Physical Audio Modeling of Passive Rigid Bodies

Real-time Interactive Audio Generation Using Linear Modal Analysis

Karl Hiner, Georgia Tech

Introduction

Goal: Estimate the time-varying Sound Pressure Level (SPL) at a "listener" point in a 3D environment with *passive rigid bodies* vibrating (and radiating acoustic waves) in response to impact forces on their surface.

Context: Interactive applications on consumer hardware (e.g. video games)

Assumptions:
- Homogeneous object material and medium of acoustic propagation.
- *Small-signal acoustics:* Amplitudes of the pressure waves are small enough to not affect the properties of the propagating medium.
- *Linear modal analysis (LMA):* Surface vibrations are a linear combination of independent oscillations at different frequencies and amplitudes. *(Can't model highly nonlinear phenomena, e.g. sharp transients following impact.)*

Linear Modal Analysis (LMA)

Linear elastodynamic equation: $\mathbf{M\ddot{u}} + \mathbf{C\dot{u}} + \mathbf{Ku} = \mathbf{f}$
- $\mathbf{u}(t)$ **: (Target quantity)** Displacements of N nodes within a volume
- $\mathbf{f}(t)$ **:** Force vector
- $\mathbf{M}, \mathbf{C}, \mathbf{K}$: Finite element mass/damping/stiffness matrices

Rayleigh damping: Assume $\mathbf{C} = \alpha\mathbf{M} + \beta\mathbf{K}$, where α and β are scalars. (Assume damping is a linear combination of inertial and elastic effects.)
Linear modal analysis: Assume solutions have the form $\mathbf{u}(t) = \mathbf{\Phi q}(t)$, where $\mathbf{\Phi}_i \in \mathbb{R}^{3N}$ are the "mode shapes" and $\mathbf{q}(t)$ are the corresponding modal amplitudes, assumed to be exponentially-decaying sinusoids.
Substituting yields the *eigenvalue problem* $\mathbf{K\Phi} = \mathbf{\Lambda M \Phi}$, where $\mathbf{\Lambda}$ is the diagonal matrix of eigenvalues and the eigenvectors $\mathbf{\Phi}$ is the modal matrix.

This allows the motions due to individual modes to be computed independently and combined by linear superposition. The system of decoupled ordinary differential equations can be written as

$$\ddot{q}_i + 2\xi_i\omega_i\dot{q}_i + \omega_i^2 q_i = \frac{Q_i}{m_i}, \quad i = 1..N,$$

- $m_i = diag\left(\mathbf{\Phi}^T\mathbf{M\Phi}\right)_i, k_i = diag\left(\mathbf{\Phi}^T\mathbf{K\Phi}\right)_i, Q_i = \left(\mathbf{\Phi}^T\mathbf{f}\right)_i$
- $\omega_i = \sqrt{\mathbf{\Lambda}_i} = \sqrt{k_i/m_i}$: Undamped frequency of vibration (rad)
- $\xi_i = \dfrac{c_i}{2m_i\omega_i} = \dfrac{1}{2}\left(\dfrac{\alpha}{\omega_i} + \beta\omega_i\right)$: Dimensionless modal damping factor

For a system starting from rest at $t = 0$, the solution for the mode i due to forcing $Q(\tau)$ is:

$$q_i(t) = \int_0^t e^{-\xi_i\omega_i(t-\tau)} \sin \omega_{di}(t-\tau)\frac{Q_i(\tau)}{m_i\omega_{di}}\, d\tau,$$

where $\omega_{di} = \omega_i\sqrt{1-\xi_i^2}$ is the (observed) damped frequency of vibration.

Integration Method (Audio DSP)

Each resonating mode is an exponentially decaying sine wave with a frequency, gain, and resonance duration. Rather than implementing them as sine waves with exponential envelopes, using resonant bandpass filters allows any signal to excite the vertices (modes). Mode filters are implemented as a parallel bank of biquad filters (***Fig. 1***), each with transfer function (***Fig. 2***)

$$H(z) = g\frac{1 - z^{-2}}{1 + \alpha_1 z^{-1} + \alpha_2 z^{-2}},\ \alpha_1 = -2\tau\cos\omega,\ \alpha_2 = \tau^2,\ \omega = \frac{2\pi f}{f_s},\ \tau = 0.001^{\frac{1}{t_{60}}}$$

g : Mode gain, f : Mode frequency (Hz), f_s : Sampling rate,
t_{60} : time it takes for the mode to decay by 60 dB from its initial amplitude.

