Abstract. Individuals with severe disabilities typically suffer from lack of muscle power and deformation of their hands. This significantly affects their ability to operate an Electric Wheelchair (EW) especially when using joystick controllers because one operation affects both the straight and turning motions of an EW. Thus, individuals with severe disabilities typically exhibit difficulties in operating an EW, especially in cases that require fine steering capabilities. In contrast to autonomous driving, the study proposes a novel shared control method using reinforcement learning considering utilizing users' residual physical functions. With the use of online learning, the shared controller provides proper assistance based on the driving environments and each user's input characteristics. The effectiveness and characteristics of the novel shared controller are discussed via simulations.

Keywords: Shared control, Electric wheelchair, Individuals with severe disabilities, Reinforcement learning, Residual physical functions

1. Introduction

Recently, the number of individuals with severe disabilities is increasing in Japan [1]. Maintaining mobility is important to improve their quality of life and dignity [2, 3]. For individuals with severe disabilities, electric wheelchairs (EWs) are almost the only tool to maintain mobility. However, individuals with severe disabilities experience difficulties in operating an EW due to two main reasons. First, some individuals with disabilities do not exhibit proper input devices due to their lack of muscle power and deformation of their hands [4, 5]. Second, even with the special designed input devices, there is no guarantee that they can accurately operate the EW as intended and especially when the straight and turning motion of an EW should be...

Received: August 19, 2019; Accepted: October 10, 2020; Published: October 17, 2020
Automatic driving partly solves the problem although excessive reliance on automatic driving is not conducive to maintaining their residual physical functions.

Shared control technologies offer another possibility wherein a user and machine work together to operate a navigation system to the destination. The problem of shared control is typically defined as "If the machine function augment or replace one or more of the human senses, the user must be able to reach his goal and also understand what the machine is doing. This two-way interaction must exist in any shared control system." Many types of shared control systems were developed for the EWs. Essentially, they are divided into three categories based on how users participate in the driving process. The first category corresponds to behavior based shared control that is typically designed for those individuals with disabilities who still continue to face constant and considerable operating ability. The system only intervenes in a few previously defined situations such as avoiding an obstacle and following a wall. The key issue of the approach is how to execute the behavior. Many of the EWs are combined with methods including the dynamic window approach (DWA) or vector field histogram (VFH) when the defined situations are detected or triggered by users. The second approach corresponds to goal based shared control. The system first estimates a potential goal or subgoal from the input signal of the user. Subsequently, several possible paths are generated by global and local planners. The most likely driving trajectory is selected by considering the user’s input. It should be noted that although the user’s intention is considered in the method, the EWs simply select the prepared trajectories calculated by the planners, thereby significantly limiting the user’s control authority.

The third approach corresponds to a continuous shared control. Users’ performance is evaluated online via some predefined cost functions, and user control authority is then determined by online evaluation results. However, driving an EW corresponds to a complicated process wherein operation is different even if in the same condition. Real-time optimization limits users’ performance, and the construction of the cost functions is always a difficult problem.

The study proposed a novel shared control design methodology of the EWs for individuals with severe disabilities based on reinforcement learning. The control authority of the user is adjusted by considering the requirements and operating characteristics of the user. In contrast to the continuous shared control discussed above, the proposed shared control evaluates user performance over considering the entire driving process wherein the shared controller gradually adapts to the user’s operating characteristics after fully understanding a user’s entire driving process. The main purpose of the study is to provide a design methodology for this type of shared controller that completely utilizes a user’s operating ability and can be adapted to various environments.

Specifically, in Section 2, driving models of individuals with severe disabilities are built
2. Driving environments and models for individuals with severe disabilities

This section initially determines the driving environments of the study based on real living environments for individuals with severe disabilities. Subsequently, the operating characteristics of individuals with severe disabilities are analyzed, and driver models based on these characteristics are designed. The driver models are used to show the validity of the proposed shared control methodology in section 5 and 6.

