

幾何音響シミュレーションと機械学習を用いた 2次元領域に対する波動伝搬予測手法の検討

Study on Wave Propagation Prediction Method for a Two-Dimensional Domain Using Geometrical Acoustics Simulation and Machine Learning

芦澤 剛^{1*} Yen Benjamin^{2,1} 中臺 一博¹

¹ 東京科学大学 システム制御系

² 理化学研究所 生命機能科学研究センター

¹ Department of Systems and Control Engineering, Institute of Science Tokyo

² RIKEN BDR

Abstract: 音響 AI モデルの学習を行う際には、音の波動的な振る舞いが表現されているデータが必要となるが、実環境において多様な条件でのデータを収集することは現実には困難である。また、有限差分時間領域 (FDTD) 法をはじめとする波動音響シミュレーションを実施する場合でも、大規模な解析モデルに対して高周波領域まで解析しようとする、依然として計算コストが高い。そこで本研究では、幾何音響 (GA) シミュレーションの一つである音線法 (Ray Tracing 法) を入力とし、画像間変換モデルを用いて波動音響シミュレーションの結果を予測する Ray2Wave を提案する。本稿では、提案手法の有効性を検証するために、2次元の矩形モデルに対して、音源や障害物の異なる 20 ケースに対して Ray Tracing と FDTD のペア画像を生成し、そのうち 16 ケースを学習データ、4 ケースをテストデータとして用いた機械学習を行った。今回学習した機械学習モデルでは、音線が存在する領域では全体的な伝搬挙動を再現でき、障害物がない解析モデルに対しては推定誤差が小さかった。その一方、障害物を有する解析モデルに対する結果では、特に障害物周囲での推定精度が低下した。これは Ray Tracing の解析結果が回折を表現できていないことに起因すると考えられる。今後はこのようなエラーに着目して、機械学習モデルの改善に取り組む予定である。

1 はじめに

ロボット聴覚における音声認識や音源定位といったタスクは、AI や機械学習の発展によって進歩している。機械学習モデルの学習を行う際には、音波は反射や回折、干渉などの影響を受けるため、これらの波動的な振る舞いを踏まえたデータが必要となる。しかしながら、例えば屋内の限られた条件下での学習データは計測可能である一方 [1]、多様な室環境や、さらに屋外まで想定した多様な環境下での大規模なデータセットを実測にて収集することは、あまり現実的ではない [2, 3]。そこで、シミュレーションで機械学習モデルの学習を行い実環境に適用する simulation-to-real (Sim2Real) が有効な代替手段となり得る [4]。

数値音響シミュレーション手法は、大きく幾何音響シミュレーションと波動音響シミュレーションに分類され

る。音線法 (Ray Tracing Method) や虚像法 (Image source method) に代表される幾何音響 (GA) 解析手法は計算コストが低い一方で、波動現象を考慮しない [5, 6, 7]。これに対して、FDTD, FEM, BEM などの波動音響 (WBA) は波の伝搬を精密に再現できるが、計算コストが高い [8, 9, 10, 11]。近年の計算資源の発展により、数キロヘルツ帯までの大規模三次元空間の解析が可能になりつつあるが、依然として GA に比べて高コストである。また、低周波数帯域を WBA、高周波数帯域を GA で扱うハイブリッド手法も提案されているが、高周波帯域に関してはあくまで GA の精度に依存することとなり、また、WBA の実施コストを削減できるものではない [12, 13]。

そこで本研究では、幾何音響の結果を入力として、画像変換モデルにより波動伝搬を予測する手法 Ray2Wave を検討する。

本論文の貢献は以下の通りである。

- 低コストな音波の表現として Ray Tracing によ

*連絡先: 東京科学大学
152-8552 東京都目黒区大岡山 2-12-1
E-mail: ashizawa@ra.sc.eng.isct.ac.jp

る解析結果をモデルの入力とすることで、簡易な機械学習モデルでも波動伝搬挙動を予測することを試みた。

- 2次元矩形領域に対して異なる音源位置や障害物を定義して解析を行い、U-netを基本とした機械学習モデルを作成した。
- 障害物のないモデルに対しては比較的精度よく予測できたが、障害物のある場合には正しく予測できなかった。これはRay Tracingが回折を表現できていないことに起因すると考えられ、今後は機械学習モデルの改善が必要である。

