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

Auston Sterling, Nicholas Rewkowski, Roberta L. Klatzky, and Ming C. Lin

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 from recorded impact sounds that probabilistically models these external factors. We represent the combined effects of material damping, support damping, and sampling inaccuracies with a probabilistic generative model, then use maximum likelihood estimation to fit a damping model to recorded data. This technique greatly reduces the human effort needed and does not require the precise object geometry or the exact hit location. We validate the effectiveness of this technique with a comprehensive analysis of a synthetic dataset and a perceptual study on object identification. We also present a study establishing human performance on the same parameter estimation task for comparison.

**Index Terms**—Damping modeling, sound synthesis, modal analysis, statistical modeling

## 1 Introduction

Interactive virtual environments are more effective when they maintain a strong sense of immersion. To preserve immersion, 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 [28]. Modal sound synthesis improves a user's immersion, 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. 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.

Traditionally, material parameters are selected through laborious human hand-tuning. We present a study evaluating human efficiency and precision at this task in section Sect. 5.2. Even with a simple, easy-to-use GUI optimized to minimize during modal analysis, the study shows that significant human effort is needed to select accurate parameters. The study also finds that humans are able to distinguish between sounds with minor differences in material parameters, suggesting that material parameters from a library may not sufficiently reproduce the sound of a specific real-world object.

Automated material parameter estimation provides a means to estimate the material parameters of a specific object while reducing required human effort. 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 mandatory knowledge about the shape and properties of the struck object, to estimate a number of material parameters [33]. The material parameters can be applied to sound synthesis of any virtual object, "virtualizing" the audio characteristics of a given material in the physical world. While recent techniques have been able to estimate material damping properties, they assume minimal 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. The interface between the object and its *support* will introduce additional damping. To account for this support damping, recordings must be made with supports that introduce minimal damping, requiring a carefully controlled recording environment using special support [29], e.g. strings or rubber bands, to suspend the object [32]. 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. Satisfying all of the assumptions made by prior work requires significant human effort.

In this paper, we present a practical and efficient probablistic algorithm to estimate material damping properties directly from recorded impact sounds that accounts for these different factors affecting damping, reducing their effects on the estimated parameters. Unlike previous work [32], this method is fast and requires no prior knowledge about the recorded object's geometry, size, or hit location(s). We are able to virtualize the specific materials of a given object in the real world and easily transfer these audio parameters to other synthetic objects in the virtual scene. Our method requires significantly less human time and

- *Auston Sterling is with UNC Chapel Hill. E-mail: austonst@cs.unc.edu.*
- *Nicholas Rewkowski is with UNC Chapel Hill. E-mail: nrewkows@live.unc.edu.*
- *Roberta L. Klatzky is with Carnegie Mellon University. E-mail: klatzky@cmu.edu.*
- *Ming C. Lin is with University of Maryland. E-mail: lin@cs.umd.edu.*

*Manuscript received xx xxx. 201x; accepted xx xxx. 201x. Date of Publication xx xxx. 201x; date of current version xx xxx. 201x. For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org. Digital Object Identifier: xx.xxxx/TVCG.201x.xxxxxxx*

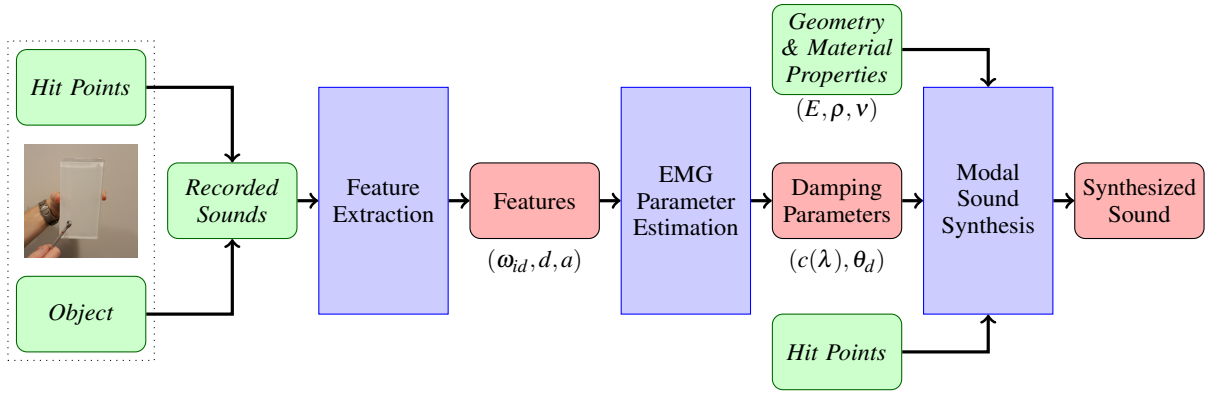


Fig. 2. Our pipeline for estimating material parameters from recorded audio and using the parameters to synthesize sound for objects of the same material. Inputs are in green with italic text. If the object and hit points are unknown, the pipeline can begin with recorded sounds instead.

effort to acquire material damping parameters than previous methods, while producing parameters of similar quality. The key contributions of this work include:

- A new probabilistic material damping model that independently considers each source of damping (Sect. 4.4);
- Application of this probabilistic model to acquisition of material damping parameters for synthesizing virtual sounds (Sect. 4.5);
- A study evaluating human effectiveness at manual estimation of material parameters from recorded audio clips (Sect. 5.2); and
- Quantitative (Sect. 5.3) and perceptual (Sect. 5.4) evaluation of captured damping parameters for virtualized sounds.

We validate our method through comparison between estimated and ground-truth damping values, an auditory perceptual study, and comparison against alternative techniques. Fig. 1 demonstrates our system in several complex virtual environments consisting of real-time interaction with virtual objects of different materials. Figure 2 shows the full pipeline for estimating material parameters and using them to synthesize sound.

## 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. For offline applications, wave-based methods produce high-quality sound [41], though in this paper we focus on real-time methods for interactive applications. Strings and drums can be simulated through physical models, such as the Karplus-Strong algorithm [20] and digital waveguide synthesis [36].

For simple objects with known analytical modes of vibration, the frequencies of their modes can be used to synthesize impact sounds [39]. 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 [10, 28]. Interactions between a sounding object and a striking tool can be modeled to better simulate the attack of the sound [1, 4]. Simulation of acoustic radiation can help spatialize synthesized modal sound [18], and can be approximated for interactive applications [23]. Acceleration of synthesis can be achieved by culling modes based on perceptual metrics [30], by performing efficient vectorization [40], through parallelism on CPUs and GPGPUs [6], and by exploiting geometric symmetries [22].

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 [5]. A contact model can be used to modify the damping matrix for sound synthesis based on forces applied by other objects [44].

