation techniques do not explicitly consider these factors and attribute all damping to the material [30]. This assumption greatly limits the recordings which can be used for accurate damping parameter estimation: they must be recorded in a carefully controlled setting. Furthermore, while these techniques are able to create realistic sound, the estimated values are not solely properties of the material. Instead, they are parameters to a line of best fit which models the combined effect of the material *and the environment*. These parameters may not properly transfer to an object of the same material, but in a different environment. With a thoroughly robust technique that can separately model environmental factors, we can better estimate the true, physically-accurate material parameters.

#### 4.3 Generative Model for Combined Damping

We now introduce a generative model for sampling damping values. The model defines the probability distribution for the extracted damping

value  $d_i$ , given the eigenvalue  $\lambda_i$  and a set of parameters  $\theta$ . This can be written as  $p(d_i|\lambda_i, \theta)$ . Some components of  $\theta$  are the damping parameters:  $\alpha_1$  and  $\alpha_2$  for Rayleigh damping and  $\gamma_1$ – $\gamma_3$  for hybrid damping. These damping parameters will be referred to as  $\theta_d$  for generality, since their quantity and meaning change depending on the chosen damping model. The remaining components of  $\theta$  are parameters of the probability density functions.

The damping value  $d_i$  is the final damping value as reported by the feature extraction step. In the *absence* of any external factors,  $d_i$  would only consist of the material damping value: the damping function  $c$  evaluated at an eigenvalue  $\lambda$  with damping parameters  $\theta$ . To account for the external factors, we model  $d_i$  as a random variable based on the sum of normally and exponentially distributed random variables.

We use a normal distribution to model the effect of some external factors. The normal distribution accounts for the slight complexity of the modes of vibration, the small sources of frequency-aligned background noise, and error in feature extraction due to spectrogram discretization. We assume that each of these factors are an additive, normally distributed random variable. Whether or not these factors are dependent or independent, their sum,  $d_i^n$  is also normally distributed:

$$p(d_i^n|\lambda_i, \theta_d, \sigma) = \mathcal{N}\left(c(\lambda_i, \theta_d), \sigma^2\right)$$

The distribution is centered on the damping function  $c$  evaluated at an eigenvalue  $\lambda$  with damping parameters  $\theta_d$ , with a standard deviation  $\sigma$  resulting from the combination of factors.

We use an exponential distribution to model the effect of the object's support.

$$p(d_i^e|\eta) = \text{Exp}(\eta) = \eta e^{-\eta d_i^e}. \quad (9)$$

$d_i^e$  is the resulting exponential damping resulting from the object's support, while  $\eta$  is the rate parameter of the exponential distribution. Zheng and James defined a model to approximate additional per-mode damping based on contacts with other objects (see Section 4.5) [37]. In this model, the additional damping is primarily a function of products of values in the eigenvector matrix  $\phi$ , with the contact point and coefficient of friction weighting certain modes more heavily. We are not aware of any prior work that has attempted to statistically model the distribution of eigenvector matrix  $\phi$  components, and it is unclear to what extent the geometry of the object impacts this distribution. We analyzed the distribution of  $\phi$  components for synthetic objects. The data failed Kolmogorov-Smirnov goodness-of-fit tests for normal, exponential, and a number of other common distributions ( $p > .05$ ). We have selected the exponential distribution as a way to approximate the support damping values in the absence of a more well-defined model. Although it is an approximation, this captures (1) the  $[0, \infty)$  support of the distribution of support damping values, and (2) the relatively long tail of the exponential distribution as compared to the normal distribution.

These final extracted damping value  $d_i$  can then be modeled as the combination of the normally-distributed factors  $d_i^n$  centered on the evaluated material damping model, and the exponentially-distributed factor  $d_i^e$ . Assuming that the factors are independent, they can be considered as two separate sources of exponential decay:

$$z_i(t) = a_i e^{-d_i^n t} e^{-d_i^e t} \cos(\omega_{id} t) \quad (10)$$

$$= a_i e^{-(d_i^n + d_i^e) t} \cos(\omega_{id} t). \quad (11)$$

The probability density function of the sum of the two random variables  $d_i = d_i^n + d_i^e$  is the convolution of their individual probability density functions. The resulting distribution is an exponentially modified Gaussian (EMG) distribution. EMG distributions have been used extensively in chromatography [12], but have also found uses in other domains. Probability density functions for each of these distributions are plotted in Figure 5. The EMG probability density function is somewhere between that of its two components: more of a Gaussian on one side and more of an exponential on the other. With the modified normal distribution being centered on the evaluated damping function,

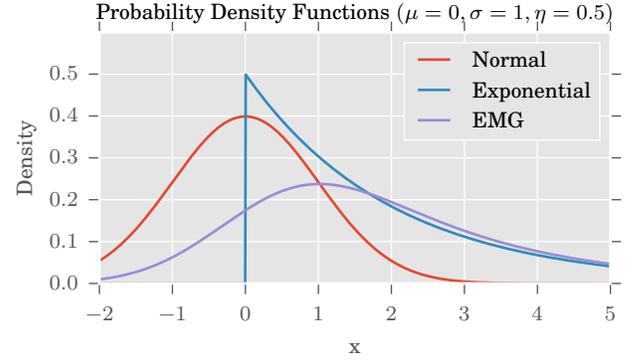


Fig. 5. Probability density functions for the exponentially modified Gaussian (EMG) distribution, which is the combination of a normal and exponential distribution. The EMG distribution behaves like an exponential distribution on the right and a normal distribution on the left, accurately capturing the distribution of extracted damping values for  $d_i$ .