2.1. Driving environments for individuals with severe disabilities

The selected environments must contain the features of lifestyles for individuals with severe disabilities. Ten individuals with severe disabilities in Shimoshizu National Hospital were investigated to understand their lifestyles. The results indicate that their lifestyle range is significantly limited, and thus most of them only move among their bedrooms, living rooms, and toilets. In order to complete the tasks with EWs, the EW should exhibit three basic movements, namely docking at an obstacle, going straight, and turning. In the study, as shown in Fig. 1, two situations, namely docking at an obstacle and going straight and then turning around a corner, are used to verify the effectiveness of the proposed shared control method.

(a). Docking (b). Going straight and then turning around a corner

Figure 1: Two basic situations for EW driving
2.2. Operating characteristics of individuals with severe disabilities

Residual finger functions are one of the most important criteria to assess the degrees of disabilities \([15]\). Additionally, individuals with severe disabilities prefer to operate input devices with hand functions as opposed to facial muscle functions \([3]\). Therefore, the study focuses on analyzing the input characteristics of individuals with severe disabilities with input examples using fingers. Figure 2 shows two examples as to how individuals with severe disabilities operate input devices that are designed based on their hand functions \([4, 5]\). Figure 2(a) shows an individual with severe disabilities operating a tiny joystick that is designed based on a quantitative evaluation of his residual hand functions and deformation of his hands. In order to understand his input characteristics, the input range is first measured. As shown in Fig. 2(b), the results indicate that the input range is significantly limited and especially for the right half. As shown in Fig. 2(c), the individual with severe disabilities is then asked to operate a tiny joystick to follow a moving target to understand dynamic properties of his input characteristics. The position of the moving target corresponds to \((-R \cos \omega t, 0)\), where \(\omega = 36 \,^\circ/s\). Figure 2(d) shows that clear differences exist between the input signal and target signal wherein the input signal is delayed or insufficient even in the left half (< 90 °). Figure 2(e) shows the previous value (before developing the tiny joystick) and the current value (after developing the tiny joystick) of the fingertip force that the user can give in X-direction. The results show that the fingertip force in the abduction direction declined by approximately 0.5 N during the development of the tiny joystick, which made this user difficult to operate the tiny joystick.

Figure 2(f) shows another situation where the individual has lost almost all the physical functions, and he can only move his thumbs up and down in a very small range. A one-dimensional input device (1DID) is developed wherein the input signal is adjusted by only detecting the extent to which the sensor is pushed. As shown in Fig. 2(f), the green bar from the input signal grows faster when users barely push the device. The user is asked to provide the input signal and subsequently stop the signal at the middle. Figure 2(g) shows one example that how occupational therapist helps this user to wear the 1DID with their rich experience. One sponge was used to support the arm and another sponge was used to reduce the distance between the thumb and other fingers.

Figure 2(h) shows one of the evaluation results, and the subject stops the signal near the target line after several practices albeit still with some deviation. It can be concluded that although 1DIDs are used to express an individual’s intention to a certain extent, it is almost impossible for the individual to operate the EWs without any assistance due to his/her extremely weak and unstable operating ability. In summary, the input characteristics of individuals with severe disabilities are summarized as follow:

- The input range is limited. This is mainly due to the following reasons. First, during the
development of the input devices, their physical functions continue to decline, and thus they are unable to properly operate the designed device after the device is developed. Second, a small disturbance, such as posture of their hands, also affects the input range due to the lack of muscle power and deformation of hands. Their input signals is always delayed or advanced when compared with the target reference. For individuals with an extremely weak operating ability, their input only intuitively represents the subjects’ intention although driving an EW in real environments corresponds to complicated behavior. More accurate and frequent input is necessary and especially for situations in which fine steering capabilities are required, and this is almost an impossible task for them.

2.3. Driving models for individuals with disabilities

In order to prove the effectiveness of the shared control method proposed in the study, two driver models are designed to describe the driving features of individuals with disabilities. The discussion in Section 2.2 indicates that individuals with disabilities experience difficulties in driving an EW and especially in situations where frequent input adjustment and precise control are required. For the docking situation, in our previous research, it is found that users usually begin to brake when the EW is closer than 30 cm to the target line [16].