**Damping** has long been a concern in the construction of buildings and other structures [27], but also plays a significant role in modal sound synthesis. There are a number of ways to model the physical phenomena behind damping to varying degrees of accuracy [35, 42]. Various *damping models* describe the damping of an object as a function of its mass and stiffness [2].

**Statistical modeling of sound** has found applications in summarizing and analyzing sound. The late reverberations of sounds in rooms have been modeled as Gaussian noise, whose summary statistics convey properties of the environment [38]. It has also been found that humans inherently use summary statistics to understand sounds [26].

**Studies on human perception of material in sound** provide important clues about perceptually important parameters. Studies have evaluated which parameters humans rely on for material identification, finding that damping rate and frequency (i.e. pitch) are particularly important [14, 21, 25]. Similar studies have focused on perception of object size from sound [13, 15]. Material perception is also affected by concurrent visual stimuli [12].

**Estimation of material properties** can be performed experimentally with specialized measurement equipment [11]. Impact sounds convey information about an object's vibrations, but it is currently impossible to fully reconstruct an object from its sound without sufficient constraints on the problem [19]. Previous analysis/synthesis techniques model deterministic features or modes of input sounds, then apply random variation to create plausible synthetic sounds [24, 34]. 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 [29]. Alternatively, the Young's modulus for small parts of the object can be individually optimized to best recreate the input sounds [43].

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 [33]. This technique has been extended to support optimization over arbitrary damping models [37]. However, both 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.

## 3 Modal Sound Synthesis

As an object vibrates, its surface deforms and oscillates. These oscillations produce pressure waves in the surrounding air 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 briefly review 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 review only the most relevant details in this section. A more detailed explanation can be found in supplementary material Section 1, but complete derivations can be found in prior work [28, 37]

Modal sound synthesis assumes that an object's vibrations can be broken down as a linear combination of its *modes of vibration*. Each mode of vibration $i$ describes vibrations at a natural frequency $\omega_{in}$ with an exponential damping rate $d_i$. Each object has a different set of modes depending on the object's shape and material. Modal sound synthesis uses this information about shape and material to perform an eigenanalysis of the object to determine the set of frequencies. On the other hand, the damping rate for each mode is determined as a function of that mode's frequency and a set of *material damping parameters*.

### 3.2 Material Damping Modeling

There are a number of *material damping models* that model the relationship between frequency and damping. These model damping of vibrations as a result of the object's material, not its geometry. While many damping models are possible, only a subset of 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 [31]. In this paper, we refer to the values $c_i = 2d_i$, as $c$ arises more naturally from the mathematical formulation and is more commonly used in related work. Once the frequencies of the object's vibration are known, the damping rate for each mode with natural frequency $\omega_{in}$ is:

$$c_i = \alpha_1 + \alpha_2 \omega_{in}^2. \tag{1}$$

In Caughey damping, each $c_i$ is a polynomial function of $\omega_{in}^2$ to an arbitrary degree [7, 8]. Generalized Proportional Damping (GPD) is an even more general model, where each $c_i$ can be an arbitrary function of $\omega_{in}^2$ as long as the function is analytical and continuous near each $\omega_{in}^2$ [3].

We consider one additional GPD-derived damping model: a hybrid model incorporating Rayleigh damping and a power law damping model [37]. The damping rates are described according to the function:

$$c_i = \gamma_1 + \gamma_2 \omega_{in}^{2\gamma_3}. \tag{2}$$

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 [37], 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 for a wide range of geometries under the Rayleigh damping model [32]; 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 solid cube with those parameters will produce a metallic sound. However, the geometry-invariance assumption has only been thoroughly tested on thick, very rigid objects [32], and the assumption may fail for thin-shelled objects [9], less rigid objects, objects with loosely-coupled points of self-collision, or objects demonstrating nonlinear vibrational behavior.

## 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. 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 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.

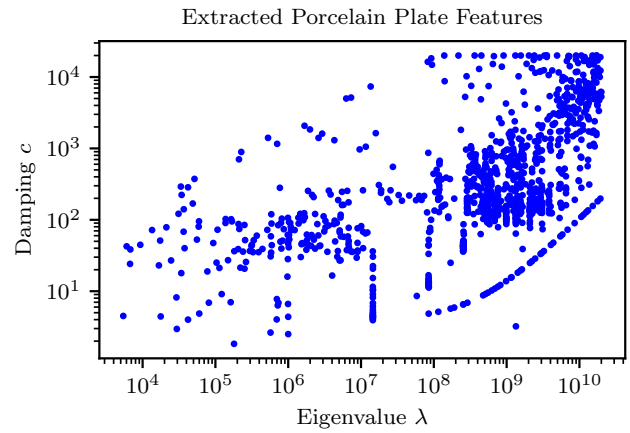


Fig. 3. Features extracted from multiple impact sounds on a porcelain plate. $\lambda$ is the eigenvalue of the mode of vibration (related to the frequency), while $c$ is the rate of exponential decay. For any value of $\lambda$, there is a range of possible $d$ values, which can be captured in a statistical model.

### 4.1 Feature Extraction from Audio

Our technique uses multiple recorded impact sounds to estimate material parameters. 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. Our method uses recorded impact sounds as input.

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 $d_i$.

To extract these features, we adopt the algorithm proposed by Ren et al. [33]. We only capture the linear behavior needed for damping parameter estimation, but their algorithm also provides a method for capturing the nonlinear residual effects for richer resynthesis.

See supplemental material Section 2 for a brief description and our additional modifications to the algorithm. The most notable modification is that we account for background noise (modeled as additive white Gaussian noise) by estimating the amplitude of the noise floor. The extracted $(\omega_{id}, a_i, d_i)$ features are converted into pairs of $(\lambda_i, c_i)$ values, where $\lambda_i$ is the eigenvalue corresponding to that mode of vibration and $c_i = 2d_i$.

### 4.2 Distributions of Damping Values

With $(\lambda_i, c_i)$ features extracted from multiple input sounds, we now interpret the results. Fig. 3 shows an example of features extracted from impact sounds on a porcelain plate. Note that for any given eigenvalue $\lambda$, there exists a range of extracted damping values. This is especially noticeable where feature points appear as a vertical line, showing that even the same mode of vibration may have a variable rate of decay. These results are inconsistent with the damping models in Equation 1 and Equation 2, which propose a one-to-one mapping between $\lambda$ and $c$. Instead, we propose that there is significant error present in the extracted damping value of each feature, and that error can be modeled with a statistical distribution.

