4Y-01

円形マイクロホンを用いたend-to-end学習による音場再現手法

川瀬 敬子 † 佐藤 元 † 津國 和泉 † 池田 雄介 †
† 東京電機大

1 はじめに

複数のスピーカを用いて所望音場を物理的に再構成する目的で数多くの音場再現技術が提案されている．近年，深層学習技術の発展に伴い，計算量の低減と高いロバスト性の実現を目的として，end-to-end 学習による音場再現手法が提案されてきている [1]．しかし，音場の集音は制御点と同形状の平面マイクロホンアレイで行われてきた．そこで，本研究では集音時のマイクロホン形状を構成がより容易な円形マイクロホンアレイを用いた end-to-end 学習によるスピーカ駆動関数決定手法を提案する．

2 Pressure Matching

音場再現の手法の一つとして知られる Pressure Matching（PM）法 [2] では，再現領域を離散化した制御点における音圧を所望の音圧と一致させるようスピーカの駆動関数を決定する．スピーカアレイの内側の音場は各スピーカからの伝達関数の重ね合わせで表現される．したがって周波数領域上において，n 番目の制御点における音圧 $p(\mathbf{r}_n)$ は，l 番目のスピーカの駆動関数 w_l と，l 番目のスピーカと n 番目の制御点の間の伝達関数 $G(\mathbf{r}_n|\mathbf{r}_l)$ を用いると以下のように表される．

$$p(\mathbf{r}_n) = \sum_{l=1}^{L} w_l G(\mathbf{r}_n|\mathbf{r}_l) \tag{1}$$

ここで $\mathbf{r}_n, \mathbf{r}_l$ はそれぞれ制御点位置とスピーカ位置を表す．式 (1) は N 点すべての制御点で成立するため，伝達関数の行列 $\mathbf{G}(\in \mathbb{C}^{N\times L})$ と駆動関数ベクトル $\mathbf{w}(\in \mathbb{C}^{L\times 1})$ によって以下のように表される．

$$\mathbf{p} = \mathbf{G}\mathbf{w} \tag{2}$$

また，スピーカの駆動関数 $\mathbf{w}$ は正則化最小二乗法によって求まる．

3 提案手法

Xi Hong らは，PM 法を end-to-end 学習で行う手法を提案している [1]．提案手法では，それらを応用し，マイクロホンアレイを制御点配置と同形状ではなく，円形マイクロホンアレイで収録した所望音場の信号を入力として，スピーカの駆動関数を出力する end-to-end のモデルを提案する．図 1 に学習の流れを示す．合成の目標となる所望音場における M 個のマイクロホン信号の実部と虚部をベクトル化した p をモデル $\mathcal{U}$ への入力とする．出力は，L 個のスピーカへの駆動関数の実部と虚部をベクトル化した w_{DNN} とする．スピーカから制御点までの伝達関数 G と推定された l 番目のスピーカの駆動関数 $w_{\mathrm{DNN},l}$ と制御点における合成音圧 $p_{\mathrm{DNN}}(\mathrm{r_n})$ と所望音圧 $p(\mathrm{r_n})$ の誤差を用いて，損失関数は次のように表す．

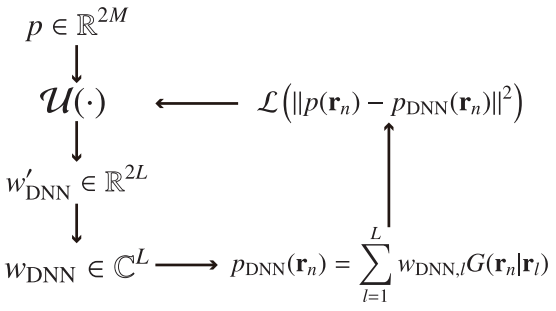

図 1: 学習の流れ

$$\|p(\mathbf{r}_n) - p_{\mathrm{DNN}}(\mathbf{r}_n)\|^2 \tag{3}$$

ただし，式 (1) より，合成音圧は

$$\|p(\mathbf{r}_n) - \sum_{l=1}^{L} w_{\mathrm{DNN},l} G(\mathbf{r}_n|\mathbf{r}_l)\|^2 \tag{4}$$

と表される．

4 シミュレーション実験

4.1 実験条件

シミュレーションデータを用いて提案システムの合成精度の評価を行った．実験条件は，表 1 に示す．また，マイクロホン，制御点，スピーカの配置を図 2(a) に示す．今回は一次音源である単一の点音源をランダムな位置に配置した音場のデータについて学習を行った．

データセット数は 1800 個，データセットにおける学習と検証の比率は 8 : 2 とした．バッチサイズは 8，エポック数は 400 とした．学習率は 0.0001 として optimizer には Adam を使用した．また伝達関数 G については既知のものとしてシミュレーション上で計算したものを使用した．ネットワークの構成は図 3 の通り

Sound field reproduction based on end-to-end learning with circular microphone array

† Keiko Kawase (20fi030@ms.dendai.ac.jp)
† Gen Sato (22fmi21@ms.dendai.ac.jp)
† Izumi Tsunokuni (21udc02@ms.dendai.ac.jp)
† Yusuke Ikeda (yusuke.ikeda@mail.dendai.ac.jp)
Tokyo Denki University (†)

