Audio Engineering Society

Conference Paper

Presented at the Conference on
Audio for Virtual and Augmented Reality
2020 August 17 – 19, Redmond, WA, USA

Fast acoustic obstruction with proximity cost differencing (PCD)

Keith W. Godin[1] and Nikunj Raghuvanshi[2]

[1]*Microsoft Mixed Reality*
[2]*Microsoft Research*

Correspondence should be addressed to Keith W. Godin (`kegodin@microsoft.com`)

ABSTRACT

We propose a fast heuristic for acoustic obstruction in virtual reality and games. Our method is designed to model the dependence of obstruction on the size of both obstructions and apertures (like doors) without explicit markup, while operating at an order-of-magnitude speedup compared to a ground-truth wave simulation. Our technique is robust to scene complexity, making it suitable for interactive applications. While this initial study is limited to two dimensions, generalization to three dimensions is possible.

1 Introduction

Immersive virtual reality and games face the challenging problem of *acoustic obstruction*, which is the loss in acoustic energy on the initial wavefront emitted from a source as it propagates through a complex scene to reach the listener's location. Acoustic obstruction is critical for maintaining audio-visual consistency in interactive applications, such as when the source and listener are separated by a doorway, or when an object obstructs line-of-sight. For instance, as a talker in another room walks into the listener's room through a door, the expectation is a smooth ascent in loudness that lets the listener anticipate when the talker might become visible.

Acoustic obstruction exhibits spatially smooth, frequency-dependent behavior. Waves can diffract efficiently around small (compared to wavelength) obstructions, while small apertures have the opposite effect. Transmitted energy progressively weakens as an aperture's size is reduced. Such effects are widespread and easily perceived for audible sound because of its wavelength in the centimeter to meter range. These behaviors are quite unlike visual obstruction which is well-approximated by binary visibility owing to the sub-micron wavelength of visible light. Visual obstruction can be efficiently checked with a single ray-cast from source to listener. Thus, fast diffraction approximation becomes the central challenge in estimating acoustic obstruction at interactive rates.

For an accurate solution, one must solve the fundamental wave equation on a grid, which prior work [1] has shown can be viable to provide obstruction and other acoustic effects in games [2] and VR [3]. However, extensive, cluster-scale resources are required for precomputation, which can be a hindrance for many practical use cases and limits to largely static scenes. Further, the simulations are band-limited to 500Hz in practice due to computation time constraints, as wave solvers must sample space and time at a resolution proportional to the maximum modeled frequency.

On the other hand, common techniques in games prioritize for speed, using geometric heuristics that sacrifice accuracy, with varying degrees of plausibility. A common approach is using visual obstruction for audio. But then a small obstructing object occludes as much as a concrete wall. We call this the "lamppost problem". A better technique is to employ shortest path ratio relative to Euclidean distance, but the effect of aperture size cannot be captured. Thus a door that is slightly ajar transmits as much energy as when fully open. This is the "pinhole problem": small gaps in geometry can result in implausible, immersion-breaking jumps in loudness.

Thus, existing methods either do computationally-expensive physical solution, or use simplified heuristics that are fast but miss easily-audible diffraction effects that depend on obstruction and aperture size in relation to wavelength. In this paper we discuss a new heuristic method that strives to strike a balance: it qualitatively models such effects in complex environments, while reducing the amount of computation required by orders of magnitude compared to a wave solution. Our method also avoids the adverse computational scaling with frequency inherent in wave solvers. This initial study restricts to 2D for simplicity, although the method generalizes well to 3D.

2 Related work

In the field of room acoustics [4], there is a large body of literature on interactive auralization [5], which addresses the broad problem of rendering the complete acoustic impulse response between a source and listener in an environment. There are two general approaches. Wave-based techniques [6, 1] are accurate, directly solving the underlying wave equation, naturally modeling diffraction to yield robust obstruction in complex scenes. But this comes at the cost of large amounts of pre-computation.