- Figure 2: Operating examples of individuals with severe disabilities [4, 5]
individuals with severe disabilities are unable to provide the signal to stop the EW in time cause the EW to collide with the obstacle while users who provide a premature signal are unable to get sufficiently close to stop at the right position. (1) is used to describe the “stopping too early” and “stopping too late” situations.

\[ v = \begin{cases} v_0, & \Delta y > y_{brake} \\ v_0 - k(t - t_{brake}), & \Delta y < y_{brake}, v_0 > k(t - t_{brake}) \\ 0, & v_0 < k(t - t_{brake}) \end{cases} \]

Where \( v \) denotes the velocity of the EW, \( v_0 \) denotes a constant velocity of the EW before the EW starts to brake, \( \Delta y \) denotes the distance to the goal, \( y_{brake} \) denotes the distance of the EW from the goal when braking begins, \( k \) denotes the braking gain, and \( t_{brake} \) denotes the time to break, respectively.

As shown in Table 1, the “stopping too early” and “stopping too late” situations are reproduced by changing the parameters in (1), and the results are shown in Fig. 3(a-b).

With respect to situations involving going straight and turning, individuals with severe disabilities experience difficulties in providing accurate turning signals when they are asked to drive an EW to the goal at a certain speed. (2) shows an example of a 1st order driver model [17].

\[ U_\psi = \left( h_L * d_L - h_R * d_R \right) (\tau_p s + 1) / (\tau_d s + 1) \]

\( U_\psi \) denotes the user input for the yaw rate, \( h_L \) and \( h_R \) denote operating gain, \( \tau_p \) denotes preview time, \( \tau_d \) denotes operation delay, \( s \) denotes the Laplacian operator, and \( d_L \) and \( d_R \) denote the distances to the left and right wall, respectively. The input and driving characteristics are adjusted by changing the parameters in (2). Figure 3(c) shows the case where the user drives the EW with insufficient input, while Figure 3(d) shows that the user turns with an oversteered input. The parameters for each situation are also shown in Table 1.

| Parameters | Stop too early | Stop too late | Insufficient input | Oversteer input |
|------------|----------------|---------------|-------------------|----------------|
| \( v_0 \)  | 0.7 m/s        | 0.7 m/s       | -                 | -              |
| \( y_{brake} \) | 0.5 m        | 0.1 m        | -                 | -              |
| \( k \)     | 1             | 0.1           | -                 | -              |
| \( h_L \)   | -             | -             | -                 | -              |
| \( h_R \)   | -             | -             | -                 | -              |
| \( \tau_p \) | 3             | 3             | -                 | -              |
| \( \tau_d \) | -             | -             | -                 | -              |
3. Shared control system using reinforcement learning

This section mainly describes the design process of the shared control system based on the input characteristics and requirements of the individuals with severe disabilities. Specifically, the modeling of EW that is used in the shared control system is first introduced. Subsequently, the design requirements of the shared control system are summarized based on the process of EW driving and input characteristics of the individuals with severe disabilities. The design concept is then proposed based on the design requirements. Finally, the method to realize the design concept using reinforcement learning is introduced.

3.1. Modeling an EW

The EW shown in Figure 4 is driven by two rear wheels and two casters are used as the front wheels. The parameters are shown in Table 2. Given the assumption that the slip between the wheels and road is neglected, the straight and rotation motion of an EW can be expressed as follows:

\[ v = r(\omega_R + \omega_L)/2 \]

\[ \dot{\psi} = r(\omega_R - \omega_L)/W \]

Table 2: Parameters of an EW

| Parameters | Meaning |
|------------|---------|
| \( \omega_R, \omega_L \) | Angular velocity of the right (left) wheel |
| \( r \) | Radius of the rear wheel |
| \( W \) | Width of the wheelchair (0.5 m) |
| \( L \) | Length of the wheelchair (0.6 m) |
| \( v, \dot{\psi} \) | Velocity of the straight (yaw) motion |
| \( x, y \) | Position in world coordinate system |
| COR | Center of rotation |
3.2. Requirements of the shared control system

- The system assists users to arrive at the destination safely and comfortably.
- The system completely utilizes residual functions of individuals with disabilities.
- The system assists the user by considering the current moment and the complete driving process.
- The system adapts to various types of users and environments.