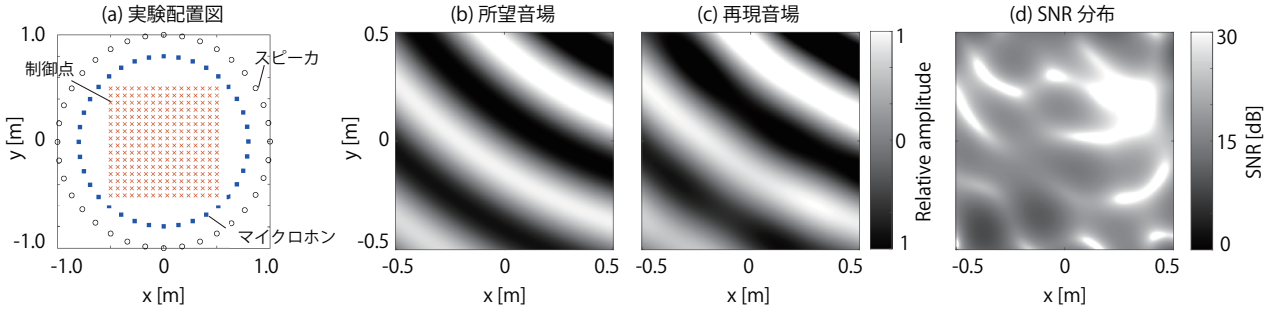

図 2: 実験配置図と実験結果

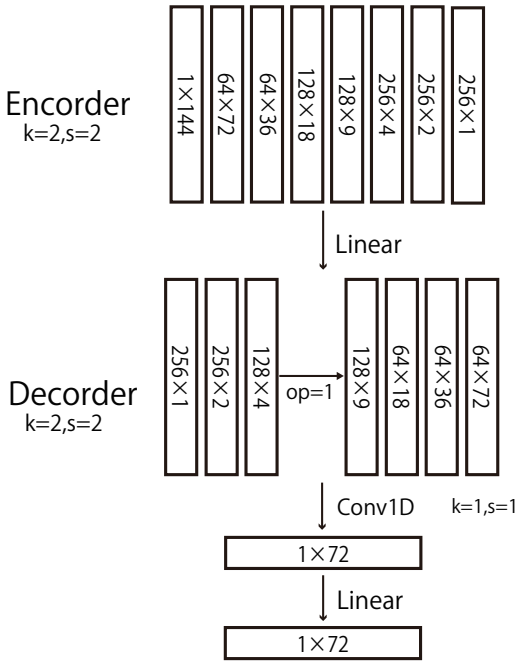

図 3: ネットワーク概要図（k=kernel size, s=stride, op=output padding）

である．この U-net を元にしたネットワークの層には主に一次元の畳み込みニューラルネットワーク（CNN）を使用した．また，以下の信号対雑音比 (SNR) を用いて定量的な評価を行った．

$$\mathrm{SNR} = 10\log_{10}\frac{|p_{\mathrm{des}}|^2}{|p_{\mathrm{des}} - p_{\mathrm{rep}}|^2} \tag{5}$$

ここで，p_{des} と p_{rep} はそれぞれ所望音場と再現音場の複素音圧である．

4.2 実験結果

所望音場を位置 (1.2, 1.8) に一次音源がある音場とし，提案手法で推定した駆動関数を用いて音場を再現する．図 2(b),(c) に所望音場と提案手法による再現音場を示し，図 2(d) に制御領域における SNR の分布図を示す．

図 2(b),(c) より，音波の到来方向の再現と振幅の再現が実現したことが分かる．また，提案手法の SNR 平均値は約 18.3 dB であった。したがって，提案手法は十分に高い精度で再現可能であることが分かる．以上のことから，従来の PM 法 [2] では，制御点とマイクロホンの配置が一致する必要があるが，提案手法ではマイクロホンと制御点の位置を自由に構成することが可能となった．

表 1: 実験条件

一次音源の位置 [m]	1.5 ~ 3.5
音源の周波数 [Hz]	750
スピーカ数	36
スピーカアレイの半径 [m]	1
制御点数	16×16
制御点の間隔 [m]	x:−0.5 ~ 0.5
	y:−0.5 ~ 0.5
制御点の範囲 [m]	0.067
マイクロホン数	36
マイクロホンアレイの半径 [m]	0.8

5 おわりに

本研究では，円形マイクロホンアレイを用いた end-to-end 学習によるスピーカ駆動関数決定手法を提案した．今後は，広帯域周波数への適用や，スピーカからの制御点への伝達関数のモデル化，実測による評価を行う．

謝辞 本研究の一部は東京電機大学総合研究所研究 Q22J-02 の助成を受けたものです．

参考文献

[1] Xi Hong, Bokai Du, Shuang Yang, Menghui Lei, Xiangyang Zeng, “End-to-end sound field reproduction based on deep learning,” J. Acoust. Soc. Am. , Vol. 153, pp.3055 , 2023.

[2] Ole Kirkeby, Philip A. Nelson, “Reproduction of plane wave sound fields,” J. Acoust. Soc. Am., vol. 94, no. 10, pp.2992–3000, 1993.