2 関連研究

本節では、音響シミュレーションと機械学習を組み合わせた先行研究について確認する。

初期の研究では、深層ネットワークを用いて波動シミュレーションを近似したり[14]、明示的な物理モデルを用いずに振動場を再構成したりする手法が提案された[15]。その後、支配方程式をネットワークに埋め込む枠組みとしてPINN (Physics-Informed Neural Networks)が登場し[16]、SIRENに代表される暗黙的ニューラル表現は、高精細な表現能力を持つことが示された[17]。また、ニューラルサロゲートを用いた波動音響シミュレーションの高速化[18, 19, 20]、PINNによる波動方程式の直接解法[21, 22]、計測データからの音場推定[23, 24, 25, 26]、さらにはインパルス応答(IR)生成やレンダリングに関する研究[27, 28]も進展している。

中でも本研究と特に関連するものとして以下の4つが挙げられる。Ratnarajahらは、GWAデータセットを用い、3Dメッシュと音源・受音点位置からIRを生成するMESH2IRを提案した[28]。Moseleyらは、時空間座標と媒体モデルから長時間の波動場を予測するPINNフレームワークを導入した[21]。Wangらは、複雑構造による散乱場をメッシュ生成なしにPINNで再構成した[22]。Fanらは、物体形状から周波数依存の散乱場をCNNで推定し、計算を大幅に高速化した[29]。

これらの研究は、いずれもサロゲートモデルや特定タスクに対する推定に焦点を置いており、大規模かつ多様な音場に対する一般的な枠組みを対象としてはいない。効率的でスケーラブルな幾何音響は、これまで波動ベース挙動を予測するための中間表現としては利用されてこなかった。Ray2Waveはこの点に着目し、GAを低コストで情報量のある中間表現として活用することで、機械学習モデルの学習が波動特有の現象に集中できるようにすることを試みる。本手法はPINNやサロゲートモデルを置き換えるものではなく、それらと相補的であり、将来的には組み合わせて利用することも可能である。

3 提案手法

3.1 Ray2Waveの概要

図1に、本研究で提案するRay2Waveの概念図を示す。本研究では、入力として幾何音響シミュレーション手法であるRay Tracing解析を用い、出力として波動音響シミュレーション手法であるFDTDを用いた。両者の結果は画像として扱い、image-to-image translationによって変換を行うことを試みた。本節では、解析モデルに加え、Ray Tracing, FDTD, および image-to-image translationの各手順について述べる。

3.2 解析モデル

図2に示すように、本研究では2次元の矩形領域を解析対象とし、その内部に複数の矩形障害物を配置したモデルを構築した。解析モデルに関するパラメータの詳細を表1にまとめる。

音源には、中心周波数 $f_0 = 500$ [Hz] のガウシアンパルス放射する点音源を使用した。音速は $c_{\text{wave}} = 343$ [m/s] と設定した。空間方向の離散化においては、この中心周波数に基づいて、グリッド間隔 $\Delta x = (c_{\text{wave}}/f_0) \cdot (1/20)$ [m] を用いた。

解析領域は x 方向に 600 グリッド、 y 方向に 500 グリッドで構成した。点音源は座標 $(x_{\text{src}}, y_{\text{src}})$ に配置し、障害物は位置 $(x_{\text{obs}}, y_{\text{obs}})$ とサイズ $(w_{\text{obs}}, h_{\text{obs}})$ を持つ矩形として設置した。音源座標・障害物座標・サイズは Table 1 に示す範囲内でランダムに割り当てた。

障害物および解析領域の境界条件は、両解析手法においてサウンドハード境界として設定した。

これらのモデルに対して、幾何音響解析としてRay Tracing法を、波動音響解析としてFDTD法を適用し、同一条件下で音波伝搬の比較を行った。

3.3 Ray tracingの実施手順

本研究で実施したRay Tracingによる解析は次のとおりである。音源は、点音源の座標から全方位へ360本の音線(Ray)を放射することとした。解析は、時間ステップ Δt ごとに音線を逐次的に進めることで行った。障害物および解析領域の境界ではサウンドハード条件を想定し、法線ベクトルに基づいて鏡面反射となるように設定した。各音線については、伝搬の各ステップにおける先端座標を記録し、時系列データとして保存した。