the probability density function for the EMG is:

$$p(d_i|\lambda_i, \theta_d, \sigma, \eta) = \frac{\eta}{2} e^{\frac{\eta}{2}(2c(\lambda_i, \theta_d) + \eta\sigma^2 - 2d_i)} \text{erfc}(s_i) \quad (12)$$

$$s_i = \frac{c(\lambda_i, \theta_d) + \eta\sigma^2 - d_i}{\sqrt{2}\sigma}, \quad (13)$$

where  $\text{erfc}$  is the complementary error function, defined as:

$$\text{erfc}(x) = 1 - \text{erf}(x) = \frac{2}{\sqrt{\pi}} \int_x^\infty e^{-y^2} dy. \quad (14)$$

This defines the probability of observing an extracted damping value, given the material damping and probability distribution parameters. This is the complete generative model for damping values, encompassing multiple sources of damping and error. Since only the modes' frequencies and damping values are needed for this model, we do not need to assume that the mode shapes remain unchanged. The full set of parameters  $\theta$  is  $(\theta_d, \sigma, \eta)$ , which together define the distribution.

#### 4.4 Parameter Estimation

With the generative model established, we now describe the estimation of damping parameters. We estimate the parameters  $\theta$  through maximum likelihood estimation (MLE). The generative model above uses known parameters to produce probability data. MLE is an optimization method that serves to reverse the process: use known data to produce optimal parameters. Given a set of extracted  $(\lambda_i, d_i)$  pairs as data and a set of parameters, we can use the generative model to compute the log-likelihood of the data given the parameters:

$$\log p(d|\lambda, \theta_d, \sigma, \eta) = \sum_i \log\left(\frac{\eta}{2}\right) + \eta c(\lambda_i, \theta_d) + \frac{\eta^2 \sigma^2}{2} - \eta d_i + \log(\text{erfc}(s_i)). \quad (15)$$

Using the log-likelihood simplifies computation, removing exponentiation and turning a product of probabilities into a sum of log probabilities. We want to find the parameters which *maximize* this log-likelihood—and hence also maximize the original probability. These maximizing parameters are those which best explain the extracted data, “fitting” the probability distribution to the data. We compute the analytical gradient of the log-likelihood function and perform gradient ascent to find these optimal parameters.

The gradient is constructed from the partial derivatives with respect to each parameter. We compute the full average derivative for the  $n$   $(\lambda_i, d_i)$  samples. We define a term  $t_i$  to simplify notation:

$$t_i = \frac{-2}{\text{erfc}(s_i)\sqrt{\pi}} e^{-s_i^2} = \frac{-2}{\text{erfc}(-s_i)\sqrt{\pi}}. \quad (16)$$

The derivatives for  $\eta$  and  $\sigma$  must be computed for all damping models. Their derivatives are as follows:

$$\frac{\partial \log p}{\partial \eta} = \frac{n}{\eta} + n\eta\sigma^2 + \sum_i c(\lambda_i, \theta_d) - d_i + t_i \frac{\sigma}{\sqrt{2}}, \quad (17)$$

$$\frac{\partial \log p}{\partial \sigma} = n\lambda^2\sigma + \sum_i t_i \left( \frac{\eta\sigma^2 + d_i - c(\lambda_i, \theta_d)}{\sqrt{2}\sigma^2} \right). \quad (18)$$

The derivatives for  $\theta_d$  will depend on the damping function itself. We will present the derivatives for Rayleigh damping here; derivatives for alternative models are not difficult to compute. For Rayleigh damping’s linear  $c = \alpha_1 + \alpha_2\lambda$  function, the derivatives for  $\alpha_1$  and  $\alpha_2$  are:

$$\frac{\partial \log p}{\partial \alpha_1} = \eta n + \sum_i \frac{t_i}{\sqrt{2}\sigma}, \quad (19)$$

$$\frac{\partial \log p}{\partial \alpha_2} = \sum_i \eta \lambda_i + \frac{t_i \lambda_i}{\sqrt{2}\sigma}. \quad (20)$$