Geometric Acoustics (GA) reduces computation by assuming an infinite-frequency limit to the wave equation, tracing rays of sound. Obstruction modeling requires diffraction, which presents a difficulty for GA since obstruction is a deterministic process involving systematic losses introduced by intervening obstructing objects and apertures as the initial wavefront propagates from source to listener. But deterministic ray-tracing has exponential complexity (a ray branches into many each time it hits an edge or surface), so that stochastic path tracing is the norm [7]. This is well-suited for the primary application in room acoustics: reverberation modeling in auditoria where line-of-sight between source (e.g., musical instrument) and listener is largely unobstructed, and reverberation is an inherently stochastic process. However, obstruction presents a challenge since it violates the basic assumption of geometric acoustics and unlike in room acoustics of single enclosures, it is of primary importance in games and virtual reality.

Stochastic approaches to diffraction within GA have been proposed, where edges probabilistically deflect rays passing nearby [8] but it is not clear how many rays may be needed for converged estimates that do not have loudness fluctuation artifacts from insufficient ray-tracing, especially in cases that might require large (and thus, improbable) deflections. Obstruction modeling with stochastic GA in complex scenes that can have arbitrary numbers of obstructions and apertures remains an open problem, see [9] for a survey of recent GA approaches.

In the past there has been work [10] on deterministic obstruction modeling using the idea of Fresnel Zones used in the radio propagation community. But the method only works for a single obstruction or aperture, whereas we wish for our heuristic to work in general scenes. Another deterministic approach is beam tracing, that combined with the Uniform Theory of Diffraction [11] could be used for deterministic modeling of obstruction. But there are two issues. Firstly, such methods rely on analytical solutions for infinite wedges, but cannot account for the effect of apertures. In these methods, a straight line path through a small aperture (pinhole) goes unimpeded, while in reality aperture loss increases smoothly as aperture size decreases. Secondly, beam tracing suffers from exponential computational scaling, same as deterministic ray tracing, which does not scale well in complex scenes common today. While scene simplification is possible in principle, automatic acoustics-aware geometry simplification is itself a difficult open problem with limited work [12].

Recently, machine learning approaches have been investigated for fast obstruction and scattering from rectangular plates using a small neural net [13] and later, using a large convolutional neural net that allows arbitrary convex prism shapes [14]. These approaches are restricted to obstruction from a free-standing single object while we are interested in the fully general class of indoor and outdoor scenes.

In the game development community, many heuristics have been proposed, but perhaps the most common method that yields spatially smooth results is to utilize the ratio of shortest path length to the Euclidean distance, e.g., as discussed in Neumann and Lawlor [15]. Our method is in spirit of such heuristics, but gets closer to the qualitative behaviors of diffraction as we show in our results, especially by addressing the small aperture (pinhole) problem which existing heuristics do not handle.

3 Problem

We define our problem as the estimation of the acoustic obstruction g_{obs}. Following Godin et al. [16], we assume a monopole source and normalize the impulse response $h[n]$ such that $\sum_n h^2[n] = 1$ at a distance of 1m in the free field, then define obstruction as:

$$g_{obs} = d\sqrt{\sum_{n=n_1}^{n_2} h^2[n]} \qquad (1)$$

where $h[n]$ is the acoustic impulse response, d is the Euclidean distance from source to listener, n_1 is the onset of the first arriving wavefront, and n_2 is the end of the perceptual integration period following initial onset, typically 10ms. We also define its decibel equivalent:

$$G_{obs} = 20\log_{10}(g_{obs}) \qquad (2)$$

In this formulation, G_{obs} is 0dB throughout space in the absence of geometry. Defined in this way, G_{obs} fits neatly into widely available interactive frameworks for audio design [16] since obstruction loss can be applied as an additional gain alongside existing processing for distance attenuation.

Diffraction plays an important role in determining G_{obs}, depending strongly on wavelength and its relation to geometric feature size. In general, most practical issues can be understood as a combination of two problems.

Obstructions (the "lamppost" problem) A given obstructing object attenuates wavelengths smaller than the object much more than it does wavelengths larger than the object. Obstruction methods that take the high-frequency (small wavelength) limit allow thin obstructions (such as lampposts) to cast deep but narrow shadows. The practical import of this problem is that, for a listener moving about the scene, the source appears to (implausibly) switch on and off rapidly. We refer to this as the "lamppost problem".