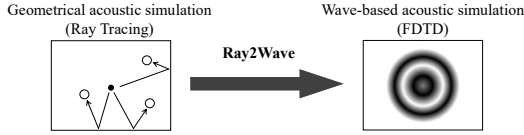


図 1: 提案手法 Ray2Wave の概念図

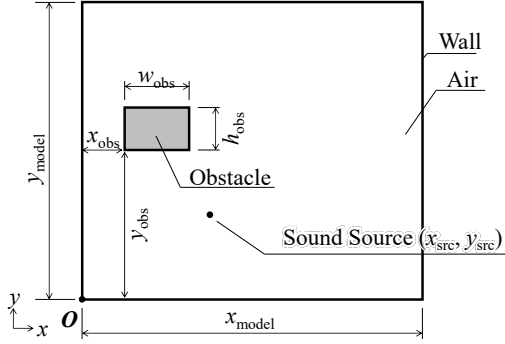


図 2: 解析モデル

表 1: 解析モデルのパラメータ

Center frequency [Hz]	f_0	500
Sound speed [m/s]	c_{wave}	343
Grid spacing [m]	Δx	$(c_{\text{wave}}/f_0) \cdot (1/20)$
Model size [Δx]	$(x_{\text{model}}, y_{\text{model}})$	(600, 500)
Source position [Δx]	$(x_{\text{src}}, y_{\text{src}})$	(50 ~ 450, 50 ~ 550)
Obstacle position x [Δx]	x_{obs}	30 ~ (600 - w_{obs} - 30)
Obstacle position y [Δx]	y_{obs}	30 ~ (500 - h_{obs} - 30)
Obstacle size [Δx]	$(w_{\text{obs}}, h_{\text{obs}})$	(30 ~ 100, 30 ~ 100)
Time Step [s]	Δt	$0.8 \Delta x / (\sqrt{2} c)$

3.4 FDTD の実施手順

次に、本研究で実施した FDTD 法による波動音響解析について述べる。解析は、以下の波動方程式を差分化することで行った。

$$\frac{\partial^2 \phi}{\partial t^2} = c^2 \left(\frac{\partial^2 \phi}{\partial x^2} + \frac{\partial^2 \phi}{\partial y^2} \right) \quad (1)$$

ここで、 ϕ は音圧 p および粒子速度 \mathbf{v} と以下の関係を満たすポテンシャルである。

$$p = -\rho_0 \frac{\partial \phi}{\partial t}, \quad \mathbf{v} = \nabla \phi = \left(\frac{\partial \phi}{\partial x}, \frac{\partial \phi}{\partial y} \right). \quad (2)$$

音源には、中心周波数 f_0 のガウシアンパルスを使用した。障害物および解析領域の境界ではサウンドハード条件とするため、Neumann 境界条件を適用した。

3.5 Image-to-Image translation

幾何音響解析 (Ray Tracing) によって得られた画像から、波動音響解析 (FDTD) による解析結果を予測するために、image-to-image translation モデル [30]

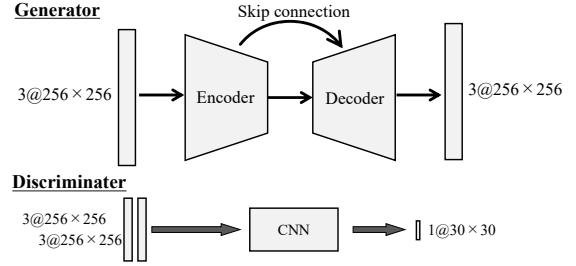


図 3: 画像変換モデル

を用いることを考える。本モデルは Conditional GAN (条件付き生成的敵対ネットワーク) に基づいて構成され、入力画像 (条件画像) から対応する出力画像を生成する。ネットワークの概要を図 3 に示す。本ネットワークは Generator と Discriminator の 2 つの主要な構成要素からなる。