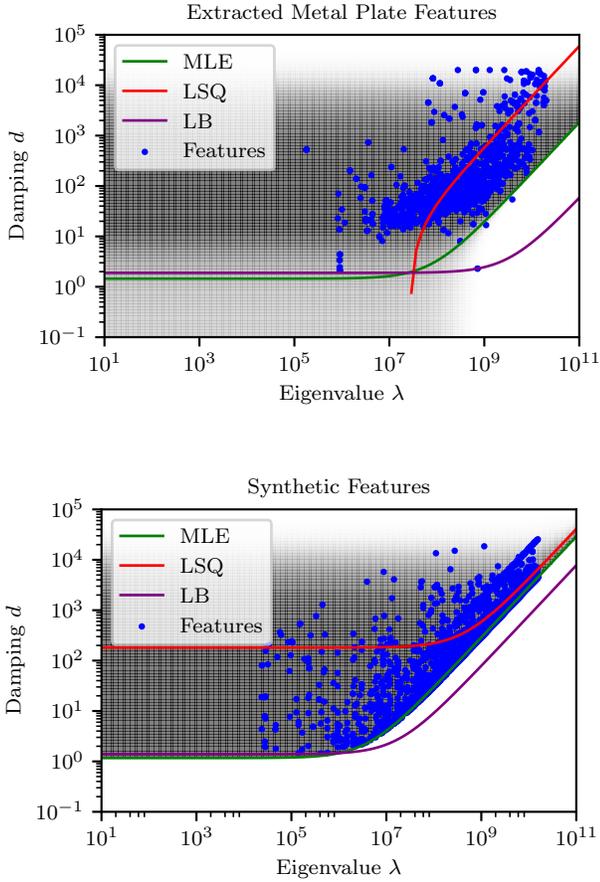


Fig. 6. Parameter estimation on sound features. Each feature consists of an eigenvalue  $\lambda_i$  and its corresponding damping coefficient  $d_i$ . Estimated Rayleigh damping curves are plotted, with the variation of the curve caused by external factors. Our method, MLE, provides the closest fit to the data and is relatively unaffected by features with abnormally high or low damping.

With the derivative established, we can perform standard gradient ascent until convergence. The final damping parameters in  $\theta_d$  are the optimal parameters for the material of the struck object. Figure 6(a) shows features extracted from 19 impact sounds on a metal plate, while Figure 6(b) shows synthetic features generated from modal synthesis (see Section 3.2). The figure shows three estimated material damping curves estimated using least squares fits (LSQ), an lower bound fit (LB) and the maximum likelihood estimation method (MLE, ours). The



Fig. 7. Three objects from our impact sound dataset: a porcelain cup (left), a small glass tile (center), and a wood block (right). Note the ways that each object is supported. These supports interfere with damping parameter estimation.

points that fall along the line are features like any others, but are particularly informative for parameter estimation. When performing sound synthesis for an object of the same material, the extracted damping parameters can be plugged in to recreate the damping effects of that material.

The effect of a support cannot be entirely removed, and in real-world situations the extracted damping values may all be much higher than the material damping function alone. As a result, this estimator is positively biased: assuming the extracted  $(\lambda, d)$  features faithfully represent the input sounds, the estimated parameters will be larger than the ground truth. By accounting for the external factors, this estimator has less bias than, say, a least-squares estimate of the parameters, and is therefore a more accurate estimator. In the case of Figure 4.4, our estimator, MLE, provides the best fit to the majority of the data, while LSQ is heavily affected by increased external damping factors and LB is heavily affected by single features with abnormally low damping. See supplementary material for additional examples.

#### 4.5 Sound Synthesis with Estimated Values

The estimated damping parameters should be relatively accurate, having accounted for the effect of the support. However, using them directly in modal synthesis may produce sounds which are slightly too reverberant. Modal sound synthesis assumes free vibrations (i.e. no support) when in most cases there will be something supporting the object. An additional step is needed to apply support damping to synthesize contact sounds due to support.

We adopt a contact model for modal sound synthesis introduced by Zheng and James [37]. This model is physically-based and uses the contact forces and modal displacements at each timestep to generate realistic viscous support damping values. The method uses an additional damping matrix  $G$  to model the additional damping resulting from each contact point  $k$  in the set of contact points  $\mathcal{C}$ :

$$G = \sum_{k \in \mathcal{C}} c_k \Phi_k^T (\mu \mathbf{I} + (1 - \mu) \mathbf{n}_k \mathbf{n}_k^T) \Phi_k, \quad (21)$$

where  $c_k$  is the magnitude of the contact force for contact  $k$ ,  $\Phi_k$  is the set of eigenvectors corresponding to the point at contact  $k$ ,  $\mu$  is the coefficient of friction, and  $\mathbf{n}_k$  is the normal direction at that point. Some mode coupling is introduced since  $G$  is not diagonal, but this coupling was found to be perceptually minor. Therefore, each damping model may be augmented by adding the corresponding diagonal component of  $G$ .

## 5 RESULTS

We have implemented the audio material reconstruction algorithm described in this paper and tested its effectiveness through both numerical analysis and perceptual validation.

To verify its effectiveness, we have created a new dataset of impact sounds on rigid objects, where the hit point and method of support are documented for each impact. Fifteen objects covering a broad range of rigid materials were struck to produce the impact sounds in this dataset. Overall, there are 698 impact sounds sampled from these fifteen objects. These impact sounds were designed to be representative of possible real-world support situations, so all objects were supported by hand,

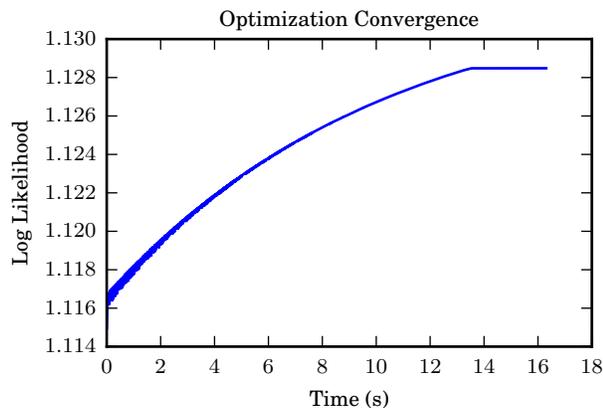


Fig. 8. Plot of log-likelihood maximization converging over the course of parameter estimation. Optimization was performed on 752 frequency-damping points extracted from porcelain plate impact sounds, and converged after 39,009 iterations for a total of 16.3s in a one-time preprocessing.

often either with an edge being pinched between two fingers or the center resting on a few fingertips.