3.3. Concept of the shared control system

Based on the requirements of the shared control system, the concept of the novel shared control system is shown in Fig. 5. The relationship between user and machine is similar to the relationship between a driving coach and learner driver. The coach should intervene at an appropriate timing based on the performance of the learner driver. The coach should also be able to intervene in different ways based on different training purposes. For example, it is possible to appropriately decrease the intervention for individuals with an error correction capability.
3.4. Framework of the proposed shared control system

In order to satisfy the requirements of the shared control system, the proposed shared control system (coach) is divided into two parts, namely the planner and online learning parts, as shown in Fig. 6(a). Users provide a velocity order \((v_{\text{user}}, \dot{\psi}_{\text{user}})\) via some type of input device. Subsequently, user signals, planner velocity signals \((v_{\text{DWA}}, \dot{\psi}_{\text{DWA}})\), and environmental information (the relative position between the EW and the environment) are used in the online learning system, and the one (user or planner) who performs better obtains more control weights. The control weight for the user is defined as \(k_v\) and \(k_{\dot{\psi}}\) while the control weight for non-expert planner is \((1 - k_v, 1 - k_{\dot{\psi}})\). The online learning part is shown in Fig. 6(b). For individuals with severe disabilities, the operating characteristics for going straight and turning are significantly different, and thus \(k_v\) and \(k_{\dot{\psi}}\) are calculated separately. The equations of shared control are expressed as follows:

\[
\begin{align*}
\dot{v} &= k_v v_{\text{user}} + (1 - k_v)v_{\text{DWA}} \\
\dot{\psi} &= k_{\dot{\psi}} \dot{\psi}_{\text{user}} + (1 - k_{\dot{\psi}})\dot{\psi}_{\text{DWA}}
\end{align*}
\]

The planner in the system corresponds to a non-expert motion planner that exhibits two purposes. First, the EW reaches the goal with the planner although the user’s input is unreliable. Second, previous studies show that proper competition increases individual’s motivation \([18]\). Thus, the planner is also used as a competitor, and users obtain more control weight when they perform better than the non-expert planner and vice versa. In the study, the dynamic window approach (DWA) \([11]\) is selected as the non-expert planner. The purpose of the online learning part involves obtaining optimal user control weight by considering the requirements of the shared control system. The combination of the two parts is used as the coach that can intervene based on different training purposes. The design of the online learning part is discussed in the next Section.
4. Shared controller design using reinforcement learning

4.1. Reinforcement learning

Reinforcement learning is considered as a long-term optimization method that is suitable for this application. It exhibits four main elements, namely the policy, reward, value function, and model of environment. The policy is used to map states to actions. The reward $R$ and state-action value function $Q(S,a)$ (or state value function $V(S)$) are used to make decisions to adopt actions on each state. The goal of reinforcement learning is to determine the optimal policy from the reward and interactions with the environment. This section introduces designing the reinforcement learning algorithm to satisfy design requirements that are discussed in Section 3.

Reinforcement learning is generally based on the Markov decision process (MDP). Specifically, MDP is represented as $<S,A,T,R,\gamma>$, where $S$ denotes a finite set of states, $A$ denotes the finite set of actions, $T$ denotes a state transition probability matrix, $R$ denotes the reward, and $\gamma$ denotes the discount factor. As discussed in Section 3, the objective of the shared control involves determining optimal control weights (optimal policy) that maximize the accumulated reward from the current state of EW to the destination. Specifically, the optimal policy $\pi'(S)$ is represented in (6) via $Q^\pi(S,a)$:

$$\pi'(S) = \arg\max_{a \in A} Q^\pi(S,a)$$  \hspace{1cm} (6)
4.2. Design of discrete states

EW exhibits three-dimensional freedoms such that the states in MDP must contain the information of positions and directions of an EW. Therefore, the state of the human-EW system is defined as \( <x, y, \psi> \), where \( x \) and \( y \) represent position while \( \psi \) denotes the direction of the entire system. Additionally, driving an EW corresponds to a continuous process, and thus it is necessary to transform sequential data into discrete data to avoid dimensional explosion. When discretizing the states, if the density of the discrete states is excessive, then the convergence of the online learning is too slow. Conversely, if the density of the states is too sparse, then it does not accurately reflect the state of the EW. Hence, it is important to accurately reflect the states of the EW with less states. The sampling frequency of this shared control system is set as 10 Hz.

