

深層学習による室形状情報を考慮した 室内音場シミュレーションの高速化

佐藤 元[†] 池田 雄介[†]

[†]東京電機大

1 はじめに

室内音場シミュレーションは、室内の音伝搬情報の把握に重要であり、室内音響設計やゲーム音響などに用いられる。室内音場シミュレーションには境界要素法 (BEM) や FDTD 法といった波動方程式に基づいた数値解析が用いられるが、解析範囲が広がるほど、計算時間が必要となるという問題がある。そのため、これまでに計算時間の短縮を目的とした研究が進められてきている [1]。近年、音場シミュレーションの推定結果を画像として教師データに用いて深層学習モデルを作成し、室内音場推定を行う手法が提案されている [2]。しかし、部屋の形状に自由度がないため、室内の形や大きさを変更して音場推定を行う場合、モデルの再学習を行う必要がある。

そこで、本研究では、室形状と音場シミュレーションデータを学習に用いることで、任意の室形状に対して室内音場推定を行う深層学習モデルを提案する。実験では 2 次元の室内音場を推定した結果を示す。

2 提案手法

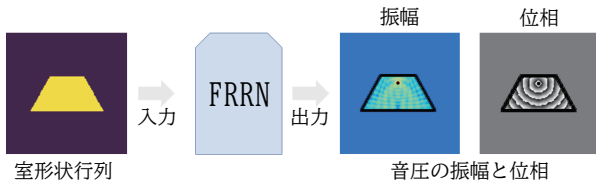


図 1: 提案手法の概要

提案手法の概要を図 1 に示す。提案手法では、室形状を表す行列を入力に音場推定を行う。音場推定を行うニューラルネットワークモデルには Full-Resolution Residual Network (FRRN) を用いる [3]。また、教師データには BEM で音場シミュレーションした結果を用いる。

2.1 データセットの作成

室形状行列と BEM による音場シミュレーション結果を用いてペアデータを作成する。室形状には、円や

ひし形など様々な形状を想定することができるが、本稿では、建築で一般的に用いられる単純な形状に着目し、台形に限定して学習データを作成する。台形は 12.8×12.8 m の 2 次元空間内に作成され、台形の上底、下底、高さはそれぞれ 3~9 m の範囲で値を持つ。台形 1 辺の間隔は 0.2 m とする。図 1 の入力、今回作成した室内形状行列の一例である。

音場シミュレーションデータには、BEM を用いて求めた複素音圧分布から振幅と位相を求め、2 チャネルの行列にしたものを用いる。また、openacoustics[4] を利用して BEM の計算を行う。表 1 に BEM の計算条件を示す。

表 1: BEM の条件

音速 [m/s]	340
周波数 [Hz]	500
音源座標	(0,1)
アドミタンス比	0.5
境界の間隔 [m]	0.1

2.2 FRRN

FRRN は、Tobias らの提案した画像のセグメント分けを行う畳み込みネットワークを用いた深層学習モデルであり、画像の低次元な特徴と高次元な特徴を同時に学習することを特色としている [3]。FRRN を用いることで音場の全体的な遷移と、局所的な遷移を学習することができる。と考える。

FRRN は、full-resolution residual unit (FRRU) という学習アルゴリズムを有している。FRRU では、解像度を変えずに残差を計算することで低次元の特徴を学習する residual stream と pooling によって解像度を低くしていくことで、高次元の特徴を学習する pooling stream の 2 つの学習を同時に行う。

3 実験

室形状行列から室内音場の振幅と位相を求める FRRN の学習を行い、推定結果の評価を行う。FRRN は、文献 [3] の FRRN A の 1 層目を畳み込みに変更したネットワークを使用する。また、入力と出力をそれぞれ 1 チャネルと 2 チャネルの 128×128 の行列とした。損失関数には平均二乗誤差を用いる。振幅については、データの正規化を行う。バッチサイズは 8 とし、

Fast sound field simulation considering room shape based on deep learning

[†] Gen Sato (22fmi21@ms.dendai.ac.jp)

[†] Yusuke Ikeda (yusuke.ikeda@mail.dendai.ac.jp)

Tokyo Denki University (†)

