euspen's 25th International Conference & Exhibition, Zaragoza, ES, June 2025
www.euspen.eu

Evaluation of Reinforcement Learning Approaches for Phase Control in Mid-air Ultrasonic Haptics

Mifuka Nakamura[1], Rin Furuya[1], Nobuya Sato[1], and Daisuke Mizushima[1]

[1]Aichi Institute of Technology, Toyota, JAPAN

v24723vv@aitech.ac.jp

Abstract
Mid-air ultrasonic haptics can provide tactile sensations without touch by applying acoustic radiation pressure. In general, ultrasonic haptic focusing is achieved by using cameras for phase control. However, in this study, we aimed to develop a focus without external devices using reinforcement learning. In previous studies, the phase was treated as a continuous value. However, because of the large amount of learning required, the focus did not converge. In this study, we attempted to form a focus by treating the phase difference as a discrete value. Furthermore, by narrowing the search range and reducing the action selection space, we efficiently learned to achieve a higher sound pressure, allowing us to create a focus that exceeds the target sound pressure.

Actuator, Artificial intelligence, Haptic device, Ultrasonic

1. Introduction

Haptic technology is an area of tactile sensation. Mid-air ultrasound haptics can provide tactile sensations without touch by applying the acoustic radiation pressure of focused ultrasonic waves. This feature can be employed to reveal tactile principles by observing the skin during tactile presentation [1, 2].

To create a strong tactile stimulation at a desired position, the phase control of multiple ultrasound waves using beamforming technology. In general, mid-air ultrasound actuators use hand-tracking technology with a camera to calculate the phase from the distance. In contrast, this study used reinforcement learning to define the phases of waves. Previous studies using reinforcement learning have employed a policy-based reinforcement learning algorithm called Proximal Policy Optimization (PPO). However, the results indicated that the policy could not discover a unique focus that maximized accumulated rewards [3]. In this study, to obtain a unique focus, we use a Deep Q-Network (DQN), a representative value-based reinforcement learning algorithm for discrete state spaces.

2. Reinforcement Learning

Reinforcement learning is a framework in which a decision-making entity, called an agent, learns behaviors through trial and error in a dynamic environment. The agent selects an action a in a given state s, and the desirability of that action in the environment is quantified by a metric called a reward R. The value of taking action a in state s is represented as the state-action value $Q(s,a)$, and a tabular listing of these values is referred to as the Q-table. In reinforcement learning, the goal is to learn policies that consistently yield high $Q(s,a)$ values. In this study, actions correspond to phase differences between speakers, and states represent the sound pressure levels received by a microphone.

A DQN is an algorithm that combines Q-learning with deep learning. The update equation for the Q-value in Q-learning is given by Equation (1).

$$Q(s_t,a_t) \leftarrow Q(s_t,a_t) + \alpha \left[R_{t+1} + \gamma \max_{a_{t+1}} Q(s_{t+1},a_{t+1}) - Q(s_t,a_t) \right] \quad (1)$$

In Q-learning, the updated of the Q-table involves selecting the action with the highest Q-value among all possible actions in the current state. However, this requires calculating the Q-value for all potential actions, which becomes computationally intensive. DQN approximates these Q-value calculations using deep learning. For actual decision-making, the algorithm follows the ϵ-greedy method to determine whether to select the action with the highest Q-value or to continue exploration.

3. Experimental setup

The Eight ultrasonic speakers were arranged concentrically with a radius of 13.9 mm as actuators. The SU1007 ultrasonic speaker from SPL (resonant frequency: 40 kHz) was used. The control system utilizes a Mini PC (Raspberry Pi 4B) to perform reinforcement learning and transmits the phase information to the Arduino Uno R3. Arduino writes the obtained phase information to the waveform generation IC AD9833. The signal output from the waveform generation IC was amplified through an amplifier circuit to a maximum voltage of 60 V and applied to the speakers, emitting ultrasonic waves.

The actuators and microphone were positioned face-to-face, with the center of the actuator circle defined as the origin in a three-dimensional coordinate system (x, y, z). The microphone was placed at the desired point within the coordinate system. The voltage obtained from the microphone was treated as a reward in the reinforcement learning. In this study, one speaker was selected as the reference speaker, and beamforming was performed by individually controlling the waves of the other speakers relative to those of the reference speaker. Specifically, the phase of the control wave is assigned and synchronized with the fundamental wave to optimize the interference, thereby achieving effective sound pressure enhancement. Furthermore, to conduct reinforcement learning more efficiently, a method to narrow the search range for the ultrasonic phase was proposed. An outline of this method is shown in Figure 3. The details of this approach are as follows.

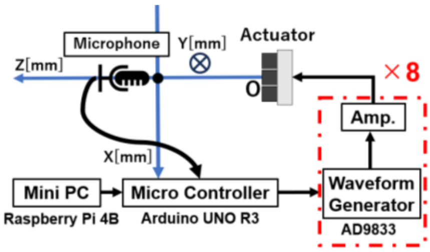

Figure 1. Experimental setup.