LMA: The mode frequencies correspond to the eigenvalues (and the damping factor), the gains are determined from the eigenvectors and the excitation force, and the t_{60} values are derived from the mode damping factors.

The impact "hammer" is modeled as a short pulse of white noise, followed by a Gaussian envelope, followed by a lowpass filter.

The full audio pipeline is compiled into Faust (Functional Audio Stream) programming language. Show below are parts of the signal-flow block diagram produced by Faust from a generated DSP program:

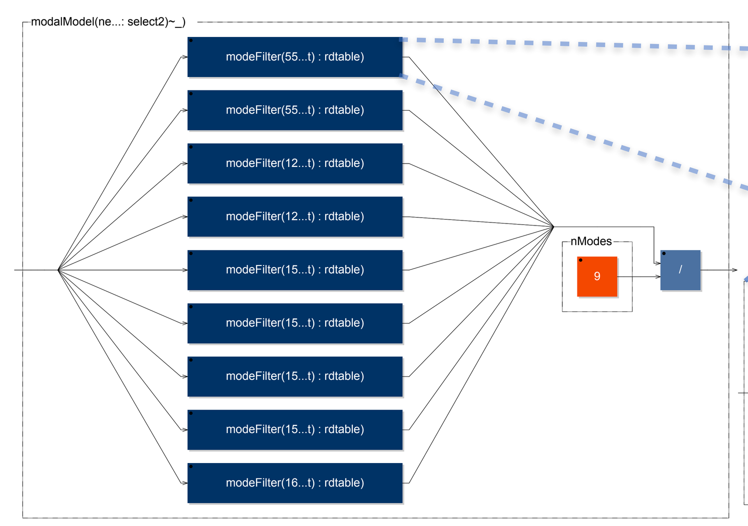

Figure 1. Parallel bank of biquad filters for each mode

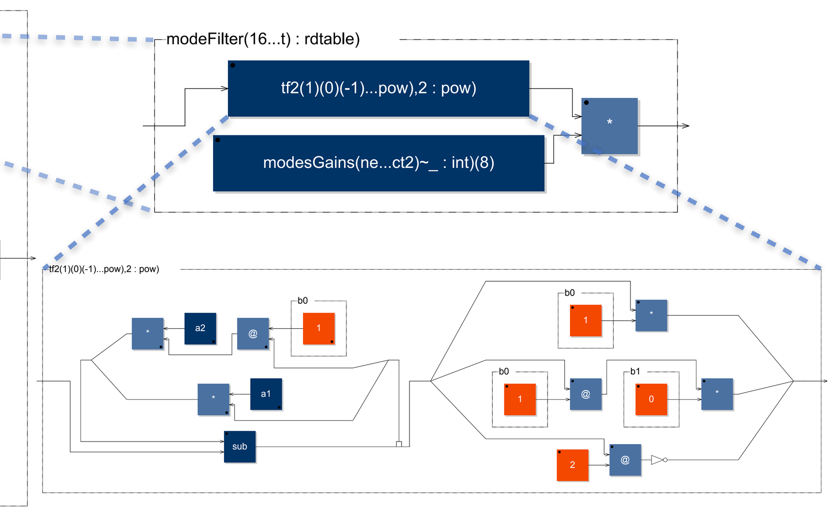

Figure 2. Single mode filter implementing $H(z)$

Mesh Audio Editor Application

Project application supports converting arbitrary surface meshes into modal audio models with interactive vertex excitation.