Given that the speed of an EW in indoor environments typically ranges between 1.5 km/h to 2.5 km/h, the EW moves approximately 4–7 cm within a cycle. The position information is presented using grid cells that correspond to squares of a certain length. We compared different learning results by using grid lengths corresponding to 10 cm, 20 cm, and 30 cm as discussed in the next section.

It should be noted here that the grid lengths here is just for pre-set user model, when the system is used by real people, it is difficult to find out exactly what the optimal value is, and only a rough range can be given.

The orientation of the EW \((\psi)\) is classified into the following three categories: correct direction, middle correct direction, and wrong direction. The method to define the categories of the direction is introduced in Section 4.3.

4.3. Design of the reward function

The reward should be defined based on the purpose of reinforcement learning. There are three main purposes of the application as discussed in Section 3. The first purpose is to drive the EW with driving indices such as comfort and safety. The second purpose is to use users’ residual physical functions as much as possible. The final purpose is to drive the EW to the final destination.

With respect to the first part, the reward is divided into the predicted reward and current reward. The current reward is determined by difference between the velocity of the current moment and a previous moment considering the index such as comfort [8] and time requirement. With respect to the situation of docking at an obstacle, the entire predicted and current reward are calculated with an inverse reinforcement learning method [16] based on the data from expert driver. The result is shown in Fig. 7 where each unit represents 0.1 m on the position classification axis while each unit represents 0.5 km/h on the velocity classification axis. The vertical axis denotes the reward for each state. It shows that users begin to decrease the speed when the distance is approximately 0.3 m from the obstacle. A Hybrid A* (HA*) [20]...
based method is designed to calculate the predicted reward when the EW is driven continuously. The HA* corresponds to a heuristic method that determines the cost from the current state to the goal state. As shown in Fig. 8, HA* method is initially used to calculate the cost of current state. Subsequently, the cost of the state for the next moment is calculated given the assumption that the entire input (user input with planner input) remains unchanged. The difference between the cost of the next state and current state is considered as the predicted reward. The items for cost function of this HA* are designed by considering the requirements and characteristics of individuals with severe disabilities. The items for cost function of this HA* are shown in Table 3, and the meaning of the parameters for each item is listed in Table 4. For example, the cost for the states $<x = 1.8 \text{ m}, y = 0.1: 2.4 \text{ m}, \psi = 15: 225^\circ >$ is shown in Fig. 9, the color denotes the cost for each state, and the cost increases when the color becomes lighter. The result is also easily considered as: the closer the state is to the goal state, the lower the cost of the state. In order to decrease the total amount of the states, we divide the direction $\psi$ into the correct direction, middle correct direction, and the wrong direction as opposed to simply classifying them by numerical values. In the study, the resolution of the direction classification is set as 30º. As shown in Fig. 9, with respect to the position $(x = 1.8 \text{ m}, y = 0.1 \text{ m})$, $\psi$ between 75–105º and 45–75º is considered as the correct direction and middle correct direction while $\psi$ in the other range is considered as the wrong direction. For the second part, the reward is used to evaluate whether the user's input is used to drive the EW. Therefore, this part is simply calculated by $P_2 * k_t - 1$ where $k_t - 1$ denotes the user control weight at the last moment and $P_2$ denotes the gain for this reward that is adjusted to the user control weight as much as possible. Because the main purpose of this paper is to discuss how to assist the user by considering their residual physical functions, the effects of $P_2$ for the shared control system will be discussed in Section 6.
With respect to the last part, a task reward is set to evaluate whether the user completes the task. The system obtains a positive reward $R_{\text{goal}}$ if the EW achieves the goal, the system obtains a negative reward $-R_{\text{collide}}$ if the EW collides, and the task reward is zero otherwise.

