

Audio Wave Radiation with 3D Rigid Body Excitations

How well can we do in interactive real-time settings?

Karl Hiner
khiner6@gatech.edu

March 9, 2024

1 Research topic and question

1.1 Research topic and physical scenario description

For my project, I am investigating physical audio modeling of rigid bodies, in the field of dynamics. Specifically, I will explore methods of estimating the time-varying sound pressure level at a point in a 3D environment with passive rigid bodies vibrating (and radiating audio waves) in a homogeneous medium (air) in response to excitations on their surface.

I am particularly interested in performance/accuracy tradeoffs, and aim to maximize (perceived) physical accuracy under the computational constraint of real-time audio generation on consumer hardware.

1.2 Research question, physics overview, and a benchmark computational model

When constrained to consumer-grade compute in a real-time interactive setting, what assumptions, models, and restrictions achieve results similar (in terms of quantitative and/or perceptual accuracy) to highly physically accurate (and computationally expensive) methods?

The Finite-Difference Time-Domain (FDTD) method is detailed, physically accurate, and highly general for modeling acoustic radiation dynamics. In this project, I will use it as a benchmark for comparison with other methods.

FDTD is a staple computational method in electrodynamics, ubiquitously used in fields such as optics to model the spatial evolution of electromagnetic radiation over time. It can accurately model arbitrary (potentially dynamic) boundary conditions, and makes no simplifying assumptions about the physics governing the dynamics other than those arising from its discretization and numerical methods. FDTD uses central differencing to discretize the first-order partial derivatives in

space and time describing the coupling of the electric and magnetic fields. It iteratively approximates solutions to these first-order PDEs by alternately solving for the electric field components, then using these results to solve for the magnetic field components, and so on in a leapfrog fashion.

Acoustic wave propagation can also be described by coupled first-order differential equations, where the temporal derivative of one field is related to the spatial derivative of another field [8]. In acoustics, the analog to the electric field is the *scalar pressure field* $P(\mathbf{x}, t)$, where \mathbf{x} is a 3D position vector, and the analog to the magnetic field is the *vector velocity field* $\mathbf{v}(\mathbf{x}, t)$, representing the velocity of the medium's particles as pressure waves propagate through it. The key material parameters are the speed of sound c_a and the medium's particle density ρ , which affects the medium's acoustic impedance.

The governing equations in three dimensions are expressed as:

$$\frac{\partial P}{\partial t} = -\rho c_a^2 \nabla \cdot \mathbf{v}, \quad (1)$$

$$\frac{\partial \mathbf{v}}{\partial t} = -\frac{1}{\rho} \nabla P. \quad (2)$$

From these equations, the wave equation can be derived (as detailed in *Schneider, 2010* [8]):

$$\nabla^2 P - \frac{1}{c_a^2} \frac{\partial^2 P}{\partial t^2} = 0.$$

A solution particularly relevant for modeling sound wave dynamics is the harmonic plane wave, described by:

$$P(\mathbf{x}, t) = P_0 e^{j(\omega t - \mathbf{k} \cdot \mathbf{x})},$$

where P_0 is the amplitude of the pressure wave, ω is the angular frequency, \mathbf{k} is the wave propagation vector.

While E&M interactions are fundamentally linear, linearity only holds generally in *small-signal acoustics*, when the amplitudes of the pressure waves are small enough that the dynamics do not significantly affect the properties of the propagating medium. Thus, when considering FDTD methods, and in fact *all* methods discussed here and investigated in this project, I make this small-scale assumption.

Since I focus on computational complexity in this project, it is also relevant to observe that acoustics modeling with FDTD is fundamentally simpler than electromagnetics, since E&M involves two vector fields carrying *six* total scalar components, while acoustics only involves one scalar and one vector field carrying *four* components [8].

2 Literature review

The following is a review and discussion on methods for achieving high physical accuracy in modeling audio radiation from a modally vibrating 3D surface, given various assumptions and restrictions.