Also acts as a dataset explorer for the *RealImpact* dataset*:
- 15 microphones x 4 distances x 10 angles = 600 listener positions
- 5 impact points per listener position = 3000 impact samples per object
- 50 objects = 150,000 recordings
- ~4 seconds @ 48 kHz = for each recording

** RealImpact: A Dataset of Impact Sound Fields for Real Objects*
Samuel Clarke et al., 2023

- Generate tetrahedral mesh and modal audio model
- Play real-world impact audio for selected vertex/listener position
- "Strike" modal audio mesh at the same vertices
- Record model audio response waveform and spectrogram, and compare with real-world recording
- View estimated modal parameters
- Set material properties and control audio DSP parameters

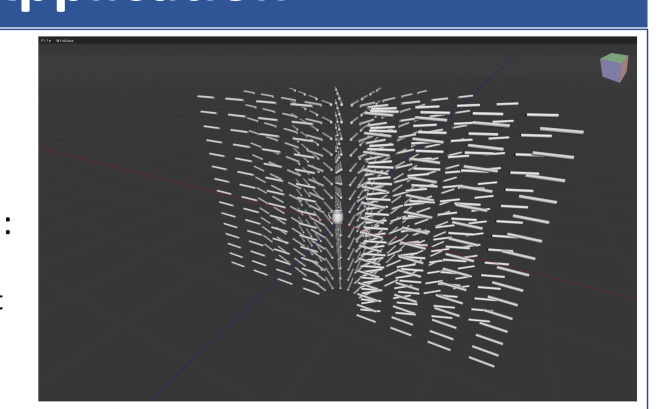

Figure 3. Microphone positions and cup

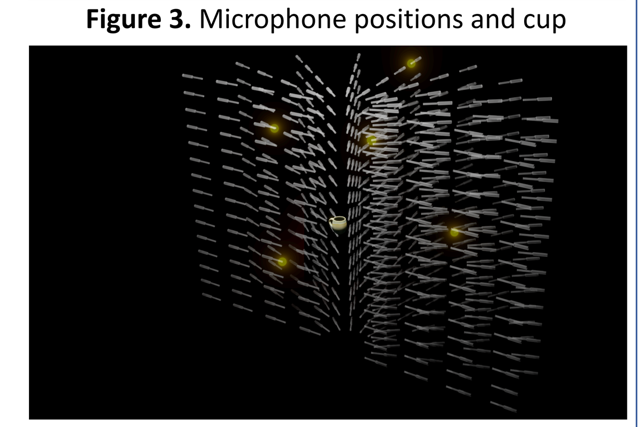

Figure 4. Comparison to *RealImpact* publication

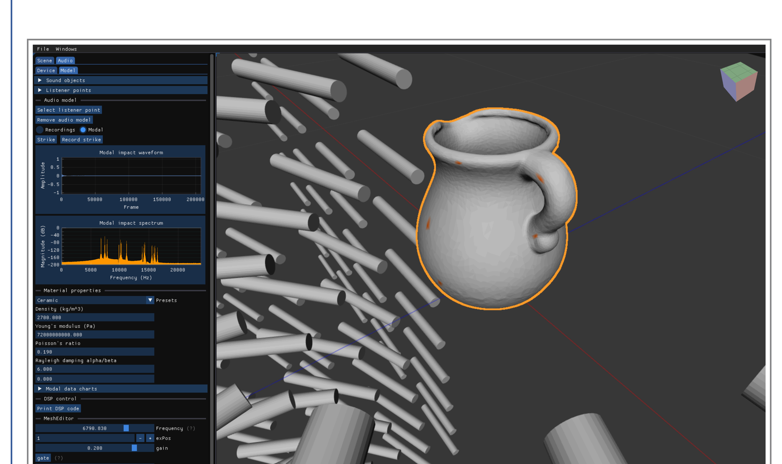

Figure 4. Ceramic pitcher with highlighted impact points

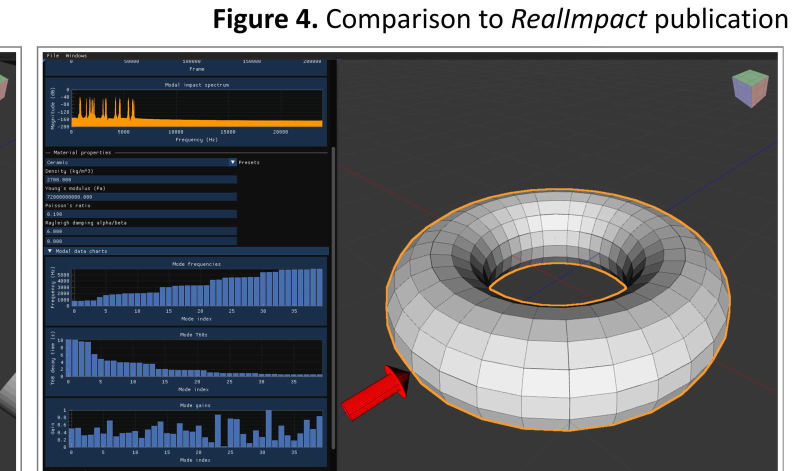

Figure 5. Modal parameters for vertex position on torus

Results

- Highly dependent on geometry/material
- Perceptually decent in many cases
- Fundamental frequency often needs tuning (affected by material properties)
- Generally, not enough damping
- Misses modes often

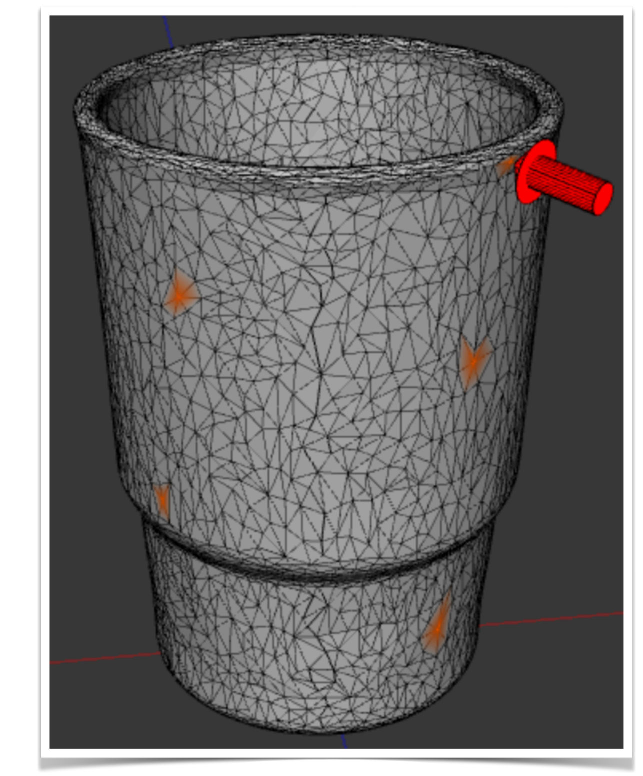

Figure 5. Ceramic cup impacted on lip

Figure 6. Comparison of resonating modes
Top: Real-world impact recording
Bottom: Modal audio model recording
Close Match <=> Not Close

Conclusions and Future Work

- Linear modal analysis is generally effective, physically plausible, and highly efficient.
- Linear assumptions can have perceptible impact on physical accuracy, especially with few estimated modes (e.g. < 50).
- Some materials are less well modeled (e.g. wood, perhaps because it is more deformable)
- Need known baselines for testing implementation.

Future work
- Estimate modal parameters from impact recordings (potential baseline)
- Model acoustic radiation
  - Simplest: Attenuation based on distance to listener, with high-frequency damping due to air absorption (assuming point source with no geometry)
  - Use Boundary Element Method (BEM) to solve the Helmholtz equation to model frequency-dependent directivity patterns and intensity variations in the radiated sound field
    - Efficient since it only requires discretization of the object's surface, not the entire volume of the space
    - Can use LMA to provide vibration characteristics of the object's surface
  - Finite-Difference Time-Domain as source-of-truth for comparison, and to handle nonlinear dynamics and time-varying geometry and boundaries
- Head-related transfer function (HRTF) to efficiently model the effect of human listener geometry on perceived audio for improved realism and 3D localization
- Collisions between objects or environment (e.g. dropping on floor)