Generator は 8 層の Encoder と、それに対応する 8 層の Decoder から構成された U-Net である。各層のチャンネル数は、入力画像 (RGB, 3 チャンネル) から始まり、Encoder の最深部では 512 チャンネルまで増加し、その後 Decoder で順に減少する構成となっている。また、各層はスキップコネクションによって接続されている。

Discriminator には PatchGAN アーキテクチャを採用し、画像をパッチ単位で分類することで、生成画像を評価する。

学習においては、バッチサイズを 4、エポック数を 100 とした。

4 実験

4.1 データセットの作成

本研究では、音源位置に加えて、障害物の数、座標、サイズを変化させることで、合計 20 ケースの解析モデルを作成した。作成した 20 個のモデルを図 4 に示す。

Case 1 から Case 16 をトレーニングデータ、Case 17 から Case 20 をテストデータとして用いた。各ケースに対して Ray Tracing と FDTD の両方を実行した。

Case 1 における Ray Tracing および FDTD の解析結果の例を図 5 に示す。本実験では Frame #0 ~ #199 の 200 フレームを生成し、出力は $10\Delta t$ ごとに行った。

以上のようにして、トレーニングデータとして $16 \times 200 = 3200$ ペア、テストデータとして $4 \times 200 = 800$ ペアを準備した。

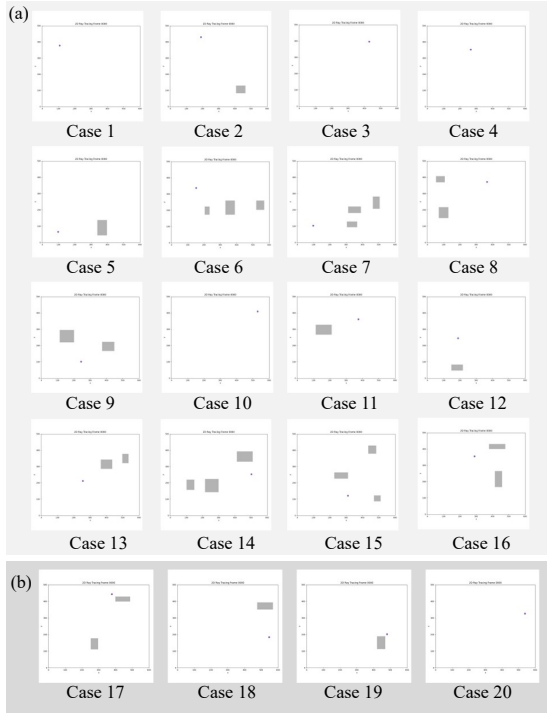


図 4: データセット. Case 1～16: 学習データ, Case 17～20: テストデータ.

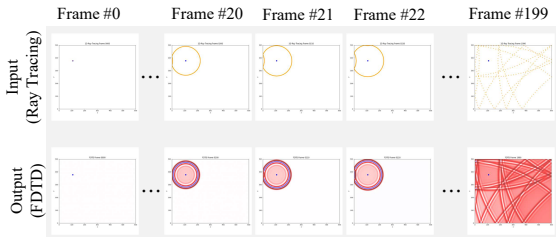


図 5: Case 1 における各フレームのペアの様子

4.2 実験結果と考察

図 6 に学習曲線を示す. Discriminator は不安定な状態が続いており, これは敵対的学習に典型的な挙動である. Generator の損失は一貫して減少したものの, 明確な収束点には達しておらず, 不完全な収束を示唆している.

図 7 に, 学習済みモデルによって得られた波動予測結果を示す. 図では, Case17～Case20 の 4 つの解析条件について, Frame #40 および Frame #80 の結果を例として示している. 左から順に, Ray Tracing による入力画像, 機械学習モデルが生成した予測画像, Ground Truth として用いた FDTD の解析結果, そして生成画像の誤差マップを示している.

まず Case 20 について考える. Case 20 では障害物が存在せず, 音源位置のみがトレーニングデータと異なる条件となっている. 生成された波動場は Ground Truth

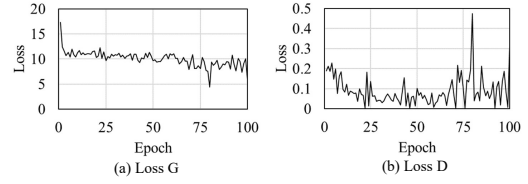


図 6: (a) Generator と (b) Discriminator の損失関数曲線.