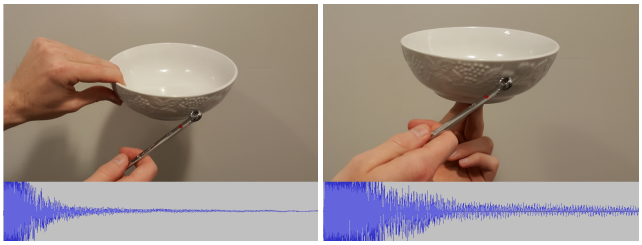

Fig. 4. A porcelain bowl struck in the same location produces different sound when supported with a tight grip (left) or supported by resting on a single point (right). Without accounting for the effect of the support, prior methods would not be able to estimate accurate material parameters from these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

The prior work of Ren et al. [33] estimates damping parameters using a least-squares metric to compare spectrograms. Statistically, a least-squares fit of a damping model is equivalent to assuming there is normally-distributed error around the model. We will refer to least-squares fitting of damping models as LSQ. However, we have found experimentally that least-squares fits tend to overestimate the material damping parameters, and resynthesized sounds all sound heavily overdamped.

Another notable property of Fig. 3 is the clear line of points forming a lower bound to the data (with a few outliers). We have found experimentally that a damping model fit to this lower bound curve results in resynthsized sounds much closer to the input sounds. If the damping model should fit the lower bound, then all error is positive and can only increase the extracted damping values. Statistically, this indicates a one-sided error distribution; e.g. half normal, exponential, or chi-square distribution. Computationally, a lower bound Rayleigh damping model can also be found as a line along the lower convex hull of the points. We refer to lower-bound fitting of damping models as LB.

However, as can be seen in Fig. 3, outliers often appear below the clean LB curve, and for other objects such a clean curve does not appear in the first place. A strict LB fit will be highly sensitive to outliers, as it must assume all error is positive. It is difficult to detect and remove outliers in extracted feature datasets. To solve this problem, we examine the physical sources of error in extracted damping values and construct an appropriate statistical distribution modeling that error. Ideally, this should produce a more robust lower bound fit which handles outliers based on their statistical probability of occurrence.

## 4.3 External Damping Factors

To accurately model error in damping values, we consider a number of physical phenomena that may affect estimates of the material damping values. These *external* damping factors are distinct from the material damping, which occurs due to the internal structure of a material.

### 4.3.1 Support Damping

An object's method of support can be varied; 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 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. A tight grip on the bowl's rim produces a more damped sound compared to gentle balancing on fingertips.

### 4.3.2 Complex Modes

Complex modes of vibration are slight deviations from normal mode behavior. Unlike normal modes, complex modes are not linearly separable: energy may be transferred between modes while vibrating. A mode that *loses* energy to others will produce higher damping values, while a mode that *gains* energy from others will produce lower damping values. Most systems only have slightly complex modes (i.e. there is little energy transfer), so normal modes are a close approximation [17], but not an exact one. Since we make the assumption of normal modes, the slight transfers of energy are a source of error in damping value estimates.

### 4.3.3 Background Noise

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" which do not match the modal exponential-decay model. We modify the feature extraction method to account for additive white Gaussian noise (Sect. 4.1). This further removes the influence of persistent background noise, though there may still be some remaining Gaussian (normal) error in the spectrograms and their resulting extracted damping values.

Acoustic reflections and reverberations from room acoustics are confounding factors. 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, but some small sources of transient noise or early reflections may be appropriately modeled by normally distributed error.

### 4.3.4 Feature Extraction Error

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 discretization of the spectrogram will produce small amounts of error. Sidelobes resulting from Fourier transforms may appear as separate peaks or affect the estimated damping rate of nearby modes.

### 4.3.5 Acoustic Radiation

Uneven acoustic radiation from the object may mean that different microphone placements will result in different initial mode amplitudes. This can be accounted for by keeping the object and 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 not change the frequencies or exponential rates of decay, and thus does not need to be accounted for in our model.

### 4.3.6 External Factor Summary

Current damping parameter estimation techniques do not explicitly consider these factors, instead attributing all damping to the material. [37]. The resulting damping parameters therefore model the combined effect of the material *and the recording environment*. These parameters may not properly transfer to an object of the same material in a *different environment*. This limits the sounds which can be used for accurate damping parameter estimation: the sounds must be recorded in a carefully controlled setting. With a thoroughly robust technique that can separately model environmental factors, we can reduce the factors' impact on the estimated parameters. The external factors cannot be fully removed, but reducing their impact may result in more physically-accurate material parameters.

## 4.4 Generative Model for Combined Damping

We now introduce a generative model for sampling damping values. The model defines the probability distribution for an extracted damping value $c_i$, given the eigenvalue $\lambda_i$ and a set of parameters $\theta$. $\theta$ contains parameters representing both the *material* and the *environment*. The material damping parameters, such as $\alpha_1$ and $\alpha_2$, are referred to as $\theta_d$ for generality. The model can be written as $p(c_i|\lambda_i, \theta)$, and asks, "given a known material and environment, what is the probability of measuring any particular damping value?"

The value $c_i$ is a damping value obtained from the feature extraction step. In the *absence* of any external factors, $c_i$ would only consist of material damping. To account for the external factors, we model $c_i$

as a random variable based on the sum of normally and exponentially distributed random variables.

#### 4.4.1 Normal Distribution

A normal distribution models the effect of some external factors. The normal distribution accounts for (1) energy transfer due to complex modes, (2) small sources of background noise, and (3) error in feature extraction due to spectrogram discretization. We assume that each of these factors are an additive, normally distributed random variable. The sum of these normally distributed factors ($c_i^n$) is also normally distributed:

$$p(c_i^n|\lambda_i,\theta_d,\sigma) = \mathcal{N}\left(c(\lambda_i,\theta_d),\sigma^2\right) \tag{3}$$

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.

#### 4.4.2 Exponential Distribution

An exponential distribution models the effect of the object's support.

$$p(c_i^e|\eta) = \mathrm{Exp}(\eta) = \eta e^{-\eta c_i^e}. \tag{4}$$

$c_i^e$ is the resulting exponential damping resulting from the object's support, while $\eta$ is the rate parameter of the exponential distribution. This distribution is an approximation, but in attempting to create a robust lower bound method, it serves the role of a one-sided distribution fitting to the lower bound of damping values.

Zheng and James defined a model to approximate additional per-mode damping based on contacts with other objects [44]. However, a statistical analysis of this model is highly dependent on the distribution of elements of the matrix of eigenvectors $\phi$. We are not aware of any prior work that has attempted to statistically model the distribution of eigenvector matrix $\phi$ elements, and our own analysis using Kolmogorov-Smirnov goodness-of-fit tests found no probable common distributions. In the absence of a more well-defined model and with the main requirement of a one-sided distribution satisfied, the exponential distribution was selected empirically based on extracted feature data.