Apertures (the "pinhole" problem) A related problem occurs on complementary geometry. In the physical world, the energy transmitted by a given aperture decreases with increasing wavelength. Obstruction methods that ignore aperture size allow the tiniest of openings to transport full spectrum loudness. For instance, with such methods, a small and barely-open window radiates sound from outside into the interior as loudly as if the window were very large and fully open. For this reason, users are limited in design choices and must also carefully ensure their meshes are watertight, which can be difficult in practice, such as game levels, or when geometry is acquired via depth scanning in mixed reality. This is the "pinhole problem".

4 Method

4.1 Motivation

The shortest path from source to listener in a visually obstructed case must necessarily detour around objects. Estimating obstruction based on this detour results in smooth variation that addresses the lamppost problem. To address the pinhole problem, we observe that paths through a small aperture necessarily travel near the objects defining the aperture. Accumulating additional attenuation on paths passing near objects should thus address the pinhole problem. Based on these intuitions, our proposed 'Proximity Cost Differencing (PCD)' approach defines a cost field where cost increases close to objects. It then determines obstruction by comparing the cost of the lowest-cost path through this field to the cost of a straight-line path.

4.2 Cost fields

We define fields on a discretized space with a regular grid of step size h ('pixels' in 2D and 'voxels' in 3D, or 'cells' in general). We introduce a parameter p

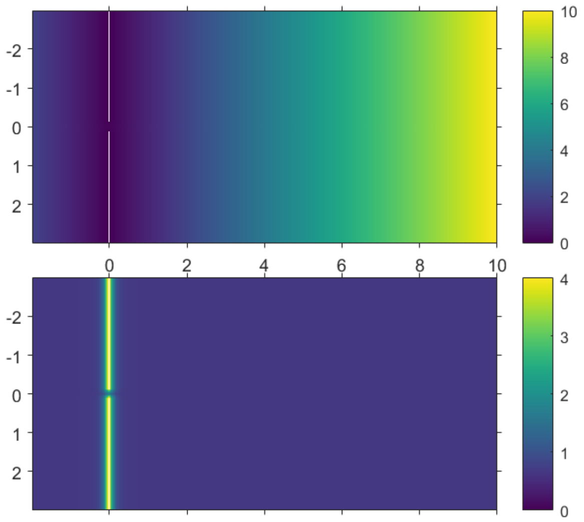

Fig. 1: A scene with two walls with 25cm aperture. Axis units of meters with origin at center of aperture. Top: Distance field D. Bottom: $\log_{10}$ of cost field U with $\lambda = 2.74$m (125Hz at $c = 343$m/s).

to control the per-cell cost of path deviations around objects and a parameter s to control the spatial contour of the resulting obstruction function. Higher values of p and s impose higher cost on deviations from the straight-line path, thus modeling higher frequencies. Normalizing by the number of cells per wavelength (λ/h), we define cost field W representing free-space propagation:

$$W = p^{1/s}\frac{h}{\lambda} \tag{3}$$

Obstruction will be estimated based on comparison of path cost through W to path cost through a geometry-derived cost field U.

We develop cost field U by first defining field D, where each cell is the Euclidean distance to the nearest cell occupied by a scene object. Normalizing D to measure distance in number of wavelengths (D/λ), we extend W with a coefficient that is 1 at infinite distance and increases exponentially when approaching objects:

$$U = p^{1/s}\frac{h}{\lambda}\left(1 + l\frac{\lambda^r}{D^r}\right) \tag{4}$$

Parameters l and r control the rate of increase. Figure 1 shows D and U for an example scene.

4.3 Computing obstruction

We define u and w as the total (integrated) costs of the lowest-cost paths from source to listener through cost fields U and W, respectively. We compute G_{obs} from their difference:

$$G_{obs} = -(u-w)^s \tag{5}$$

Our proposed definitions of U and W guarantee $u \geq w$.