Figure 8 shows the entire reward function. Specifically, the predicted reward corresponds to the difference between the cost of the state for the next moment and current moment, and it represents the evaluation of the current behavior of the EW. The current cost is calculated by the change in EW speed because acceleration corresponds to the main factor that affects comfort.

Participation rewards are used to evaluate whether a user is actively involved in the driving process. Finally, the task reward aids the system in evaluating driving behavior from the complete driving process.

\[
\text{Reward} = \frac{\text{Cost}_{\text{current}} - \text{Cost}_{\text{next}}}{\text{Predicted reward}} + P_1 \cdot \text{abs}(v_t - v_{t-1}) + P_2 \cdot k_{t-1} + R_{\text{goal}} - R_{\text{collision}} \]

\[
= \frac{\text{Cost}_{\text{current}} - \text{Cost}_{\text{next}}}{\text{Predicted reward}} + P_1 \cdot \text{abs}(v_t - v_{t-1}) + P_2 \cdot k_{t-1} + R_{\text{goal}} - R_{\text{collision}} \]

\[
\text{Cost}_{\text{current}} \text{ using Hybrid A* method}
\]

\[
\text{Cost}_{\text{current}} \text{ using Hybrid A* method}
\]
4.4. Choice of the online reinforcement learning algorithm

Temporal difference (TD) methods are typically used to solve adaptive optimal control problems in reinforcement learning fields \[19\]. In TD methods, the value function is always updated at each time step. Thus, Q-learning is an off-policy TD method that is successfully used to solve many robot control problems \[21\]. In contrast to Q-learning, Sarsa-learning corresponds to an on-policy TD method wherein the action-value function \(Q(S,a)\) is estimated for the current policy and state-action pair \[19\]. The main difference between Sarsa-learning and Q-learning is that real observed data is used to revise the action value function in Sarsa-learning as opposed to updating the Q-values by using the maximum action value function for the next state.

Sarsa-learning is widely used in the field of dynamic planning in an unexperienced environment \[22-23\]. In the study, Sarsa-learning is applied to calculate the optimal policy that is used to select optimal user control weight in different states. The state sequence \(S_t, A_t, R_t\), action value function \(Q\), and optimal policy are designed in the aforementioned parts of the section. The step to update the action-value function is shown in (8)

\[
\alpha_t = \frac{e^{Q(A_t)/\tau}}{\sum_{i=1}^{n} e^{Q(A_i)/\tau}}
\]

4.5. Reinforcement learning based shared control algorithm

At each time step, the system obtains a set of data \(\{S_t, A_t, R_t\}\) after each learning episode. Subsequently, the action value function \(Q(s)\) is updated using (8). The complete algorithm is shown below.

Algorithm 1. \(t = 0, \gamma = 0.8, \alpha_t = 0.05, R_t \text{ from reward function}\)

Algorithm 2. Recursively compute until the goal is reached

Algorithm 2.1. Recursively compute until the stop condition is met
1. Obtain current state $S_t$

2. Decide action $A_t = (k_v, k_{\dot{\psi}})$ by Sarsa-learning

3. Send $(k_v, k_{\dot{\psi}})$ to the system, calculate output $(v, \dot{\psi})$

4. Send output $(v, \dot{\psi})$ to the EW

5. Calculate the next state $S_{t+1}$, Update $Q$ table, Load $R_t$

- $Q(s, a)$
- $Q(s, a)$
- $Q(s, a)$
- $Q(s, a)$
- $Q(s, a)$

5. Simulation studies

| Simulation ID | Environment  | Driver model |
|---------------|--------------|--------------|
| 1             | Turning left | Insufficient input |
| 2             | Docking at an obstacle | Early/delayed stop |
| 3             | Turning left | Oversteer input |

5.1. Designing the density of discrete states

- The stop condition of the algorithm is set as follows:
  - $Q(s, a)$ for each state only changes in a certain value.
  - Rank of $Q(s, a)$ for each state only changes in a certain value.
  - The latest trajectory does not include new states.
  - Consecutive successes of the above condition.
  - Number of training exceeds a certain number.