とよく一致しており, 良好な再現性が確認できた.

一方, Case 17～19 では音源位置と障害物配置の両方が異なっている. 全体的な波動パターンは類似していたものの, 障害物による回折が再現されず, その結果として差異が生じた.

図 8 に, 全フレームで平均した MSE(平均二乗誤差)を示す. 障害物がない Case 20 では誤差は小さかったが, 障害物を含む Case 17～19 では誤差が大きくなった. これは, Ray Tracing が回折をモデル化せず, 音線が直進のみを扱うため, 入力画像には回折が反映されないことによる. その結果, 音線情報の存在しない領域では予測が困難となった. しかしながら, 音線が到達している領域においては, モデルは波動パターンを再現することができた.

今後, 機械学習モデルの改善を行い, 回折や干渉などの波動的な振る舞いをより正確に推定できるようにすることを試みる予定である.

5 おわりに

本研究では, image-to-image translation モデルを用いて, 幾何音響から波動音響場を予測するフレームワーク Ray2Wave を提案した. 2次元の矩形モデルに対して, Ray Tracing と FDTD を実施し, それらを用いた機械学習を実施したところ, 障害物がない場合や, 音線が明示されている領域では波動伝搬の再現が見られた一方で, 音線が表現しきれない障害物による回折については大きな誤差が生じる結果となった.

今後は機械学習モデルの改善に加え, 広帯域や複数音源の条件へ拡張し, より大規模なシミュレーションデータセットの活用や, 回折の扱いと汎化性能向上を目的とした物理インフォームドモデルの構築を目指す.

参考文献

- [1] Shoichi Koyama, Tomoya Nishida, Keisuke Kimura, Takumi Abe, Natsuki Ueno, and Jesper Brunnström. MeshRIR: A dataset of room impulse responses on meshed grid points for

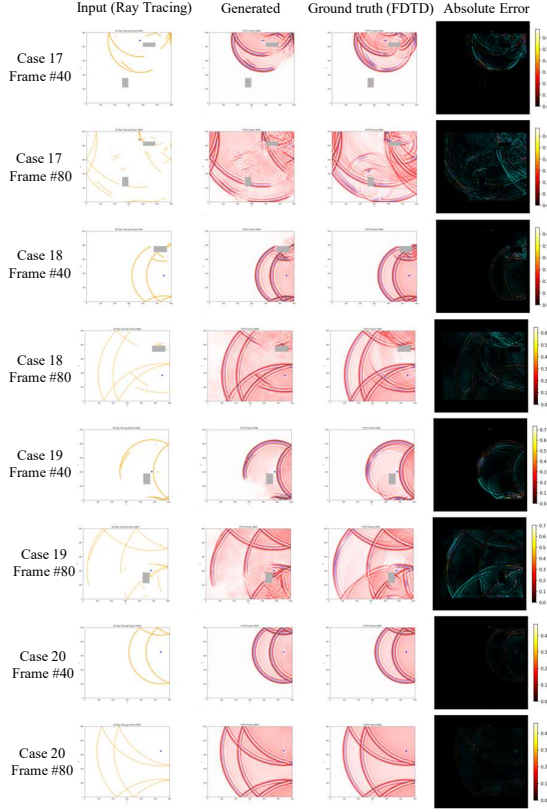


図 7: テストデータに対する予測結果と FDTD との比較 (左から: GA 入力/予測結果/FDTD/誤差)。

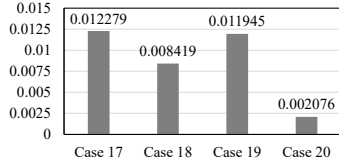


図 8: 各ケースに対する生成画像の MSE

evaluating sound field analysis and synthesis methods. In *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2021.