4.4 Field and path cost computation

We compute D, u, and w using the Fast Marching Method (FMM) [17]. FMM bears strong similarities to Dijkstra's more well-known shortest-path graph search method but estimates solutions on the continuous field that underlies the discretized grid. To compute D, all occupied elements of the grid are set as simultaneous starting points. For U and W, FMM's slowness field F is set to $F = U^{-1}$ and $F = W^{-1}$ respectively.

5 Evaluation

5.1 Evaluation data

We constructed canonical occluder and aperture examples in two dimensions as representatives of the lamppost and pinhole problems, respectively. We analyzed these in three octave bands centered at 125, 500, and 2k Hz. We used two sizes for the occluder/aperture: 1-meter and 0.25 meters. The depth of the occluder and of the walls surrounding the aperture was 1 cell. All measurements of G_{obs} with all methods were performed with the same step size $h = 2.13$cm, providing 8x oversampling of the shortest wavelength of interest.

5.2 Ground truth

The Finite Difference Time Domain (FDTD) method of discretizing and numerically solving the acoustic wave equation was used to generate reference fields of G_{obs} over a discretized grid. Separate simulations were run for each octave band of interest. We used a soft source with a stimulus derived from a Parks-McClellan filter designed to fill the octave band. The n_1 value from Eq. 1 at each cell in the discretized grid was found using FMM, with n_2 defined as $n_1 + \frac{10e-3}{T}$, with $T = 44.2\mu$s the temporal simulation step. For 2D simulation, we modify Eq. 1 to account for circular rather than spherical power loss with distance:

$$g_{obs} = \sqrt{d\sum_{n=n_1}^{n_2} h^2[n]} \tag{6}$$

5.3 Path ratio baseline

We implemented a shortest path ratio obstruction metric for comparison, similar to [15]. It is defined as:

$$G_{obs} = -s\left(\frac{d}{r} - 1\right) \quad (7)$$

for shortest path length d, Euclidean distance r, and obstruction factor s. We tuned the value of s separately for each frequency band of interest based on a subjective comparison to FDTD of its performance on the 1-meter lamppost problem. The chosen s values for the respective frequency bands were 80, 160, and 320.

We measured shortest path length d using FMM by setting the speed field to 1 in free space regions. As with PCD, the line-of-sight path was also measured using FMM, rather than Euclidean distance, because small deviations in FMM from Euclidean distance (on the order of 1e-5) are amplified by the G_{obs} formulation to perceptually significant levels. Because this free-space field is computed once for all scenes and stored, its runtime cost is only that of a lookup table.

5.4 PCD parameter selection

The PCD parameters p, l, r and s were tuned by visual comparison with the FDTD reference on the 1-meter pinhole case. The values were $p = 24$, $l = 1e-5$, $r = 4$, and $s = 0.5$. These values were then held constant in all test cases, i.e., no per-case tuning was performed.

6 Results

Figures 2 and 3 compare G_{obs} fields computed using the shortest path ratio method, the proposed PCD method, and FDTD on the 1-meter lamppost and pinhole problems, respectively, for three octave bands, centered at 125 Hz, 500 Hz, and 2 kHz. Figures 4 and 5 compare the three methods on 0.25-meter obstructions and apertures, respectively.

All three methods successfully address the lamppost cases shown in Figures 2 and 4, in that they have soft shadow transition regions. However, the shadow computed by the shortest path method fades significantly sooner than in the PCD and FDTD cases. Further, the conical shape of the FDTD obstruction loss region, especially in the 125Hz case, shows that diffraction causes audible obstruction loss even in some line-of-sight regions that are near obstructing objects. PCD is capable of modeling some of these effects as seen in the 6-10 meter region of the 125Hz case, due to the cost field extending past the geometry to affect the cost of otherwise line-of-sight paths.

In the pinhole cases, the shortest path ratio method has a region of full brightness of width equal at least to the aperture size. In contrast, in the 0.25m aperture case for 125 and 500Hz, PCD and FDTD restrict energy through the aperture and lack a full brightness region.

6.1 Cost comparison

The goal of the PCD method is to estimate obstruction at lower cost than numerical solution of the wave equation. FMM needs to only track the first wavefront as it sweeps through the scene, while any volumetric solver must update the entire volume at each time-step.