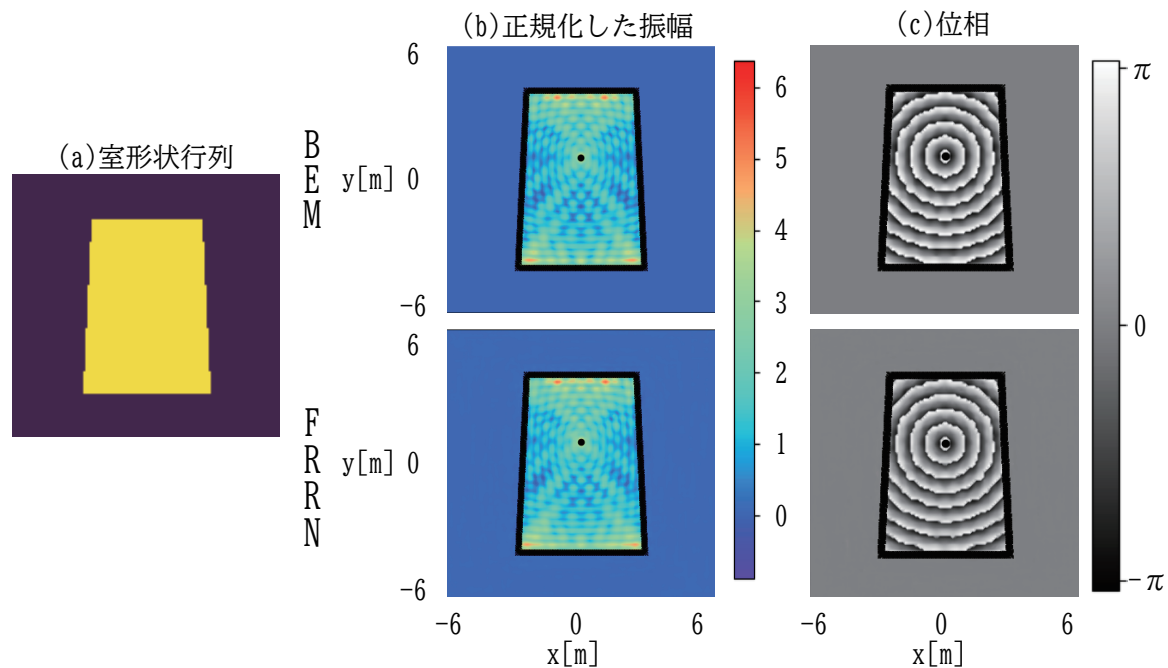


図 2: 入力室内形状と BEM (Ground Truth) と推定した振幅, 位相の比較

optimizer には学習率を 1.0×10^{-4} , モーメンタムを 0.99, 重み減衰を 5.0×10^{-4} に設定した stochastic-gradient-descent を用いる. 10000 サンプルのデータセットのうち, 学習データとして 9800, 検証データとして 100, テストデータとして 100 サンプルを利用する. イテレーション数は 122500 とする.

4 結果

図 2 は上底 5.2 m, 下底 6.0 m, 高さ 8.4 m の台形を推定し, 可視化した結果である. 座標 (0, 1) の黒点が音源位置を示し, 黒線が壁面の位置を示している. 振幅は正規化した値を可視化しているため相対的な表現となっている. BEM と FRRN の振幅と位相の推定結果を比較すると, 概ね近い音場が推定可能であることが分かる. また, 振幅における root mean squared error (RMSE) の平均は 0.200, 分散は 0.001 であった. 位相の RMSE は, 平均は 0.514 で分散は 0.006 となった. 振幅と位相の RMSE の分散より, 形状に左右されずに推定を行えることが分かった. 一方で, 音場の可視化を目的とする場合には, 十分な推定精度であると考えられるが, 可聴化の目的では, より高精度な推定が必要となると考えられる.

また, ワークステーション PC (AMD EPYC 7543P CPU, NVIDIA RTX A6000 GPU, 256GB DDR4 RAM) で図 2 の台形を BEM でシミュレーションする場合約 480 s かかった. しかし, 提案手法では約 1.7 s で推定が可能のため, 計算時間を約 280 倍短縮でき

ることが分かった.

5 おわりに

本稿では, 自由な室内形状に対して室内音場推定を行う深層学習モデルの作成を目指し, 室形状行列を入力に, BEM で推定した複素音圧の振幅と位相のを教師データとして用いて, FRRN の学習を行った. 作成したモデルの推定結果から, 可視化には十分な精度で, 振幅と位相を求められることが分かった. 今後は, 可聴化を目的として, より高精度な推定を検討する.

謝辞 本研究は JSPS 科研費 22K12099 の助成を受けたものです.

参考文献

- [1] Seiya Kishimoto *et al.*, “Analysis of Instantaneous Acoustic Fields Using Fast Inverse Laplace Transform,” *IEICE Transactions on Electronics*, 2022, Volume E105.C, Issue 11, Pages 700-703.
- [2] Ziqi Fan, *et al.*, “Fast Acoustic Scattering Using Convolutional Neural Networks,” *ICASSP 2020 – 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2020, pp. 171-175.
- [3] Tobias Pohlen *et al.*, “Full-Resolution Residual Networks for Semantic Segmentation in Street Scenes,” *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 3309-3318.
- [4] openacoustics, <http://www.openacoustics.org/>, 最終閲覧日 2023/01