- [2] Tony F. W. Embleton. Tutorial on sound propagation outdoors. *The Journal of the Acoustical Society of America*, 100(1):31–48, 1996.
- [3] Shiguang Liu and Jin Liu. Outdoor sound propagation based on adaptive fdtd-pe. In *2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pages 859–867, 2020.
- [4] Changan Chen, Carl Schissler, Sanchit Garg,

Philip Kobernik, Alexander Clegg, Paul Calamia, Dhruv Batra, Philip W Robinson, and Kristen Grauman. Soundspaces 2.0: A simulation platform for visual-acoustic learning. In *NeurIPS 2022 Datasets and Benchmarks Track*, 2022.

- [5] Adil Alpkocak and Malik Kemal Sis. Computing impulse response of room acoustics using the ray-tracing method in time domain. *Archives of Acoustics*, 35(4):505–519, 2010.
- [6] Fabian Brinkmann, Lukas Aspöck, David Ackermann, Steffen Lepa, Michael Vorländer, and Stefan Weinzierl. A round robin on room acoustical simulation and auralization. *The Journal of the Acoustical Society of America*, 145(4):2746–2760, 04 2019.
- [7] Haibo Wang, Huimin Gao, and Ming Cai. Simulation of traffic noise both indoors and outdoors based on an integrated geometric acoustics method. *Building and Environment*, 160:106201, 2019.
- [8] Takumi Yoshida, Takeshi Okuzono, and Kimihiro Sakagami. Binaural auralization of room acoustics with a highly scalable wave-based acoustics simulation. *Applied Sciences*, 13(2832), 2023.
- [9] Takao Tsuchiya. Recent techniques on sound field simulation. *Japanese Journal of Applied Physics*, 61(SG0801), 2022.
- [10] Brian Hamilton and Stefan Bilbao. FDTD methods for 3-D room acoustics simulation with high-order accuracy in space and time. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 25(11):2112–2124, 2017.
- [11] Giulia Fratoni, Brian Hamilton, and Dario D’Orazio. Feasibility of a finite-difference time-domain model in large-scale acoustic simulations. *The Journal of the Acoustical Society of America*, 152(1):330–341, 07 2022.
- [12] Masahiro Toyoda and Yuta Sakayoshi. Filter and correction for a hybrid sound field analysis of geometrical and wave-based acoustics. *Acoustical Science and Technology*, 42(4):170–180, 2021.
- [13] Zhenyu Tang, Rohith Aralikatti, Anton Ratnarajah, and Dinesh Manocha. GWA: A LARGE