For a square 2D scene of edge length N meters, a discretized grid has $(N/h)^2$ elements. For PCD, the global distance field D needs to be computed only once, so its cost is amortized across all source probe locations of interest, of which there may be scores or hundreds. PCD runs FMM once for each frequency band for each source probe. FMM visits every element once and maintains a min heap to select elements for update, for a total cost of $O\left(\left(\frac{N}{h}\right)^2 \log N\right)$ operations [17].

FDTD must update each cell on each time step. On average it takes $O(N/h)$ time steps for the initial wave to reach every corner of the scene, for a total 2D FDTD cost of $O\left(\left(\frac{N}{h}\right)^3\right)$.

The constants involved in both techniques are similar order of magnitude. Thus, our technique gives a speedup of $O\left(\frac{N}{h \log N}\right)$. For our simulations going up to 2kHz octave band, FDTD required a cell size of $h = 2.13cm$. Assuming the smallest practical scene size of $N = 10m$, our technique saves computation by roughly a factor of 100, increasing with N.

The resulting speedups may be even larger, because wave solvers require h to decrease in inverse proportion to modeled frequencies, while our method instead requires only that h be based on the size of features of interest such as the smallest aperture. In addition, wave solvers require an absorbing boundary condition at the edge of the simulation region that is not required by our technique.

7 Conclusion

Acoustic obstruction systems for interactive media and virtual reality experiences must be robust and reliable on complex scenes. They must fit practical use cases and avoid corner cases that unpleasantly surprise designers and users alike who are not experts in acoustics. We have characterized these requirements in part by defining instances of narrow occluders and apertures. We have demonstrated on 2D scenes that a path cost difference through a distance-related cost field can be used to estimate acoustic obstruction that approximates diffraction. Our method produces obstruction that is robustly smooth to small occluders and small apertures, at an order-of-magnitude speedup compared to ground truth wave simulation.

Future work may include investigation of the technique on three-dimensional scenes, further refinement or alternative proposals for methods to compute the cost fields U and W, such as to avoid obstruction on paths travelling parallel to walls in close proximity, and use of numerical modeling to tune the parameters of the proposed method.

References

[1] Raghuvanshi, N. and Snyder, J., "Parametric Directional Coding for Precomputed Sound Propagation," *ACM Trans. Graph.*, 37(4), pp. 108:1–108:14, 2018.

[2] Raghuvanshi, N., Tennant, J., and Snyder, J., "Triton: Practical pre-computed sound propagation for games and virtual reality," *J. Acoustical Soc. of Am.*, 141(5), pp. 3455–3455, 2017.

[3] Godin, K. W., Rohrer, R., Snyder, J., and Raghuvanshi, N., "Wave Acoustics in a Mixed Reality Shell," in *2018 AES International Conference on Audio for Virtual and Augmented Reality (AVAR)*, 2018.

[4] Kuttruff, H., *Room Acoustics*, Taylor & Francis, 4th edition, 2000.

[5] Vorländer, M., *Auralization: Fundamentals of Acoustics, Modelling, Simulation, Algorithms and Acoustic Virtual Reality*, Springer, 1 edition, 2007.

[6] Raghuvanshi, N., Snyder, J., Mehra, R., Lin, M. C., and Govindaraju, N. K., "Precomputed Wave Simulation for Real-Time Sound Propagation of Dynamic Sources in Complex Scenes," *ACM Trans. on Graph.*, 29(3), 2010.

[7] Schröder, D., *Physically Based Real-Time Auralization of Interactive Virtual Environments*, Logos Verlag, 2011.

[8] Stephenson, U. M. and Svensson, U. P., "An improved energetic approach to diffraction based on the uncertainty principle," in *19th Int. Cong. on Acoustics (ICA)*, 2007.

[9] Savioja, L. and Svensson, U. P., "Overview of geometrical room acoustic modeling techniques," *J. Acoustical Soc. of Am.*, 138(2), pp. 708–730, 2015.

[10] Tsingos, N. and Gascuel, J.-D., "Soundtracks for computer animation: sound rendering in dynamic environments with occlusions," in *Proc. Graph. Interface 1997 Conf.*, pp. 9–16, 1997.