#### 4.4.3 Exponentially Modified Gaussian

The combined damping value $c_i$ can then be modeled as the combination of (1) the normally-distributed factors $c_i^n$ due to complex modes, noises, and other sources of errors, and (2) the exponentially-distributed factor $c_i^e$ due to the support damping. Assuming that the factors are independent (for mathematical feasibility), they can be formulated as two separate sources of exponential decay of the mode amplitude $z_i$:

$$z_i(t) = a_i e^{-c_i^n t} e^{-c_i^e t}\cos(\omega_{id}t) \tag{5}$$

$$= a_i e^{-(c_i^n + c_i^e)t}\cos(\omega_{id}t). \tag{6}$$

The probability density function of the sum of the normal and exponential distributions ($c_i^n + c_i^e$) 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 [16], but have also found uses in other domains. The probability density function for the EMG is:

$$p(c_i|\lambda_i,\theta_d,\sigma,\eta) = \frac{\eta}{2} e^{\frac{\eta}{2}(2c(\lambda_i,\theta_d)+\eta\sigma^2-2c_i)}\,\mathrm{erfc}(s_i) \tag{7}$$

$$s_i = \frac{c(\lambda_i,\theta_d)+\eta\sigma^2 - c_i}{\sqrt{2}\sigma}, \tag{8}$$

where erfc is the complementary error function, defined as:

$$\mathrm{erfc}(x) = 1 - \mathrm{erf}(x) = \frac{2}{\sqrt{\pi}}\int_x^\infty e^{-y^2}dy. \tag{9}$$

This defines the probability of observing an extracted damping value, given the material damping and environmental damping parameters.

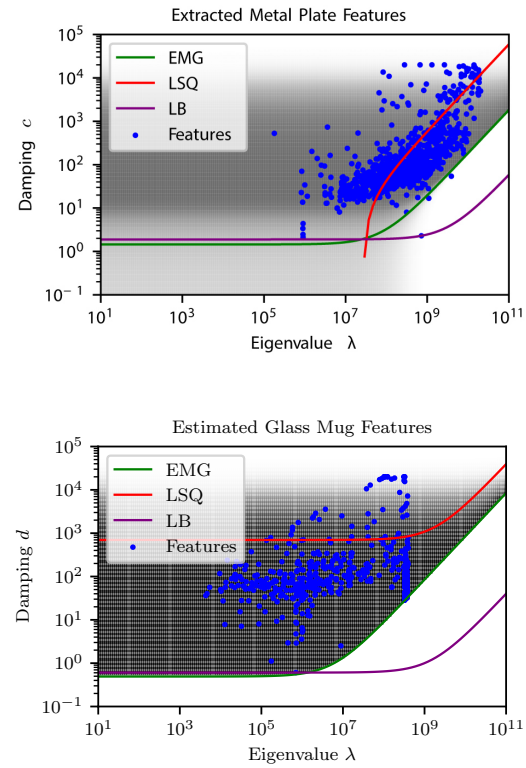


Fig. 5. Parameter estimation on sound features. Each feature consists of an eigenvalue $\lambda_i$ and its corresponding damping coefficient $c_i$. Estimated Rayleigh damping curves are plotted, with the variation from the curve caused by external factors. Our method, using the EMG distribution, provides the closest fit to the lower bound of the data while being relatively unaffected by outliers.

This is the complete generative model for damping values, encompassing multiple sources of damping and errors. 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.

### 4.5 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 data from a distribution. MLE is an optimization method that reverses the process: use known data from a distribution to produce best-fitting parameters. Given a set of extracted $(\lambda_i, c_i)$ 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:

$$\log p(\mathrm{d}|\lambda,\theta_d,\sigma,\eta) = \sum_i \log\left(\frac{\eta}{2}\right) + \eta c(\lambda_i,\theta_d) + \frac{\eta^2\sigma^2}{2} - \eta c_i + \log\left(\mathrm{erfc}(s_i)\right). \tag{10}$$

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.

We compute the full average derivative for the $n$ $(\lambda_i, c_i)$ samples. We define a term $t_i$ and use the scaled complementary error function

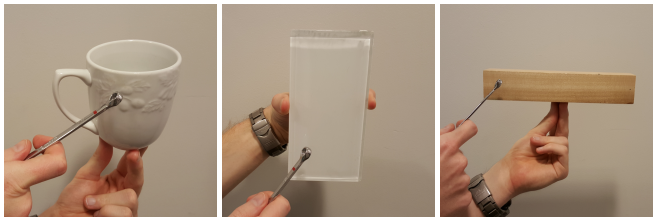

Fig. 6. Three objects from our impact sound dataset: a porcelain cup (left), a small glass tile (center), and a wood block (right). Note the ways that each object is supported. These supports interfere with damping parameter estimation.

$\mathrm{erfcx}(s_i) = \exp(s_i^2)\,\mathrm{erfc}(s_i)$ to simplify notation:

$$t_i = \frac{-2}{\mathrm{erfc}(s_i)\sqrt{\pi}} e^{-s_i^2} = \frac{-2}{\mathrm{erfcx}(s_i)\sqrt{\pi}}. \tag{11}$$

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

$$\frac{\partial \log p}{\partial \eta} = \frac{n}{\eta} + n\eta\sigma^2 + \sum_i c(\lambda_i, \theta_d) - c_i + t_i \frac{\sigma}{\sqrt{2}}, \tag{12}$$

$$\frac{\partial \log p}{\partial \sigma} = n\lambda^2\sigma + \sum_i t_i \left( \frac{\eta\sigma^2 + c_i - c(\lambda_i, \theta_d)}{\sqrt{2}\sigma^2} \right). \tag{13}$$

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:

$$\frac{\partial \log p}{\partial \alpha_1} = \eta n + \sum_i \frac{t_i}{\sqrt{2}\sigma}, \tag{14}$$

$$\frac{\partial \log p}{\partial \alpha_2} = \sum_i \eta\lambda_i + \frac{t_i\lambda_i}{\sqrt{2}\sigma}. \tag{15}$$

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. These damping parameters can be used to represent the recorded material for modal sound synthesis, with other effects (e.g. room acoustics, supports) modeled separately [44].

Figure 5(a) shows features extracted from 19 impact sounds on a metal plate, while Figure 5(b) shows features extracted from 40 impact sounds on a glass mug. The figure compares our EMG fit with MLE optimization against the baseline LSQ and LB methods (see Sect. 4.2). In each case, LSQ overfits the data, while LB is strongly affected by low outliers and underfits the data.

Overall, The effect of external damping factors cannot be entirely removed, and in real-world situations the extracted damping values may all be much higher than the material damping function alone. This positively biases the estimator: the estimated parameters will often be larger than the ground truth. By accounting for external factors, this estimator has less bias than other methods, and is therefore more accurate.