Let's consider an interactive digital media environment, such as a video game or digital audio workstation. In this setting, we may only require a sound pressure level estimate at a *single* listener position. Furthermore, the accuracy of our estimate need only be perceptually "close" [7] to the real thing. What combinations of model assumptions, environment limitations, and restrictions on dynamics enable real-time generation of audio streams with high perceived physical realism in the computationally restricted (but capable!) setting of e.g. a current-generation consumer laptop? What if we help ourselves to current generation discrete GPUs?

In this section, we start with the simplest case of a point-like audio source with a single listener, and gradually add detail, lift environmental constraints, or generalize the computational model, to widen the range of applicability and improve physical accuracy.

2.1 A point listens to a noisy point

If we ignore the surface geometry of the sound-emitting object, and instead assume a sound originates from a single point \mathbf{x}_o in a homogenous medium, we could fully and accurately model how the sound radiates to a single (also point-like) listener point \mathbf{x}_l using only *attenuation* and *delay* (phase):

$$P(\mathbf{x}_l, t) = \frac{P_0}{4\pi r^2} \cdot e^{j\omega(t - \frac{r}{c})}$$

- $P(\mathbf{x}_l, t)$ is the sound pressure at the listener's location at time t .
- P_0 is the initial sound pressure at the source.
- r is the distance between the source and listener, $r = \|\mathbf{x}_l - \mathbf{x}_o\|$.
- ω is the angular frequency of the sound wave.
- c is the speed of sound in the medium.
- The term $\frac{P_0}{4\pi r^2}$ represents the attenuation factor.
- The term $e^{j\omega(t - \frac{r}{c})}$ accounts for the phase change (time delay) as the wave travels from the source to the listener, with $\frac{r}{c}$ being the time it takes for the sound to cover distance r at speed c .

This represents the case of modeling radiation *only*, with no environmental geometry whatsoever, including the sound-generating object or the listener, which both have zero spatial extent.

2.2 Two linear time-invariant filters listen to a noisy point (HRTFs)

Ignoring the nonphysical notion of a point magically generating sound, let's first improve on the point-like audio measurement apparatus.

For a long time [11], it has been recognized that the way sound interacts with the shape of our ears, head, shoulders, and torso has an important effect on how we *localize* sounds, or estimate their

relative position. Specifically, the shape of this "head-related" anatomy causes small frequency-dependent deviations in arrival time to the two ear canals of a listener, and the auditory system uses this characteristic pattern to inform its localization estimates.

If we place small microphones in both ears of a subject, and record their frequency-dependent response to an audio signal played in the environment, we can measure the *head-related transfer functions* (HRTFs) for both ears. We can improve the perceived realism of our model with low computational cost by changing the single listening point into two linear time-invariant (LTI) filters designed to match these transfer functions, placed the correct distance apart in the scene [10].

2.3 A contact microphone records an object's surface

Sound waves originate from a vibrating object via interactions between the object's surface and the particles in the surrounding medium. The directionality and intensity of the pressure waves emanating from the object are determined by the normal velocity (normal to the surface geometry) of each point on the object's surface. In this section, we ignore the details of sound radiation, and consider what a listener directly on the surface of a single vibrating object would "hear". This is sometimes referred to as "direct sampling", and is analagous to recording an instrument using a contact microphone attached directly to its surface instead of a microphone placed nearby.

A popular approach to modeling the distribution of normal velocities around a vibrating object's surface is *linear modal analysis* (LMA) [3, 9, 11]. Similar to the *small-scale acoustics* assumption (described above) of linearity in the *medium's* response to pressure waves propagating through it, LMA assumes an object's surface vibrations are a linear combination of independent oscillations at different frequencies and amplitudes. (The object is said to vibrate *modally*, or in *pure harmonic oscillation*. Note that by the Fourier Theorem, this assumption is physically accurate at the limit of spatial resolution and number of sinusoids, but LMA is concerned with low-rank decompositions.)