[11] Tsingos, N., Funkhouser, T., Ngan, A., and Carlbom, I., "Modeling acoustics in virtual environments using the uniform theory of diffraction," in *SIGGRAPH '01*, pp. 545–552, ACM, New York, NY, USA, 2001.

[12] Siltanen, S., *Geometry Reduction in Room Acoustics Modeling*, Master's thesis, Helsinki University of Technology, 2005.

[13] Pulkki, V. and Svensson, U. P., "Machine-learning-based estimation and rendering of scattering in virtual reality," *The Journal of the Acoustical Society of America*, 145(4), pp. 2664–2676, 2019, doi:10.1121/1.5095875.

[14] Fan, Z., Vineet, V., Gamper, H., and Raghuvanshi, N., "Fast Acoustic Scattering Using Convolutional Neural Networks," in *ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 171–175, 2020.

[15] Neumann, T. and Lawlor, S., "Overwatch - The Elusive Goal: Play by Sound," Talk at Game Developers Conference, 2016, available at: `https://www.gdcvault.com/play/1023317/Overwatch-The-Elusive-GoalPlay`.

[16] Godin, K. W., Gamper, H., and Raghuvanshi, N., "Aesthetic modification of room impulse responses for interactive auralization," in *AES Conf. on Immersive and Interactive Audio (IIA)*, 2019.

[17] Sethian, J. A., "A fast marching level set method for monotonically advancing fronts," *Appl. Math.*, 1996.

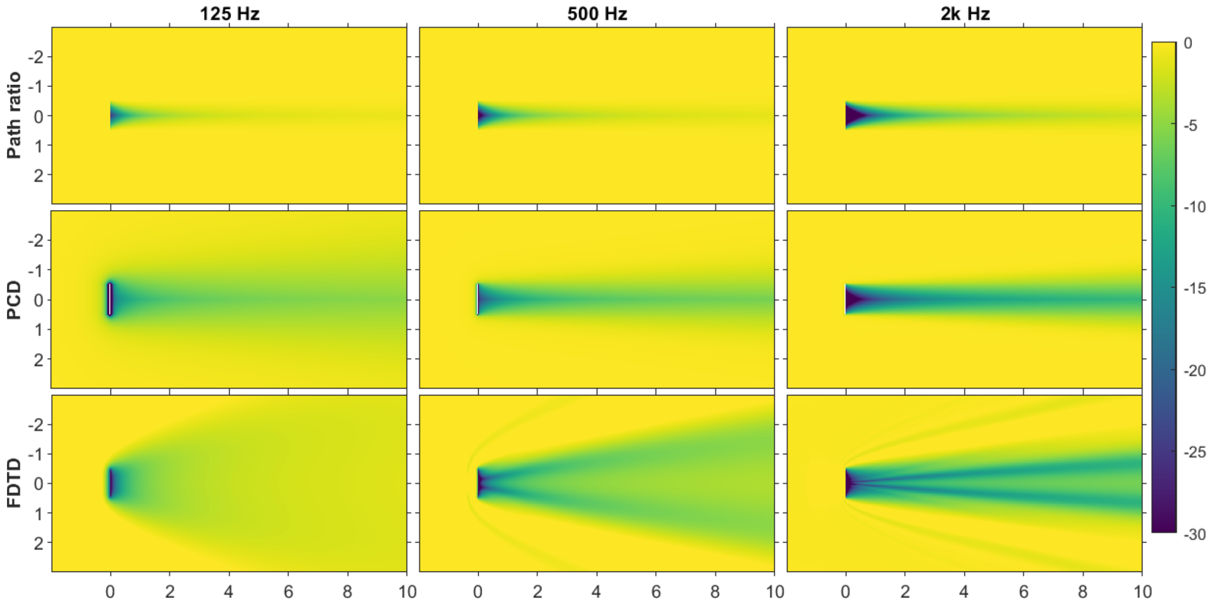

Fig. 2: G_{obs} on the 1-meter lamppost problem. Axes are in meters, origin at the center of the occluder. Source location is $(-13.64, 0)$, which is 5λ for $f = 125$ Hz at $c = 341$ m/s.