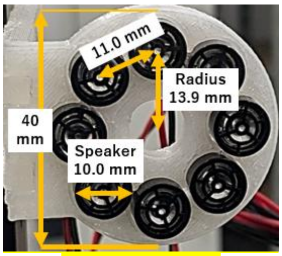

Figure 2. Actuator

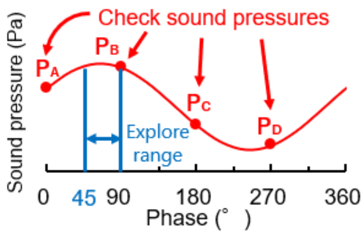

Figure 3. Diagram of directional localization

I. The phases of the ultrasonic wave were set to 0°, 90°, 180°, and 270°, and the sound pressures P_A, P_B, P_C, P_D were measured for each phase.

II. The total sound pressures $P_A + P_B, P_B + P_C, P_C + P_D$, and $P_D + P_A$ were calculated, and the combination with the maximum sound pressure was identified (e.g., P_A and P_B in Figure 3).

III. Among the combinations that resulted in the maximum sound pressure, the phase difference with the higher sound pressure was selected (e.g., P_B in Figure 3).

IV. The selected phase difference and the phase difference of the other combination are explored at 45° range.

This method efficiently narrows the exploration range, while determining the optimal phase difference.

4. Results and discussion

The DQN was implemented on the Mini PC, and the sound pressure was measured at the focal point. Prior to implementation, a simulation was conducted to optimize the learning and decay rates of. The focal point was set as (x, y, z) = (0, 0, 30). The total sound pressure when each speaker was operated individually was set as the target sound pressure. The target sound pressure at this focus point was 4839 Pa. Figure 4 shows the sound pressure distribution when learning was performed over the entire range from 0° to 360°. Overall, the sound pressure was low and the focus was not clearly defined. Calculating the focal diameter from the half-value width yielded a vertical size of 20 mm and horizontal size of 8 mm. The theoretical half-value width was 9.28 mm [4], indicating that was very broad and not sufficiently formed. Additionally, the sound pressure was 3196 Pa, which was significantly smaller than the target sound pressure. This result can be attributed to the extensive exploration range during training, which impedes adequate learning progress.

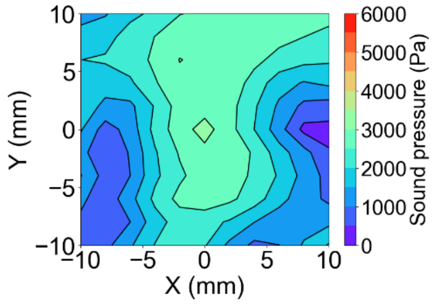

Figure 4. Sound pressure distribution without directional localization.

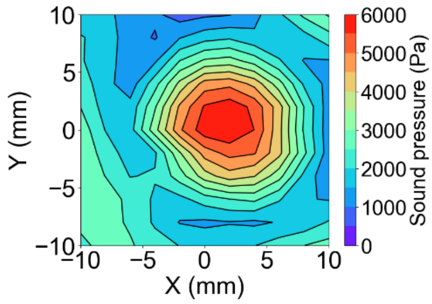

Figure 5. Sound pressure distribution with directional localization.

The sound pressure distribution when using Directional Localization is shown in Figure 4. Calculating the half-value width yields a vertical size of 12 mm and a horizontal size of 12 mm, which are very close to the theoretical values. Furthermore, the sound pressure was 5945 Pa, which exceeded the target sound pressure. This improvement is attributed to the narrowing of the exploration range, which allows the system to focus on specific phase differences that result in stronger sound pressure, leading to more efficient learning.

5. Conclusion

In this study, the phase differences of the sound sources were assumed to be discrete states, ranging from 0° to 360°. The optimal phase difference is learned using the value-based reinforcement learning DQN algorithm. Consequently, the sound pressure at the focus center was only 66% of the target sound pressure. In addition, the half-value width was significantly larger than the theoretical value. It seems that the large exploration range during learning prevented sufficient learning. Therefore, we devised a method to narrow the exploration range and efficiently learn higher sound pressures. The sound pressure distribution was evaluated when the learning range was restricted using Directional Localization. The central sound pressure of the focus exceeded the target, and the half-value width approached the theoretical value, indicating accurate focus formation.

In the future, speakers will be equipped with the ability to receive waves reflected from objects. The sound intensity of the reflected waves was learned using reinforcement learning, enabling ultrasonic focusing on the target location without the need for external equipment.

References

[1] Frier W, Abdouni A, Pittera D, Georgiou O, and Malkin R 2022 Simulating airborne ultrasonic vibrations in human skin for haptic applications *IEEE Access* **10** 15443-15456

[2] Mizushima D and Sato N 2024 Characteristics of Displacement Behavior of Dummy Skin Using Convergent Ultrasonic Haptic Actuators *Sensors and Materials* **36** 2669–2683

[3] Ryusei S, Kohei O 2023 *Proceedings of the 56th Academic Conference of the College of Industrial Technology, Nihon University* 638-639

[4] O'Brien W D Jr 2007 *Prog. Biophys. Mol. Biol.* **93** 212–255