Audio was recorded using a Zoom H4 in a padded sound recording booth which reduced, but did not eliminate, acoustic effects and background noise. Objects were struck with a small metal wrench, the wrench itself being tightly gripped to minimize its own vibrations. Figure 7 shows a sample of these objects, with various hit locations and methods of support.

We implemented the parameter optimization algorithm in Python and NumPy. On a laptop with a dual core 2.53 GHz Intel Core i5-540M processor, optimization over thousands of features from tens of input sounds and hundreds of thousands of iterations takes 1-5 minutes to complete. In many cases, convergence is achieved within 20 seconds. See Figure 8 to see an example of convergence behavior. Note that we are attempting to *maximize* the log-likelihood, as the parameters which maximize the log-likelihood also maximize the underlying probability. However, the log-likelihood values themselves are not as intuitively understood as probability, for example, it is not bound between 0 and 1.

The eigenvalues and damping values are each normalized, but the data are not shifted or centered. With this normalization, the estimated damping values need to be unnormalized for application to other materials. Although we cannot guarantee that the optimization problem in this context is always convex, especially for higher order damping functions, in multiple runs from different starting points on multiple datasets, all optimization processes have converged on the same parameters. The optimal value of  $\sigma$  tends to be very small, indicating that the distributions of damping values tend to be closer to exponential distributions than to normal distributions.  $\sigma$  and  $\eta$  are used to guide optimization of the damping parameters, but they are not needed for sound synthesis.

Table 1 contains results from extraction on some of the objects in the dataset. As previously mentioned, when  $\gamma_3 = 1$ , the model is identical to Rayleigh damping. Even small changes in  $\gamma_3$  can have a large impact on the resulting damping. For example, a 10 kHz mode on the Porcelain plate has a damping coefficient  $d = 20$  with the provided parameters ( $\gamma_3 = 1.027$ ), but changing  $\gamma_3$  to exactly 1 reduces the damping coefficient to  $d = 12$ .

In general for these damping models, larger parameters create virtual materials with more damping and shorter sounds. For example, the two objects with the most damping are the wood block and plastic bowl, whose materials are known to be naturally heavily damped. The porcelain plate, travertine tile, and glass tile all had similar estimated parameters, reflecting their similar molecular compositions.

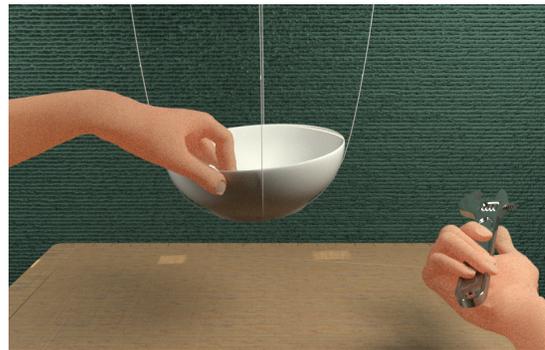


Fig. 9. A simulated porcelain bowl is struck in multiple locations, with and without a supporting grip.

## 5.1 Real-time Synthesis and Rendering

Each sounding object must be preprocessed (see Section 3.1) before running any interactive application. Preprocessing time depends primarily on the number of tetrahedra in the input mesh; a mesh with 2,000 tetrahedra takes under a minute to preprocess while a mesh with 30,000 tetrahedra can take many minutes. Once each sounding object has been preprocessed, modal synthesis is performed in real time at 44 kHz. The supplementary video contains example videos and sounds generated using our method.

Like previous work [26], we are able to synthesize sound using an interactive rigid-body physics simulation in real time. We have implemented our method for sound synthesis with support damping in C++ as a module for Unreal Engine 4. Figure 1 shows multiple scenes from our real-time demo, with multiple objects of various shapes and materials. Our demos have been integrated with an HTC Vive headset and Leap Motion controller. The user’s hands were tracked with the Leap Motion, with the Vive controllers used to represent tools that could be picked up and used to strike objects. Users can walk in the virtual environment and strike objects, immediately hearing the resulting synthesized sound. Figure 9 shows another scene, where a bowl is supported by either strings or a hand, producing different sounds depending on the hit point and support type.

## 5.2 Comparisons with Ren et al. [26]

The most closely-related prior work to our own is that of Ren et al. [26], which operates by minimizing the difference between a recorded sound and a synthesized sound while optimizing the parameters for synthesized sound. Both works attempt to estimate material damping parameters using recorded sounds as input, but Ren et al. use a somewhat different set of inputs and outputs. Ren et al. use a single recorded sound, the known geometry and size of the object, and the hit location on the object to produce the material damping parameters and ratio of Young’s modulus to density. Our work can use one sound but multiple recorded sounds improve the results. It does not assume knowledge of geometry, size, or hit location to produce the material damping parameters.

The algorithm by Ren et al. [26] and its effectiveness heavily depends on the implementation. A complete copy of the code by Ren et al. [26] is not available to perform numerical comparisons with this work. As a result, direct comparisons between re-synthesized sound are infeasible. However, a list of estimated parameters for a number of materials is given in their paper [26] and a set of input audio recordings are available in its supplementary video. Significant variation exists within each category of material, e.g. our wood block is not directly comparable to the wooden plate used in previous work. Perceptual comparisons against Ren et al. are performed in the auditory perception user study (Section 6). The following synthetic validation study evaluates accuracy instead, by comparing against other possible alternative methods.