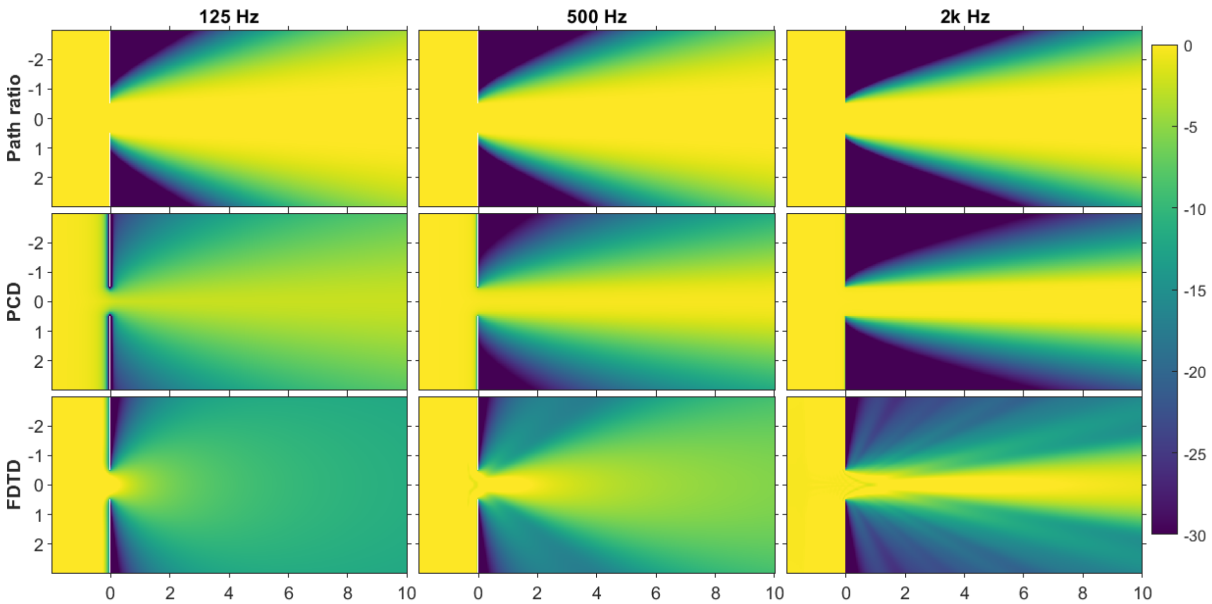

Fig. 3: G_{obs} on the 1-meter pinhole problem. Axes are in meters, origin at the center of the aperture. Source location is $(-13.64, 0)$, which is 5λ for $f = 125$ Hz at $c = 341$ m/s.