- HIGH-QUALITY ACOUSTIC DATASET FOR AUDIO PROCESSING. IN *Proceedings of the ACM SIGGRAPH Conference*, PAGES 1–9, 2022.
- [14] BENJAMIN MOSELEY, ANDREW MARKHAM, AND TARJE NISSEN-MEYER. FAST APPROXIMATE SIMULATION OF SEISMIC WAVES WITH DEEP LEARNING. *arXiv preprint arXiv:1807.06873*, 2018.
- [15] MARCO OLIVIERI, FABIO ANTONACCI, AND AUGUSTO SARTI. NEAR-FIELD ACOUSTIC HOLOGRAPHY ON ARBITRARY SHAPES USING CONVOLUTIONAL NEURAL NETWORKS. IN *Proc. INTER-NOISE 2021*, 2021.
- [16] SIFAN WANG, SHYAM SANKARAN, HANWEN WANG, AND PARIS PERDIKARIS. AN EXPERT’S GUIDE TO TRAINING PHYSICS-INFORMED NEURAL NETWORKS. *arXiv preprint arXiv:2308.08468*, 2023.
- [17] VINCENT SITZMANN, JULIEN N. P. MARTEL, ALEXANDER W. BERGMAN, DAVID B. LINDELL, AND GORDON WETZSTEIN. IMPLICIT NEURAL REPRESENTATIONS WITH PERIODIC ACTIVATION FUNCTIONS. *arXiv preprint arXiv:2006.09661*, 2020.
- [18] HE MING YAO AND LI JUN JIANG. MACHINE LEARNING BASED NEURAL NETWORK SOLVING METHODS FOR THE FDTD METHOD. IN *2018 IEEE International Symposium on Antennas and Propagation & USNC/URSI National Radio Science Meeting*, PAGES 2321–2322, 2018.
- [19] H. M. YAO AND W. JIANG. MACHINE LEARNING BASED PML FOR THE FDTD METHOD. *Journal of Electromagnetic Waves and Applications*, 32(13):1729–1742, 2018.
- [20] LI SIAHKOHI, MATHIAS LOUBOUTIN, AND FELIX J. HERRMANN. NEURAL NETWORK AUGMENTED WAVE-EQUATION SIMULATION. *arXiv preprint arXiv:1910.00925*, 2019.
- [21] B. MOSELEY, A. MARKHAM, AND T. NISSEN-MEYER. SOLVING THE WAVE EQUATION WITH PHYSICS-INFORMED DEEP LEARNING. *arXiv preprint arXiv:2006.11894*, 2020.
- [22] HAO WANG, JIAN LI, LINFENG WANG, LIN LIANG, ZHOUMO ZENG, AND YANG LIU. ON ACOUSTIC FIELDS OF COMPLEX SCATTERS BASED ON PHYSICS-INFORMED NEURAL NETWORKS. *Ultrasonics*, 128:106872, 2023.
- [23] SHOICHI KOYAMA, JULIANO G. C. RIBEIRO, TOMOHIKO NAKAMURA, NATSUKI UENO, AND MIRCO PEZZOLI. PHYSICS-INFORMED MACHINE LEARNING FOR SOUND FIELD ESTIMATION: FUNDAMENTALS, STATE OF THE ART, AND CHALLENGES. *IEEE Signal Processing Magazine*, 41(6):60–71, 2024.
- [24] MARCO OLIVIERI, XENOFON KARAKONSTANTIS, MIRCO PEZZOLI, FABIO ANTONACCI, AUGUSTO SARTI, AND EFREN FERNANDEZ-GRANDE. PHYSICS-INFORMED NEURAL NETWORK FOR VOLUMETRIC SOUND FIELD RECONSTRUCTION OF SPEECH SIGNALS. *EURASIP Journal on Audio, Speech, and Music Processing*, 2024(42):1–14, 2024.
- [25] JULIANO G. C. RIBEIRO, SHOICHI KOYAMA, RYOSUKE HORIUCHI, AND HIROSHI SARUWATARI. SOUND FIELD ESTIMATION BASED ON PHYSICS-CONSTRAINED KERNEL INTERPOLATION ADAPTED TO ENVIRONMENT. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 32:4369–4382, 2024.
- [26] MIRCO PEZZOLI, FABIO ANTONACCI, AND AUGUSTO SARTI. IMPLICIT NEURAL REPRESENTATION WITH PHYSICS-INFORMED NEURAL NETWORKS FOR THE RECONSTRUCTION OF THE EARLY PART OF ROOM IMPULSE RESPONSES. IN *10th Convention of the European Acoustics Association (Forum Acusticum)*, TURIN, ITALY, SEPTEMBER 2023.
- [27] ZITONG LAN, CHENHAO ZHENG, ZHIWEI ZHENG, AND MINGMIN ZHAO. ACOUSTIC VOLUME RENDERING FOR NEURAL IMPULSE RESPONSE FIELDS. IN *Advances in Neural Information Processing Systems (NeurIPS)*, 2024.
- [28] ANTON RATNARAJAH, ZHENYU TANG, ROHITH ARALIKATTI, AND DINESH MANOCHA. MESH2IR: NEURAL ACOUSTIC IMPULSE RESPONSE GENERATOR FOR COMPLEX 3D SCENES. IN *Proceedings of the 30th ACM International Conference on Multimedia*, MM ’22, PAGE 924–933, NEW YORK, NY, USA, 2022. ASSOCIATION FOR COMPUTING MACHINERY.

- [29] ZIQI FAN, VIBHAV VINEET, HANNES GAMPER, AND NIKUNJ RAGHUVANSHI. FAST ACOUSTIC SCATTERING USING CONVOLUTIONAL NEURAL NETWORKS. IN *ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, PAGES 171–175, 2020.
- [30] PHILLIP ISOLA, JUN-YAN ZHU, TINGHUI ZHOU, AND ALEXEI A EFROS. IMAGE-TO-IMAGE TRANSLATION WITH CONDITIONAL ADVERSARIAL NETWORKS. *Proc. CVPR*, 2017.