## 5.3 Synthetic Validation

Synthetic validation provides a numerical comparison against ground-truth damping parameters. We synthesized a variety of sounds with

		Porcelain Plate	Travertine Tile	Wood Block	Steel Wrench	Plastic Bowl	Glass Tile
<b>Rayleigh</b>	$\alpha_1$	3.9	1.3	39.0	2.3	39.8	2.0
	$\alpha_2$	1e-8	2.5e-8	1.3e-7	6.9e-8	1.3e-7	7.8e-8
<b>Hybrid</b>	$\gamma_1$	3.9	1.3	39.0	2.4	34.83	1.9
	$\gamma_2$	5.2e-9	2.5e-8	2.1e-7	5.5e-8	4.1e-7	1.5e-7
	$\gamma_3$	1.027	1.001	0.978	1.011	0.95	0.974

Table 1. Damping parameters estimated using our technique. The listed objects are a subset of those used in our impact sound dataset. These parameters are described in Section 3.3. When  $\gamma_3 = 1$ , the remaining hybrid damping parameters are equivalent to their Rayleigh damping counterparts. These parameters can be used to virtually recreate the material of the real-world object.

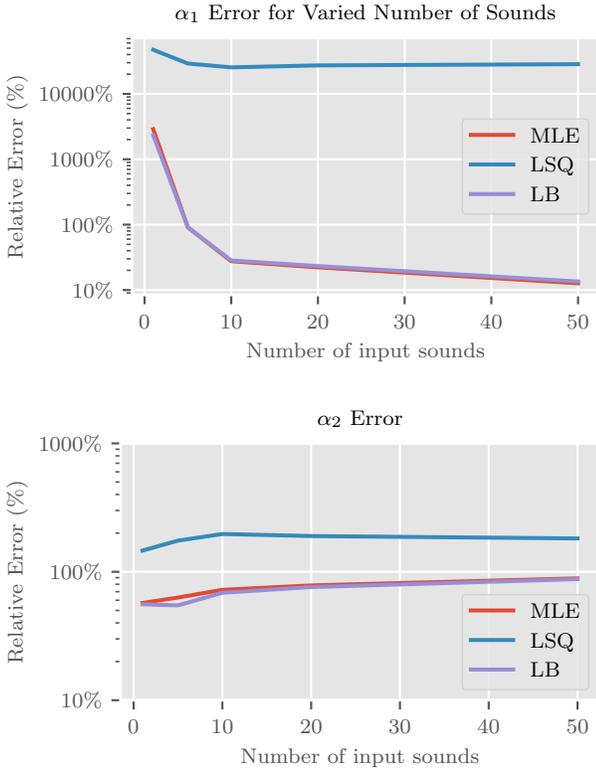


Fig. 10. As the number of sounds used for parameter estimation increase, error in Rayleigh damping parameter  $\alpha_1$  decreases while error in  $\alpha_2$  does not change much. Relative error for  $\alpha_1$  is low, while error for  $\alpha_2$  is somewhat overfit. Overall, our method outperforms the other two alternatives.

known damping parameters and passed the resulting sounds through the parameter estimation process to see if the original ground-truth values could be recovered using our algorithm. Sounds were synthesized from the geometry of 18 models, ranging from small, hollow cups to desktop vases and large sculptures. Five materials were chosen by randomly sampling material parameters from a range of realistic values. For each object, ten support points were sampled at random on the surface of the object, each with a random amount of contact force ranging from a light support to a moderate support. Then, 100 sounds were synthesized for each combination of object and material. Each sound sampled its impact point randomly on the exterior surface and picked one support point to be active. The resulting sounds were passed through the feature extraction process for Rayleigh damping, and extracted features from a varying number of sounds were used to estimate the original parameters.

Parameter estimation was performed with three different estimators:

- Our maximum-likelihood method (MLE),
- A least squares fit to the data (LSQ), and
- A lower bound fit to the data by selecting a line along the lower

convex hull (LB).

We compared the error between the ground-truth parameters and the estimated parameters while using a varying number of input sounds. For each tested number of impact sounds, 30 different sets of sounds of that cardinality were sampled, and the resulting errors averaged.

### 5.3.1 Discussion

Figure 10 shows the relative error for each parameter and each estimator. For all materials in this synthetic data, both the MLE and LB estimators significantly outperformed the LSQ estimator ( $p < .05$ ). With real-world data, the MLE and LB estimators more frequently decouple, as the MLE estimator’s statistical model better adapts to noise and other artifacts of recording. These synthetic sounds without noises and other effects are the ideal situation for the LB estimator, and do not leverage the full capabilities of the MLE estimator.

$\alpha_1$  scales well to larger amounts of data, but  $\alpha_2$  demonstrates some overfitting and relatively higher error. There are multiple sources of bias that could introduce error, for these purely synthetic sounds. The effect of the support cannot be entirely eliminated, and as a result the estimators often slightly overestimate the material parameters. The MLE estimator expects a certain amount of normally-distributed error, and when the data is unrealistically clean (as it is in synthetic data) the parameters may be inaccurate. The feature extraction step introduces its own error which is more difficult to quantify as positive or negative bias. Therefore, some error may be unavoidable, and the amount of unavoidable error is difficult to quantify.