## 5 Results

We have implemented the damping parameter estimation method described in this paper and tested its effectiveness through both numerical analysis and perceptual validation. With this method, the process for material damping parameter estimation involves striking an object repeatedly, ideally with varying hit locations and support methods. This approach has less strict requirements about the recording environment than previous work; sounds can be recorded in a quiet room, as long as there are few transient sounds and the room is not heavily reverberative. Full recording and implementation details can be found in supplementary material Section 5.1.

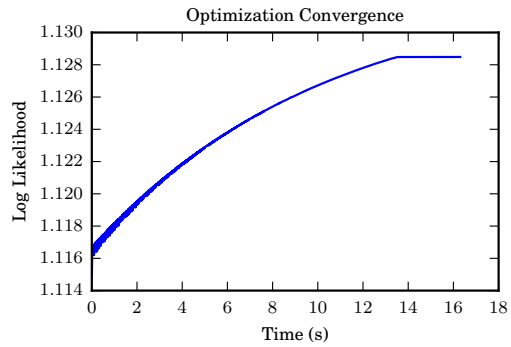


Fig. 7. Plot of log-likelihood maximization converging over the course of parameter estimation. Optimization was performed on 752 frequency-damping points extracted from porcelain plate impact sounds, and converged after 39,009 iterations for a total of 16.3s in an one-time preprocessing.

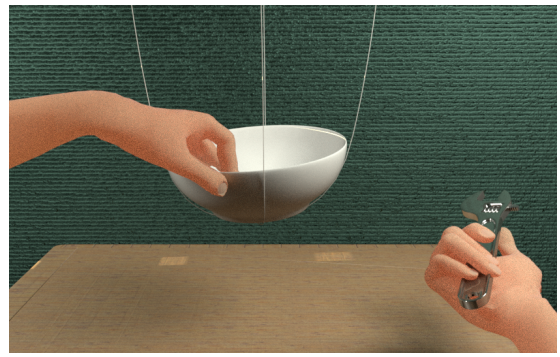

Fig. 8. A simulated porcelain bowl is struck in multiple locations, with and without a supporting grip.

We have recorded numerous impact sounds on a set of fifteen rigid objects, where the hit points and the method of support are documented for each impact. Figure 6 shows a sample of these objects, with various hit locations and methods of support. There are an average of nearly 50 impact sounds sampled per object. All objects were supported by hand, often either with an edge being pinched between two fingers or the center resting on a few fingertips.

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. See Figure 7 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. Upon convergence, the optimized $\theta_d$ parameters model the damping properties of the recorded material.

Table 1 contains results from extraction on some of the objects. 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.

### 5.1 Real-time Synthesis and Rendering

Each sounding object is preprocessed using estimated material parameters, then synthesis of sounds can be performed in real time. During synthesis, support damping is simulated when appropriate; refer to sup-

| | | Porcelain Plate | Travertine Tile | Wood Block | Steel Wrench | Plastic Bowl | Glass Tile |
|---|---|---|---|---|---|---|---|
| **Rayleigh** | $\alpha_1$ | 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 |
| **Hybrid** | $\gamma_1$ | 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 3.2. 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.

plemental material Section 4. Figure 1 shows multiple scenes from our real-time demo, with multiple objects of various shapes and materials. Figure 8 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 Human Hand-Tuning Evaluation

In the absence of an automated method for damping parameter estimation, parameters have traditionally been estimated by hand. We present a study evaluating the effectiveness of human damping parameter estimation, using human subjects to hand-tune material parameters for multiple objects. Specifically, we are interested in the tuning of the damping parameters and the specific modulus, defined as the ratio of Young's modulus to density. We seek to evaluate the distributions of subjects' selected material parameters. For example, are subjects able to agree on a single unique set of material parameters, and if so, to what degree of precision? We also seek to determine the time and sound samples needed for subjects to reach their conclusions.

#### 5.2.1 Experimental Setup

We constructed an easy-to-use GUI enabling interactive adjustment of material parameters for sounds produced through modal sound synthesis. For each object in the study, we created a corresponding 3D model by hand (a laborious process requiring precise measurements) and performed modal analysis on that model once (a few hours of computation time). The damping parameters and specific modulus for an object can be adjusted as a post-processing step, without needing to repeat the lengthy modal analysis step. With these optimizations, resynthesis with modified parameters took less than two seconds, allowing for rapid iteration. In the interface, each parameter was controlled with a slider, with a range of plausible realistic values presented on normalized scales from 0–100.

Subjects were recruited primarily through mailing lists and were not required to have any background in parameter tuning or impact sound analysis. Subjects were compensated financially for their participation. Subjects were given real-world objects, placed on small foam blocks to reduce support damping. For each object, the subjects' task was to tune material parameters such that the synthesized sound produced by the application most closely matched the sound they heard when striking the real object. Subjects were instructed to find the most accurate parameters possible, regardless of the time needed. Subjects first performed this task with a training object, in order to reduce learning effects. The six objects evaluated were all disk-shaped objects of approximately the same radius and thickness. Every subject hand-tuned material parameters for all six objects in a random order.

The study was divided into two sections. The first 20 subjects hand-tuned three parameters for each object: the two Rayleigh damping parameters and the specific modulus. The following 20 subjects hand-tuned two parameters for each object: just the two Rayleigh damping parameters. For the two-parameter section, the specific modulus was set to the mode of the subject-selected specific moduli from the three-parameter section. The three-parameter section models the real-world case where all three parameters must be picked in order to virtualize an object.

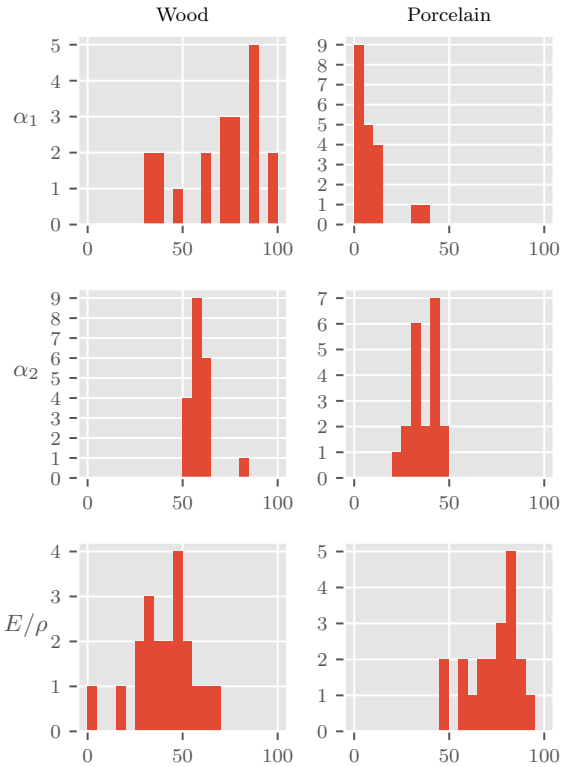


Fig. 9. Distributions of human-tuned material parameters for wood and porcelain discs. $\alpha_1$ and $\alpha_2$ are Rayleigh damping parameters, $E/\rho$ is the specific modulus, and all parameters were tuned on normalized scales from 0–100. The observed distributions indicate that subjects had difficulty finding a unique optimal solution for the specific modulus, and for $\alpha_1$ in the case of highly-damped objects.