Simulation studies

| Simulation ID | Environment  | Driver model |
|---------------|--------------|--------------|
| 1             | Turning left | Insufficient input |
| 2             | Docking at an obstacle | Early/delayed stop |
| 3             | Turning left | Oversteer input |

Table 5: Simulation purpose and conditions
5.2. Verification of the effectiveness in various environments

The results for simulation 2 are shown in Fig. 10. Specifically, Figure 10(a) shows the results for users wherein input signal displays delay characteristics such that they cannot stop the EW in time. Figure 10(b) shows the results for users wherein input signal displays early characteristics that stop the EW too early. After several trails of training, the controller gradually grasps the characteristics wherein user’s input is insufficient or delayed when approaching the obstacle. The control authority is gradually adopted by the DWA planner to ensure safe docking at the obstacle.

The results for simulation 3 are shown in Figs. 11 and 12. Figure 11 shows the trajectory change during the training process. After a few trials of training, the EW gradually makes a safe left turn. Figure 12 shows the changes in $v$, $\dot{\psi}$, and $k$ during the training. The results indicate that the changes in $v$, $\dot{\psi}$, and $k$ decrease when the training increases. When compared with the trajectory change, the results indicate that more training is necessary even if the EW makes a safe left turn, and thus parameters (for e.g., $v$, $\dot{\psi}$, and $k$) are smoother, and the shared controller can be further adapted to the characteristics of the user and environment.

(a). Situation when stopping too late
(b). Situation when stopping too early

Figure 10: Trajectory change during the training process
6. Discussion

The study proposed a shared control system based on reinforcement learning that completely utilizes a user’s operating ability through training. Three problems should be discussed here.
The first problem involves discussing how the shared controller is adapted to different users and environments. The second problem involves understanding how parameters of reward function affect the training. The last problem involves how to use the shared system in real life.

As indicated by the simulation studies in Section 5, the proposed shared control method is proven as effective for various individuals with severe disabilities and environments. The DWA can be considered as a minimum guarantee wherein if the user’s input is not sufficient to operate the EW to the destination, the DWA obtains a greater control authority. Simultaneously, users are encouraged to operate better than the DWA to take more control weights.

By designing the parameters for each part of the reward function, the training process is controlled to a certain extent. Because the main purpose of this paper is to propose a novel shared controller that completely utilizes a user’s operating ability and can be adapted to various environments, the impact of $P_2$ which control the user participation is discussed here.

As shown in (7), if the proportion of participation reward $P_2$ increases, as shown in Fig. 1, then the system attempts to increase the user’s participation. However, user control weight $P_2$ changes more frequently, and it usually requires more trainings when compared with the result in Fig. 1. Thus, if $P_2$ is excessively high, then it causes overfitting problems that affect convergence.

With respect to how to use the proposed system in real life, the simulation results indicate that the controller gradually adapts to users’ operating characteristics via several trainings. Before the controller fully understands the user’s operating characteristics, DWA planner is also regarded as a competitor to motivate the users in the designing process in addition to having the assistance function. Therefore, the effect of the shared control system on user operations will be discussed in future experiments.

7. Conclusions and future work

The study proposed a novel online learning shared control approach to assist individuals with severe disabilities to drive EWs. Considering the target users are individuals with severe disabilities, this system is first used in the simple indoor environments such as docking, going straight and turning situations to verify the effectiveness and find out its characteristics. It is found that this system is effective in these simple situations considering utilizing users’ residual physical functions.

A future study is warranted and should discuss more situations such as avoiding a dynamic object. Subsequently, the proposed method should be applied in a real EW system in the future. In the present study, the proposed system is first applied to a Virtual Reality (VR) based driving simulator that allows individuals with severe disabilities to train their skills and adapts the controller to users’ characteristics. The trained controller is then used for real EW driving.
References

[1] Annual report on government measures for persons with disabilities. Ministry of Health, Labour and Welfare, 2012, 114-119.

[2] M. Liu, K. Mieno, K. Hanayama, T. Fujiwara, N. Chino: Practical problems and management of seating through the clinical stages of Duchenne's muscular dystrophy. Arch Phys Med Rehabil. 84:6 (2003), 818-824.