While the error in  $\alpha_2$  is relatively higher, no prior work has performed a similar validation for comparison. Prior work uses optimization metrics based on the squared difference between spectrograms. This would produce results most similar to the LSQ estimator, which was outperformed by our method. In light of this, our method provides a significant benefit when input sounds are affected by external factors, while not requiring any information about object geometry. Finally, the error is mostly important as it affects users’ perception of the material. The perceptual evaluation below provides an analysis of whether our estimated parameters are accurate when evaluated by humans.

## 6 PERCEPTUAL EVALUATION

Numerical comparisons against previous work are difficult since our method is the first work to estimate damping parameters given only input audio with no knowledge of geometry, size, or hit point. The closely related work of Ren et al. [26] (see Section 5.2) performs a comparison based on auditory perception of material between real and re-synthesized audio material. To best compare against their work, we perform a similar perceptual validation study to determine if there is a perceptually noticeable and practical difference between estimated parameters.

This study was designed to compare the perceived materials of objects to one another based on their impact sounds. In a comparison between impact sounds on two similar materials, subjects should be able to recognize that the sounds come from the same material, even if the geometry is different. Users should also be able to notice a discrepancy when presented with sounds that come from objects with different materials. If our material parameters are sufficiently accurate, sounds synthesized from objects with the same material should sound more similar to one another than sounds synthesized from objects with different materials. Virtual objects of a certain material should also

		Rayleigh	Hybrid	Ren
Same	$\mu$	6.8	6.7	6.0
	$\sigma$	3.0	3.0	3.0
Diff	$\mu$	5.1	5.4	3.5
	$\sigma$	3.0	3.3	2.7
Stats	$p$	.007	.013	4e-6
	$d$	.56	.39	.93

Table 2. Population information for comparisons with a matching (same) material or a different (diff) material. The “same” comparisons are significantly higher than the “diff” comparisons at the  $p < 0.05$  level. Cohen’s effect size  $d$  is provided for each parameter set. Overall our estimated material parameters produced similar, if not slightly better, results than those by the state of the art [26].

compare favorably to sounds of the real-world object that was struck to estimate the material’s parameters.

## 6.1 Experimental Setup

The study was conducted online through a web questionnaire, and subjects were asked to wear headphones or earbuds to ensure a consistent and high-quality auditory environment. 39 subjects participated in the study, and more specific demographic information was not collected.

Each subject listened to 20 pairs of impact sounds, comparing one real-world sound to a real or synthesized sound. (The supplementary zip files userstudy subdirectory contains the sounds from the user study, the user study results, and Python code to analyze the results.) The subjects were asked to rate the similarity of the *materials* of the impacted objects on a scale from 1 to 11, where 1 indicates a low similarity and 11 indicates a high similarity. Since perception of sound is highly subjective and each subject evaluates “similarity” differently, the subjects were shown training examples alongside images of the objects, with text explaining why one pair of sounds is highly similar while another pair is not. These training examples may have introduced slight bias, but this task is challenging for untrained listeners, and based on prior work [25, 26] we believe that the benefit gained from training outweighs the possible bias introduced. The subjects were expected to (possibly subconsciously) use their personal similarity metric to determine their choice of rating, and they were allowed to re-rate previous pairs as they revise their preferences.

The recorded real-world sounds were taken from our impact sound dataset by selecting one sound from five objects of different materials. The synthesized sounds were created on two virtual objects: a large vase and a decimated Stanford bunny. On each object, sound was synthesized for each material category using estimated hybrid and Rayleigh damping parameters estimated by our method, and estimated Rayleigh damping parameters from the work of Ren et al. [26]. In total, there were 30 synthesized sounds, although each subject listened to only a subset chosen at random, each compared against a random recorded sound—in a similar setup as described by [25].

## 6.2 Results and Analysis

First, we divide the results into two categories. In “same material” comparisons, the two provided sounds are made of the same material, and ideally should be rated with high similarity. In “different material” comparisons, the two sounds are made of different materials, which may sound similar in some cases (e.g. plastic vs. wood) and different in others (e.g. wood vs. glass). We test the hypothesis that the ratings for the “same material” comparisons have a different mean than the ratings for the “different material” comparisons, assuming normally distributed populations. During study design, we did not assume that the number of comparisons or their variances would be identical across categories. Therefore, these categories are compared using a two-sample Welch’s unequal variances t-test. See Table 2 for population and statistical data.

For each of the three material sets—our Rayleigh parameters, our hybrid parameters, and the Rayleigh parameters from Ren et al.—subjects rated the similarity of “same material” comparisons significantly higher

( $p < 0.05$ ) than they rated the similarity of “different material” comparisons. This indicates that synthesized sounds tended to sound more like the material they were based on. Each material set’s “same material” and “different material” comparisons had high standard deviations. This is to be expected: perception of sound can vary significantly between individuals. However, in aggregate the subjects demonstrated they could tell the difference between “same” and “different” material comparisons despite the large variance.

We now look specifically at similarity results for “same material” comparisons, where a real recording is compared to a synthesized sound on an object of the same material. A consistently high similarity score here indicates that the reconstructed material parameters accurately capture the audio characteristics of the real-world sounding materials. The mean similarity score was 6.83 out of 11 for Rayleigh parameters using our methods on “same material” comparisons, 6.7 for our hybrid parameters, and 6.05 for Ren et al.’s estimated parameters [26]. The difference between these scores is not statistically significant as determined by one-way ANOVA ( $p > 0.05$ ). This indicates that our method performs at least equally well (and possibly better) as the prior state of the art at reconstructing the object’s audio materials, despite the rather limited input information.