LMA approximates small vibrations in a low-dimensional basis, as a set of uncoupled oscillators reacting to external forces $f(t)$:

$$\ddot{\mathbf{q}}(t) + \tilde{C}\dot{\mathbf{q}}(t) + \tilde{K}\mathbf{q}(t) = U^T f(t),$$

where $\mathbf{q} \in \mathcal{R}^M$ are the modal displacements, \tilde{C} and \tilde{K} are the $M \times M$ reduced damping and stiffness matrices, and $U \in \mathcal{R}^{3N \times M}$ is the time-invariant eigenmode matrix. We can use the Finite Element Method (FEM) to compute the mass and stiffness matrices, and precompute the matrix $U_n \in \mathcal{R}^{N \times M}$ of the eigenmode displacements using an eigenvalue solver, from which we can quickly compute surface accelerations with a sparse $\mathbf{u}_i^T \mathbf{q}$ lookup [13].

2.4 HRTFs listen to vibrating objects in an anechoic chamber

However, we ignored sound radiation completely and directly sampled the acceleration of a point (or points) on the object's surface. Next, we discuss modeling acoustic radiation from the object's surface to a listener.

Audio waves emanating from an object interact with its surface geometry, which introduces frequency-dependent directivity and intensity of sound radiation. A resonating object produces audio waves with wavelengths comparable in scale to features of its surface geometry, the radiated audio will interact with itself and can affect the audio it produces in sometimes dramatic ways (consider the sound of bowl spinning upside-down on a surface) [12].

We can accurately approximate the pattern of acoustic radiation using the *frequency-domain Helmholtz equation* and the *Boundary Element Method* (BEM) [13]. The Boundary Element Method (BEM) solves the Helmholtz equation for problems involving sound radiation and scattering. BEM models the object's surface by discretizing it into a mesh of boundary elements, and computes the sound field at any point in space, given the vibration characteristics of the object's surface, which we can provide using the linear modal analysis method described above.

This method can accurately model frequency-dependent directivity patterns and intensity variations in the radiated sound field, and it can model sound radiation from *multiple objects with complex geometries*. BEM only requires discretization of the object's surface, not the entire volume of the space, and is thus efficient for settings like this with one or more objects in an anechoic environment.

It's also worth noting that we can add multiple listener positions "for free" with this method, since BEM solves for the field across the entire domain.

However, many perceptible acoustic phenomena cannot be captured using this approach ([13]). It assumes linearized small-amplitude modal dynamics, and thus cannot model highly nonlinear phenomena such as sharp transients following contact events. (Changxi Zheng and Doug James propose improvements to the LMA model to specifically address this issue by introducing contact-dependent modal damping and separately treating frictional contact [15].) Nearby objects cannot influence each other acoustically, and it is not practical for enclosed spaces with many reflections. It also requires re-solving whenever geometries change, so it is not well suited for dynamic environments.

Finally, BEM solvers are computationally expensive. Precomputed Acoustic Transfer methods were introduced to solve this problem [5], and later FFAT Maps [2], which encode angular radiation fields in textures precomputed using FastBEM [6]. FFAT Maps enable constant-time transfer evaluation, but can use hundreds of megabytes of storage [12].

Despite the limitations in capturing complex or high-velocity impact events, friction forces, etc., this model can produce accurate results for low-velocity point impacts on the surface of stiff objects. At this point we are at the forefront of acoustic modeling as of 2010! The audio simulation pipelines used in two papers by Changxi Zheng and Doug James from 2010 and 2011 [14, 15], both with impressive results, share the core architecture laid out up to this section:

- tetrahedral mesh objects
- FEM/eigenvalue-precomputed damped modal oscillators
- excitations produced by interactive contact impulses
- sound radiation approximated by Helmholtz and solved by BEM

- HRTFs for sound rendering.

2.5 FDTD: Anything you want if you have the time

I discussed the FDTD method in the introduction, and I point to *Schneider, 2010* [8] for a detailed discussion of how to discretize the 3D pressure and velocity PDEs into a grid and solve numerically.

Simulating the wave equation is expensive enough to be out of the question for real-time sound rendering at useful resolutions even with a modern GPU (although real-time GPU performance on bandlimited signals or 2D problems was demonstrated as far back as 2015 [1]). Jui-Hsien Wang and Doug James present a highly specialized and fine-tuned GPU-accelerated FDTD transfer solver in their KleinPAT work [12], but it is far from general.

In this project, I plan to use FDTD (if time allows) as a source-of-truth to compare other methods against.