[3] Easy-to-use switch controller design for Duchenne Muscular Dystrophy, Ministry of Health, Labour and Welfare, 2005. (in Japanese)

[4] M. Shino, Y. Yamamotoa, T. Mikatab, T. Inouec: Input device for electric wheelchairs considering physical functions of persons with severe Duchenne Muscular Dystrophy, Technology and Disability, 26(2014), 105-115.

[5] L. Xi, Y. Yamamotoa, M. Shino: One Dimensional Input Device of Electric Wheelchair for Persons with Severe Duchenne Muscular Dystrophy, AAATE Conference, 2017, 770-777.

[6] D. Vanhooydonck, E. Demeester, M. Nuttin, H. V. Brussel: Shared Control for Intelligent Wheelchairs: an Implicit Estimation of the User Intention, 1st International Workshop on Advances in Service Robotics, 2003, 176-182.

[7] P. Aigner, B. McCarragher: Shared Control framework Applied to a Robotic Aid for the Blind, IEEE Control System, 1991, 40–46.

[8] R. A. Cooper: Intelligent Control of Power Wheelchair, IEEE Engineering in Medicine and Biology, 1995, 423–431.

[9] A. Poncela, C. Urdiales, E. J. Perez, F. Sandoval: A New Efficiency-Weighted Strategy,
H. A. Yanco: Shared User-Computer Control of a Robotic Wheelchair System. Ph.D. Thesis, Massachusetts Institute of Technology, 2000, 64-81.

D. Foxey, W. Burgardy, S. Thrun: The Dynamic Window Approach to Collision Avoidance, IEEE Robotics and Automation Society, 4:1 (1997), 1-23.

R. C. Simpson, S. P. Levine, D. A. Bell, L. A. Jaros, Y. Koren, J. Borenstein: NavChair: an assistive wheelchair navigation system with automatic adaptation, Assistive Technology and Artificial Intelligence, 1998, 235-255.

E. Demeester, A. Hntemann, D. Vanhooydonck, G. Vanacker, H. V. Brussel, M. Nuttin: User-adapted plan recognition and user-adapted shared control: A Bayesian approach to semi-autonomous wheelchair driving, Autonomous Robots, 2008, 24:2 (2008), 193-211.

Q. Li, W. Chen, J. Wang: Dynamic shared control for human-wheelchair cooperation, IEEE International Conference on Robotics and Automation, 2011, 4278-4283.

Handbook of Physical Disabled persons, Nagano Rehabilitation Center, 2016, 1-3.

L. Xi, M. Shino: Online Learning Shared Control System of Electric Wheelchair for Persons with Severe Disabilities, 11th International Workshop on Human Friendly Robotics (HFR), 2018, 60-65.

K. Masaichi: On Fundamental Relations Existing between Steering Behavior Model including Prediction, JSAE Transaction, 5(1958), 40-43. (in Japanese)

E. L. Deci, G. Betley, J. Kahle, L. Abrams, J. Porac: When Trying To Win: Competition and Intrinsic Motivation, Personality and Social Psychology Bulletin, 1981, 79-83.

R. S. Sutton, A. G. Barto: Reinforcement learning: an introduction, The MIT press, 1998.

N. D. Richards, M. Sharma, D. G. Ward: A Hybrid A*/Automaton Approach to Online Path Planning with Obstacle Avoidance, AIAA 1st Intelligent Systems Technical Conference, 2004, 1-17.

P. Stingu, F. Lewis: Adaptive Dynamic Programming Applied To a 6DOF Quadrotor. Computational modeling and simulation of intellect: Current state and future perspectives. IGI Global, 2011, 102-130.

W. Xu, J. Huang, Y. Wang, C. Tao, L. Cheng: Reinforcement learning-based shared control for walking-aid robot and its experimental verification. Advanced Robotics, 2015, 29:22(2015), 1463-1481.

D. Ramachandran, R. Gupta: Smoothed Sarsa: Reinforcement Learning for Robot Delivery Tasks, 2009 IEEE International Conference on Robotics and Automation, 2009, 2125-2132.