### 6.2.1 Discussion of Perceptual Study

1. In general, our estimated material parameters produced similar, if not slightly better, results than those of the state of the art.
2. Synthesized sounds across all material sets were consistently reported to be most similar to objects of the material from which they were derived.
3. In each of the tests, none of the material sets were statistically more accurate than another set.

In contrast to prior work [26], our estimation technique does not require knowledge of the geometry or the size of the struck object, the hit location, or the mechanism of support. Therefore, we are able to produce synthesized sound of similar quality with little assumption on the input information. It is able to automatically take into account the additional damping due to support without needing an elaborate setup to minimize the effects of support.

## 7 CONCLUSION

We have presented a method for estimating material damping parameters using only recorded impact sounds as input. We have validated these contributions through parameter estimation on a new dataset of impact sounds on rigid objects, using both an auditory user study and synthetic validation. These methods can extract real-world material parameters from audio recording and recreate virtualized materials and their rich sound effects arising from dynamic interaction in virtual environments.

**Limitations and Future Work:** For parameter estimation, there exists a tradeoff between the amount of assumptions on the required inputs and the quality of outputs. This work assumes that only a given set of audio recordings would be provided. We do not assume prior knowledge on the object geometry, size, material parameters, or the impact location—just audio is sufficient. However, this technique cannot estimate Young’s modulus, Poisson’s ratio, density, or geometric properties of the object. The damping parameters estimated are the most difficult to obtain through alternative means (e.g. values from handbooks), but a method that can optimize all parameters simultaneously simplifies the pipeline from recording to synthesis. Future work may explore if some small amount of additional inputs can result in a much greater increase in the number of estimated parameters.

There are some limitations of the method. A single sound is not enough to estimate parameter  $\alpha_1$  with sufficient accuracy; upwards of 10–20 sounds may be needed. In synthetic validation, estimated  $\alpha_2$  values had relatively larger error, but it is unclear to what extent this is a limitation of the method or a fundamental limitation of the external factors allowed in the input sounds.

The damping parameters estimated with our technique may be useful as input to estimation of other parameters. Deep learning may also be able to extend the predictive power of a small number of inputs, identifying patterns and correlations that are otherwise difficult to discover.

Finally, the recording and feature extraction process produces features which are not fully independent and identically distributed. Further evaluation and design of feature extraction algorithms may make the features more independent, while mixture models may help account for spurious features.