3 Computational methods

When applying the FDTD, acceleration boundary conditions for each cell on the object boundary can be computed using the FEM-derived modal vibration model described above. The section above discusses many other computational details.

4 Simulation setup

For sound synthesis using the FDTD method, the spatial domain may be relatively small (e.g., 80^3 cells), but millions of sequential timesteps [13]. This limits opportunities for parallelization. However, I will be investigating several computational methods other than FDTD, all within an interactive application, simulating environments smaller than 10 meters in any dimension.

5 Quantities to inspect

The primary quantity to inspect in this project is the sound pressure level (SPL) over time. The time-varying SPL at the listener's position directly corresponds to the perceived audio signal. This quantity must be sampled at a rate at least double the highest resolved frequency, which should be reasonably close to the upper bound of typical human hearing, commonly estimated as 20kHz. 44.1kHz is a common sample rate for high-quality audio, and 16kHz may be considered a reasonable lower bound for meaningful audio synthesis.

I plan to measure accuracy using common time-domain loss functions like MSE, as well frequency-domain losses. I also plan to use perceptual loss functions such as the "multi-scale

spectral loss" described in Google Magenta's DDSP paper [4], which simply combines multiple STFT results with variable resolution.

References

- [1] A. Allen and N. Raghuvanshi. Aerophones in flatland: Interactive wave simulation of wind instruments. *ACM Transactions on Graphics*, 34(4), Aug 2015. Proceedings of SIGGRAPH 2015.
- [2] J. N. Chadwick, S. S. An, and D. L. James. Harmonic shells: A practical nonlinear sound model for near-rigid thin shells. *ACM Transactions on Graphics*, August 2009.
- [3] P. R. Cook. Sound production and modeling. *IEEE Computer Graphics & Applications*, 22(4):23–27, July/Aug. 2002.
- [4] Jesse Engel, Lamtharn (Hanoi) Hantrakul, Chenjie Gu, and Adam Roberts. Ddsp: Differentiable digital signal processing. In *International Conference on Learning Representations*, 2020.
- [5] Doug L James, Jernej Barbic, and Dinesh K Pai. Precomputed acoustic transfer: Output-sensitive, accurate sound generation for geometrically complex vibration sources. *ACM Transactions on Graphics (TOG)*, 25(3):987–995, 2006.
- [6] Y. J. Liu. *Fast Multipole Boundary Element Method: Theory and Applications in Engineering*. Cambridge University Press, Cambridge, 2009.
- [7] T. Painter and A. Spanias. Perceptual coding of digital audio. *Proceedings of the IEEE*, 88(4):451–515, 2000.
- [8] John B. Schneider. *Understanding the Finite-Difference Time-Domain Method*, chapter 12. 2010.
- [9] A. A. Shabana. *Theory of Vibration: An Introduction*. Springer Science & Business Media, 2012.
- [10] R.H.Y. So, N.M. Leung, J. Braasch, and K.L. Leung. A low cost, non-individualized surround sound system based upon head-related transfer functions: An ergonomics study and prototype development. *Applied Ergonomics*, 37:695–707, 2006.
- [11] Daniel Starch. *Perimetry of the Localization of Sound*. Ph.d. dissertation, State University of Iowa, 1908.
- [12] Jui-Hsien Wang and Doug L. James. KleinPAT: optimal mode conflation for time-domain precomputation of acoustic transfer. *ACM Transactions on Graphics*, 38(4), August 2019.
- [13] Jui-Hsien Wang, Ante Qu, Timothy R. Langlois, and Doug L. James. Toward wave-based sound synthesis for computer animation. *ACM Trans. Graph.*, 37(4):109:1–109:16, 2018.

- [14] Changxi Zheng and Doug L. James. Rigid-body fracture sound with precomputed sound-banks. *ACM Transactions on Graphics (Proceedings of SIGGRAPH 2010)*, 29(3), July 2010.
- [15] Changxi Zheng and Doug L. James. Toward high-quality modal contact sound. *ACM Transactions on Graphics (Proceedings of SIGGRAPH 2011)*, 30(4), August 2011.