#### 5.2.2 Results and Analysis

We first consider the distributions of subjects' selected material parameters. Fig. 9 shows results from the three-parameter section of the study for a few selected objects: a wood disc and a porcelain disc. For highly reverberant objects such as the porcelain disk, subjects could generally agree on Rayleigh damping's $\alpha_1$ parameter for each object. However, for highly damped objects such as the wood disk, $\alpha_1$ responses were less consistent. The distributions of $\alpha_2$ parameters for each object show agreement between subjects, indicated by the relatively low standard deviations and frequently unimodal distributions. The specific modulus, which modifies the pitch of the synthesized sound, often resulted in multimodal distributions.

We also consider the time and number of sounds needed for subjects to reach their conclusions. Fig. 10 contains histograms for the amount of time and number of synthesized sounds needed for subjects to finalize their selections. The median time needed was 165 s to tune three parameters, and 145 s to tune two parameters. The median number of synthesized sounds needed was 35 to tune three parameters, and 28 to tune two parameters. The range of times (33–614 s) and sounds (5–182) was highly variable, and the effect of parameter count on times

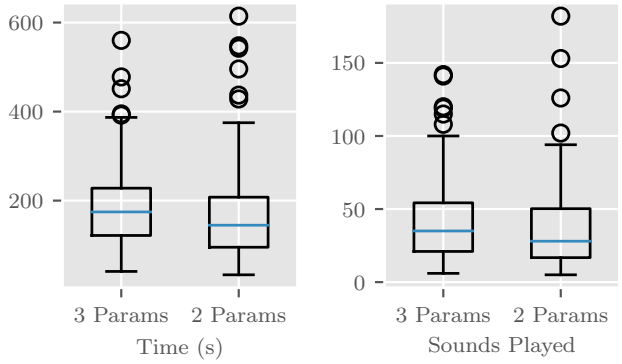


Fig. 10. Box-and-whisker plots of the time and number of synthesized sounds needed for subjects to reach their final hand-tuned material parameters, with deviant observations plotted as outliers. Our method is an automated version of the 2-parameter study, and significantly reduces the human labor needed.

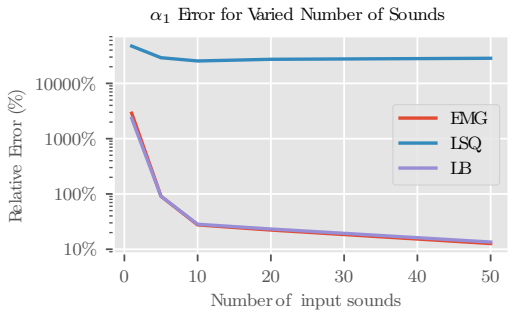


Fig. 11. As the number of sounds used for parameter estimation increase, error in Rayleigh damping parameter $\alpha_1$ decreases. Overall, our method outperforms the other two alternatives.

and sounds did not reach statistical significance by t-test.

Human hand-tuning requires around 145 s and 28 sounds per object, requiring dedicated human attention for the entire duration. Hand-tuning also requires creating an accurate 3D model of the object and performing modal analysis, possibly adding hours of extra human effort. In contrast, our automated method operates effectively with 10–20 sounds and does not require a 3D model of the object. Parameter estimation then takes a few minutes, during which no human attention is required. Overall, our method significantly reduces the amount of human labor needed to create virtualized objects. Compared to prior work, our method reduces human labor by not requiring carefully controlled recording environments, creation of a 3D model, and knowledge of object geometry and hit points.

### 5.3 Synthetic Validation

Synthetic validation provides a numerical comparison against ground-truth damping parameters. We synthesized a variety of sounds with 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 geometric models. 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. 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: EMG (our method, see Sect. 4.4) and the two baselines LSQ and LB (see Sect. 4.2). Direct comparison against the algorithm of Ren et al. [33] is infeasible due to the significant differences in inputs and outputs. However, their method will produce results most similar to the least squares (LSQ) estimator. 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

Fig. 11 shows the relative error for each parameter and each estimator. For all materials in this synthetic data, both the EMG and LB estimators significantly outperformed the LSQ estimator ($p < .05$). With real-world data, the EMG and LB estimators more frequently decouple, as the EMG estimator's statistical model better adapts to noise and other artifacts of recording. These synthetic sounds, without noise or other effects, are the ideal situation for the LB estimator, and do not leverage the full capabilities of the EMG estimator.

$\alpha_1$ estimates have minimal error, especially with larger amounts of data. For discussion of $\alpha_2$ estimates, please refer to supplementary material Section 5.3. However it is difficult to determine from strictly quantitative results what effect any error in estimation will have on users in a virtual environment. The actual effect of the error can be evaluated through perceptual evaluation.

### 5.4 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. In this study, we considered recorded real-world sounds, and sounds synthesized using three sets of damping parameters: parameters from Ren et al. [33], parameters from the human hand-tuning study (Sect. 5.2), and parameters estimated using our method to create 4 datasets. We sought to evaluate how well the synthesized sounds recreate the real-world sounds. Subjects evaluated sounds individually, answering questions about qualities of the sounds and estimating properties of the object or impact. Synthesized sounds that more accurately recreate properties of the real-world sounds should produces similar patterns of answers to questions.

#### 5.4.1 Experimental Setup

The study was conducted in an online web questionnaire, and subjects were recruited through mailing lists and online posts, but no financial compensation was offered. No prior experience in auditory perception was expected. Subjects were asked to wear headphones or earbuds to ensure a consistent auditory environment. All sounds were scaled to the same volume, though difference in sound playback devices may have affected perception. Subjects listened to a series of impact sounds, answering questions about each. Variables involved are sound datasets (4: as listed above), object shape (2: disc or rod), and material class (5: wood, metal, plastic, glass, porcelain). All together, this produces a total of 40 sounds to evaluate.