## REFERENCES

- [1] S. Adhikari. *Damping models for structural vibration*. PhD thesis, University of Cambridge, 2001.
- [2] S. Adhikari. Damping modelling using generalized proportional damping. *Journal of Sound and Vibration*, 293(12):156–170, 2006. doi: 10.1016/j.jsv.2005.09.034
- [3] S. Adhikari and J. Woodhouse. Identification of damping: Part 1, viscous damping. *Journal of Sound and Vibration*, 243(1):43–61, 2001. doi: 10.1006/jsvi.2000.3391
- [4] G. Bharaj, D. I. W. Levin, J. Tompkin, Y. Fei, H. Pfister, W. Matusik, and C. Zheng. Computational design of metallophone contact sounds. *ACM Trans. Graph.*, 34(6):223:1–223:13, Oct. 2015. doi: 10.1145/2816795.2818108
- [5] N. Bonneel, G. Drettakis, N. Tsingos, I. Viaud-Delmon, and D. James. Fast modal sounds with scalable frequency-domain synthesis. *ACM Transactions on Graphics (SIGGRAPH Conference Proceedings)*, 27(3), August 2008.
- [6] T. Caughey. Classical normal modes in damped linear dynamic systems. *Journal of Applied Mechanics*, 27(2):269–271, 1960.
- [7] T. Caughey and M. Okelly. Classical normal modes in damped linear dynamic systems. *Journal of Applied Mechanics*, 32(3):583–588, 1965.
- [8] J. N. Chadwick, C. Zheng, and D. L. James. Precomputed acceleration noise for improved rigid-body sound. *ACM Transactions on Graphics (Proceedings of SIGGRAPH 2012)*, 31(4), Aug. 2012.
- [9] P. R. Cook and G. P. Scavone. The synthesis toolkit (stk). In *In Proceedings of the International Computer Music Conference*, 1999.
- [10] E. Coumans. Bullet physics simulation. In *ACM SIGGRAPH 2015 Courses*, SIGGRAPH '15. ACM, New York, NY, USA, 2015. doi: 10.1145/2776880.2792704
- [11] A. Davis, K. L. Bouman, J. G. Chen, M. Rubinstein, O. Bykztrk, F. Durand, and W. T. Freeman. Visual vibrometry: Estimating material properties from small motions in video. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(4):732–745, April 2017. doi: 10.1109/TPAMI.2016.2622271
- [12] E. Grushka. Characterization of exponentially modified gaussian peaks in chromatography. *Analytical Chemistry*, 44(11):1733–1738, 1972. PMID: 22324584. doi: 10.1021/ac60319a011
- [13] M. Imregun and D. J. Ewins. Complex Modes - Origins and Limits. In *Proceedings of the 13th International Modal Analysis Conference*, vol. 2460, p. 496, 1995.
- [14] D. L. James, J. Barbič, and D. K. Pai. Precomputed acoustic transfer: output-sensitive, accurate sound generation for geometrically complex vibration sources. In *ACM Transactions on Graphics (TOG)*, vol. 25, pp. 987–995. ACM, 2006.
- [15] M. Kac. Can one hear the shape of a drum? *The American Mathematical Monthly*, 73(4):1–23, 1966.
- [16] K. Karplus and A. Strong. Digital synthesis of plucked-string and drum timbres. *Computer Music Journal*, 7(2):43–55, 1983.
- [17] T. R. Langlois, S. S. An, K. K. Jin, and D. L. James. Eigenmode compression for modal sound models. *ACM Transactions on Graphics (Proceedings of SIGGRAPH 2014)*, 33(4), Aug. 2014. doi: 10.1145/2601097.2601177
- [18] D. Li, Y. Fei, and C. Zheng. Interactive acoustic transfer approximation for modal sound. *ACM Trans. Graph.*, 35(1), 2015. doi: 10.1145/2820612
- [19] D. B. Lloyd, N. Raghuvanshi, and N. K. Govindaraju. Sound synthesis for impact sounds in video games. In *Proceedings of the Symposium on Interactive 3D Graphics and Games 2011*. ACM, 2011.
- [20] A. D. Nashif, D. I. Jones, and J. P. Henderson. *Vibration damping*. John Wiley & Sons, 1985.
- [21] J. F. O'Brien, C. Shen, and C. M. Gatchalian. Synthesizing sounds from rigid-body simulations. In *Proceedings of the 2002 ACM SIGGRAPH/Eurographics Symposium on Computer Animation*, SCA '02, pp. 175–181. ACM, New York, NY, USA, 2002. doi: 10.1145/545261.545290
- [22] D. K. Pai, K. v. d. Doel, D. L. James, J. Lang, J. E. Lloyd, J. L. Richmond, and S. H. Yau. Scanning physical interaction behavior of 3d objects. In *Proceedings of the 28th annual conference on Computer graphics and interactive techniques*, pp. 87–96. ACM, 2001.
- [23] N. Raghuvanshi and M. C. Lin. Interactive sound synthesis for large scale environments. In *Proceedings of the 2006 Symposium on Interactive 3D Graphics and Games*, I3D '06, pp. 101–108. ACM, New York, NY, USA, 2006. doi: 10.1145/1111411.1111429
- [24] J. W. S. B. Rayleigh. *The theory of sound*, vol. 2. Macmillan, 1896.
- [25] Z. Ren, H. Yeh, R. Klatzky, and M. C. Lin. Auditory perception of geometry-invariant material properties. *Visualization and Computer Graphics, IEEE Transactions on*, 19(4):557–566, 2013.
- [26] Z. Ren, H. Yeh, and M. C. Lin. Example-guided physically based modal sound synthesis. *ACM Trans. Graph.*, 32(1):1:1–1:16, Feb. 2013. doi: 10.1145/2421636.2421637
- [27] G. P. Scavone and P. R. Cook. Rtmidi, rtaudio, and a synthesis toolkit (stk) update. In *In Proceedings of the International Computer Music Conference*, 2005.
- [28] X. Serra and J. Smith. Spectral modeling synthesis: A sound analysis/synthesis system based on a deterministic plus stochastic decomposition. *Computer Music Journal*, 14(4):12–24, 1990.
- [29] J. C. Slater, W. Keith Belvin, and D. 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*, pp. 1508–1508. SPIE INTERNATIONAL SOCIETY FOR OPTICAL, 1993.
- [30] A. Sterling and M. C. Lin. Interactive modal sound synthesis using generalized proportional damping. In *Proceedings of the 20th ACM SIGGRAPH Symposium on Interactive 3D Graphics and Games*, I3D '16, pp. 79–86. ACM, New York, NY, USA, 2016. doi: 10.1145/2856400.2856419
- [31] T. L. Szabo. Time domain wave equations for lossy media obeying a frequency power law. *The Journal of the Acoustical Society of America*, 96(1):491–500, 1994. doi: 10.1121/1.410434
- [32] K. van den Doel and D. K. Pai. The sounds of physical shapes. *Presence*, 7:382–395, 1996.
- [33] B. Wang, L. Wu, K. Yin, U. Ascher, L. Liu, and H. Huang. Deformation capture and modeling of soft objects. *ACM Trans. Graph.*, 34(4):94:1–94:12, July 2015. doi: 10.1145/2766911
- [34] J. Woodhouse. Linear damping models for structural vibration. *Journal of Sound and Vibration*, 215(3):547–569, 1998.
- [35] K. Yamamoto and T. 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, pp. 231–240. Eurographics Association, Aire-la-Ville, Switzerland, Switzerland, 2016.
- [36] S. Yang and M. C. Lin. Materialcloning: Acquiring elasticity parameters from images for medical applications. *IEEE Transactions on Visualization and Computer Graphics*, 22(9):2122–2135, Sept 2016. doi: 10.1109/TVCG.2015.2505285
- [37] C. Zheng and D. L. James. Toward high-quality modal contact sound. *ACM Transactions on Graphics (Proceedings of SIGGRAPH 2011)*, 30(4), Aug. 2011.