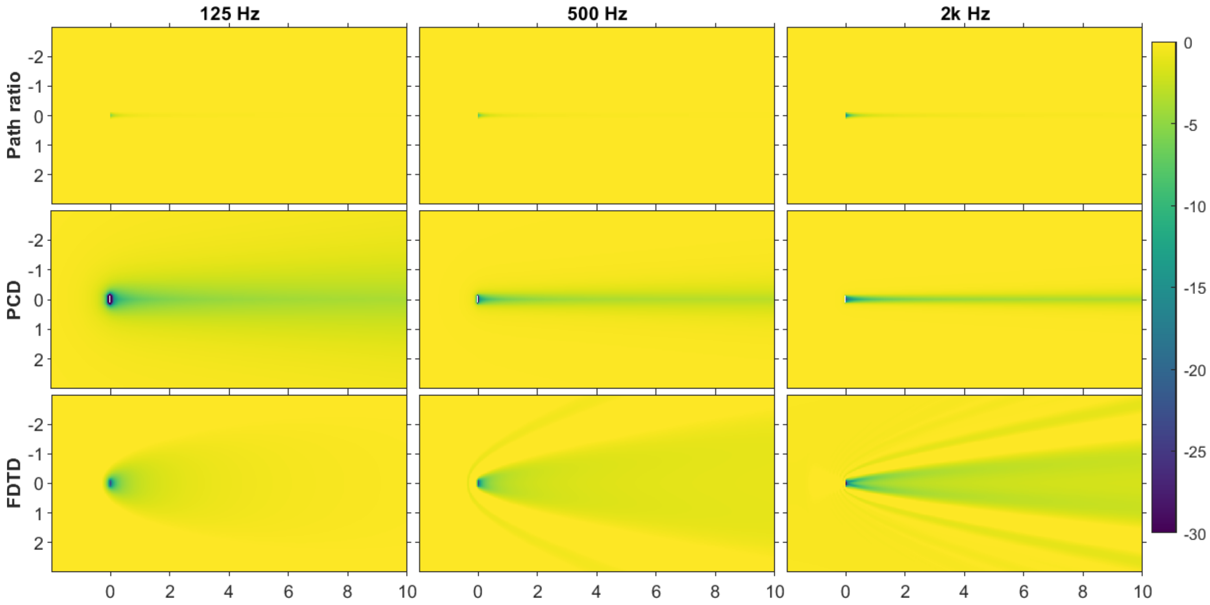

Fig. 4: G_{obs} on the 0.25-meter lamppost problem. Axes are in meters, origin at the center of the 0.25-meter obstruction. Source location is $(-13.64, 0)$, which is 5λ for $f = 125$ Hz at $c = 341$ m/s.

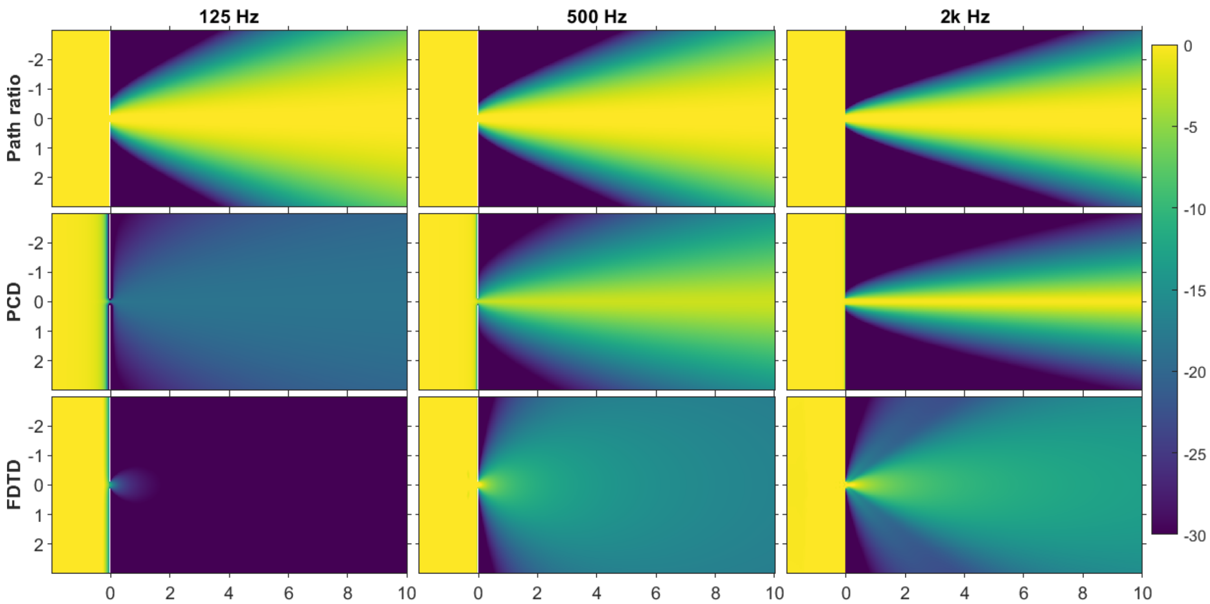

Fig. 5: G_{obs} on the 0.25-meter pinhole problem. Axes are in meters, origin at the center of the 0.25-meter aperture. Source location is $(-13.64, 0)$, which is 5λ for $f = 125$ Hz at $c = 341$ m/s.