24 subjects participated in the study, but more specific demographic information was not collected. Each subject listened to all 40 impact sounds in randomized order. For each sound, subjects were asked which object shape and material class they suspected created the sound. Subjects also were asked to rate descriptive qualities of the sound—the duration, ringiness, tonality, and pitch—on 7-point ordinal scales. The extreme ends of the scales were descriptively labeled, e.g. tonality ranged from "mixed tones" to "pure tone". Subjects could listen to each sound multiple times as needed. A brief training section provided example sounds and definitions for the descriptive qualities.

#### 5.4.2 Results: Confusion Matrices

Even with recorded sounds, user identification of material is not always accurate. In evaluation of synthetic datasets, we compare the pattern of errors to those of the recorded sounds, with a closer match suggesting

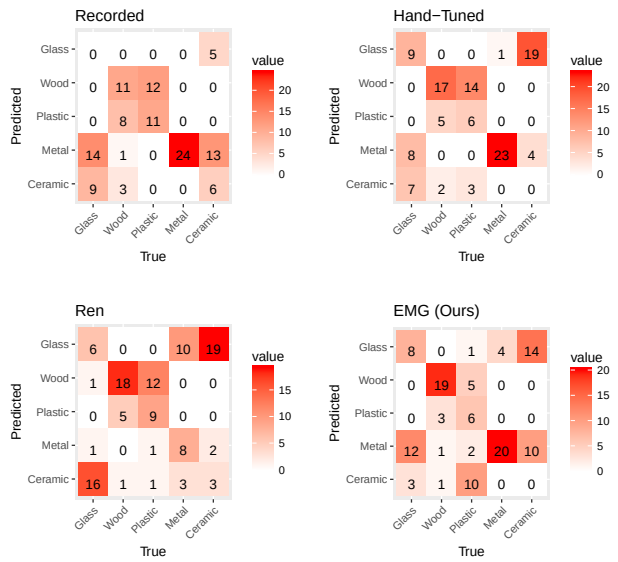


Fig. 12. Material confusion matrices for the disc-shaped objects in our study. All tested datasets (including Recorded) show significant labeling errors. However, our method (EMG) replicates the pattern of errors seen in the recorded and hand-tuned datasets more closely than the Ren dataset, suggesting more accurate recreations of the real-world sounds.

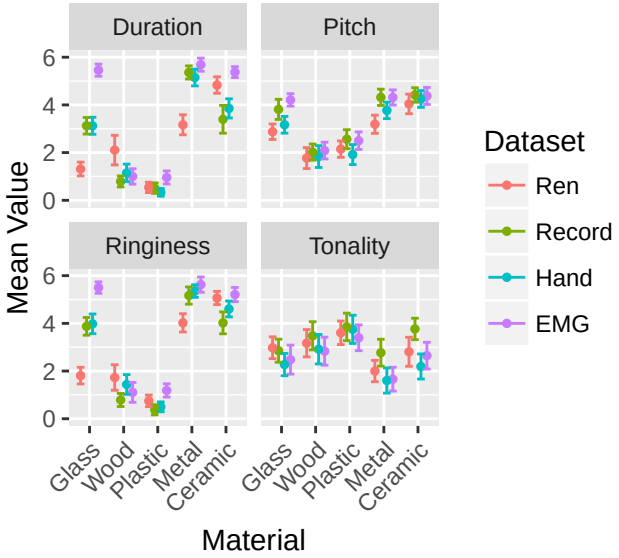


Fig. 13. The mean selected value for each descriptive quality, material, and dataset. Most datasets are tightly clustered for pitch and tonality, with more differences in duration and ringiness. While our EMG dataset contains some differences in duration and ringiness, it is closer to the recorded means than the Ren dataset.

more realistic synthesized sounds. Fig. 12 shows confusion matrices for material class identification for the disc-shaped objects.

The recorded dataset demonstrates mis-labelings such as heavy confusion between wood and plastic, perception of the glass and ceramic discs as metal, and high accuracy on the metal disc. The hand-tuned dataset differs primarily in perception of its glass and ceramic objects; these differences could be due to human error while hand-tuning or due to inherent assumptions in the underlying modal synthesis model. The Ren dataset largely reproduces the matrix from their original paper [33], although it does not recreate the error patterns (particularly metal) seen in our recorded or hand-tuned datasets. Our dataset (EMG) most closely resembles the hand-tuned results, with the exception of plastic being identified as ceramic by some subjects.

We evaluate the pairwise similarity between these matrices by computing the Frobenius norm of the element-wise difference of the two. The two most similar matrices are the hand-tuned and EMG dataset results, with a difference norm of 16.03. In comparison, between Ren and the hand-tuned data, the norm is 22.09. Against recorded sounds, EMG's norm was 23.11, while Ren's norm was 31.85 and hand-tune's norm is 22.83. The high similarity (low difference norm) between our EMG results and the hand-tuned results suggests that our method automatically produces sounds perceptually similar to what would be selected by human hand-tuning.

#### 5.4.3 Results: Descriptive Qualities

We evaluate the descriptive quality ratings by performing a multi-factorial repeated measures ANOVA. Each of the variables (sound dataset, object shape, and material) is considered an ANOVA factor, each a repeated measure across subjects. The main effects of dataset, shape, and material are all significant for all four descriptive qualities. For example, for perception of pitch, the effects of material ($F_{4,92} = 95.61, p < .05$), shape ($F_{1,23} = 30.35, p < .05$), and dataset ($F_{3,69} = 22.79, p < .05$) are all significant. This is not surprising, as each of these effects alone can dramatically change the sound. Almost all interaction effects are significant, with the exceptions of material*dataset on tonality ($F_{12,276} = 2.814, p = .107$). A full table of effects can be found in supplementary material Table 1.

Fig. 13 contains the mean values for each descriptive quality, material, and dataset (using combined results for shapes). For pitch and tonality, most dataset means are closely-clustered. Duration and ringiness show more difference: while hand-tuned and recorded are closely aligned, Ren and EMG occasionally display more variance.

Synthetic datasets that produce more realistic sounds should have descriptive qualities similar to the recorded dataset. To evaluate this, we look at the absolute error between each subject's recorded and synthetic ratings for each sound. Across all materials and shapes, hand-tuned sounds were closest to the recorded sounds for duration and ringiness. For pitch and tonality, all datasets produced more similar results. Comparing these with a one-way ANOVA produces no significant effect of dataset on average error ($F_{2,117} = 2.81, p = .064$). A full analysis can be found in supplementary material Section 5.4.

Hand-tuned parameters produce the closest ratings to the recorded dataset on these perceptual scales. The Ren dataset contains many discrepancies from the real dataset, shown in the duration, pitch, and ringiness of the more reverberative materials. Our EMG dataset properly reproduces the perceptions of the pitch and tonality of the recorded objects, but in some cases produces higher duration and ringiness. Our EMG method demonstrates an improvement over the Ren dataset in the fit to recorded sounds.

## 6 Conclusion

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

### 6.1 Limitations and Future Work

Our method removes a number of assumptions used by prior damping parameter estimation techniques [33]. For example, our method does not require knowledge of the object's geometry, and it reduces the strict assumptions on the object's support and the presence of background noise. However, some common assumptions of prior works remain: (1) application to rigid objects and their vibrations can be accurately modeled by linear analysis. (2) difficulty to fully remove all external damping factors—the presence of loud transient noises, a tightly-coupled support, or a highly reverberative room may still impose residual effects.

We do not assume prior knowledge on the object geometry, size, material parameters, or the impact location. However, this technique currently does not estimate properties of the object other than damping. Generalization of this work or use of learning algorithms can potentially estimate these parameters automatically. A single sound is not enough to estimate parameter $\alpha_1$ with sufficient accuracy; upwards of 10–20 sounds may be needed. In our human parameter tuning study, subjects were initially untrained; experts may produce slightly different parameter distributions.

## References


[1] Numerical modeling of collisions in musical instruments. *Acta Acustica united with Acustica*, 101(1), 2015.

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

[3] S. Adhikari. Damping modelling using generalized proportional damping. *Journal of Sound and Vibration*, 293(1–2):156 – 170, 2006. doi: 10.1016/j.jsv.2005.09.034

[4] F. Avanzini and D. Rocchesso. Modeling collision sounds: Non-linear contact force. In *In Proc. COST-G6 Conf. Digital Audio Effects (DAFx-01*, pp. 61–66, 2001.

[5] 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

[6] S. Bilbao, B. Hamilton, A. Torin, C. Webb, P. Graham, A. Gray, K. Kavoussanakis, and J. Perry. Large scale physical modeling sound synthesis. In *Proceedings of the Stockholm Musical Acoustics Conference/Sound and Music Computing Conference*, 8 2013.

[7] T. Caughey. Classical normal modes in damped linear dynamic systems. *Journal of Applied Mechanics*, 27(2):269–271, 1960.

[8] T. Caughey and M. O'kelly. Classical normal modes in damped linear dynamic systems. *Journal of Applied Mechanics*, 32(3):583–588, 1965.

[9] J. N. Chadwick, S. S. An, and D. L. James. Harmonic shells: A practical nonlinear sound model for near-rigid thin shells. In *ACM SIGGRAPH Asia 2009 Papers*, SIGGRAPH Asia '09, pp. 119:1–119:10. ACM, New York, NY, USA, 2009. doi: 10.1145/1661412.1618465

[10] J. Derek Morrison and J. M Adrien. Mosaic. a framework for modal synthesis. 17, 09 1994.

[11] A. E756. Standard test method for measuring vibration-damping properties of materials. 2017.

[12] W. Fujisaki, N. Goda, I. Motoyoshi, H. Komatsu, and S. Nishida. Audio-visual integration in the human perception of materials. *Journal of Vision*, 14(4):12, 2014. doi: 10.1167/14.4.12

[13] B. L. Giordano and S. McAdams. Material identification of real impact sounds: Effects of size variation in steel, glass, wood, and plexiglass plates. *The Journal of the Acoustical Society of America*, 119(2):1171–1181, 2006. doi: 10.1121/1.2149839

[14] B. L. Giordano, D. Rocchesso, and S. McAdams. Integration of acoustical information in the perception of impacted sound sources: The role of information accuracy and exploitability. *Journal of Experimental Psychology: Human Perception and Performance*, 36(2):462–479, 2010.

[15] M. Grassi. Do we hear size or sound? balls dropped on plates. *Perception & Psychophysics*, 67(2):274–284, Feb 2005. doi: 10.3758/BF03206491

[16] E. Grushka. Characterization of exponentially modified gaussian peaks in chromatography. *Analytical Chemistry*, 44(11):1733–1738, 1972. PMID: 22324584. doi: 10.1021/ac60319a011

[17] 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.

[18] 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.

[19] M. Kac. Can one hear the shape of a drum? *The American Mathematical Monthly*, 73(4):1–23, 1966.

[20] K. Karplus and A. Strong. Digital synthesis of plucked-string and drum timbres. *Computer Music Journal*, 7(2):43–55, 1983.

[21] R. L. Klatzky, D. K. Pai, and E. P. Krotkov. Perception of material from contact sounds. *Presence: Teleoperators and Virtual Environments*, 9(4):399–410, 2000. doi: 10.1162/105474600566907

[22] 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

[23] D. Li, Y. Fei, and C. Zheng. Interactive acoustic transfer approximation for modal sound. *ACM Trans. Graph.*, 35(1), 2015. doi: 10.1145/2820612

[24] 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.

[25] S. McAdams, V. Roussarie, A. Chaigne, and B. L. Giordano. The psychomechanics of simulated sound sources: Material properties of impacted thin plates. *The Journal of the Acoustical Society of America*, 128(3):1401–1413, 2010. doi: 10.1121/1.3466867

[26] J. H. McDermott, M. Schemitsch, and E. P. Simoncelli. Summary statistics in auditory perception. *Nature neuroscience*, 16(4):493, 2013.

[27] A. D. Nashif, D. I. Jones, and J. P. Henderson. *Vibration damping*. John Wiley & Sons, 1985.

[28] 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

[29] 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.

[30] 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

[31] J. W. S. B. Rayleigh. *The theory of sound*, vol. 2. Macmillan, 1896.

[32] 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.

[33] 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

[34] 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.

[35] 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.

[36] J. O. Smith. Physical modeling using digital waveguides. *Computer Music Journal*, 16(4):74–91, 1992.

[37] 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

[38] J. Traer and J. H. McDermott. Statistics of natural reverberation enable perceptual separation of sound and space. *Proceedings of the National Academy of Sciences*, 113(48):E7856–E7865, 2016. doi: 10.1073/pnas.1612524113

[39] K. van den Doel and D. K. Pai. The sounds of physical shapes. *Presence*, 7:382–395, 1996.

[40] M. van Walstijn and S. Mehes. An explorative string-bridge-plate model with tunable parameters. In *Proceedings of the 20th International Conference on Digital Audio Effects (DAFx-17)*. Edinburgh, UK, Sept. 2017.

[41] J.-H. Wang, A. Qu, T. R. Langlois, and D. L. James. Toward wave-based sound synthesis for computer animation. *ACM Trans. Graph.*, 37(4), 2018.

[42] J. Woodhouse. Linear damping models for structural vibration. *Journal of Sound and Vibration*, 215(3):547–569, 1998.

[43] 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.

[44] C. Zheng and D. L. James. Toward high-quality modal contact sound. *ACM Transactions on Graphics (Proceedings of SIGGRAPH 2011)*, 30(4), Aug. 2011.