Proceedings of the 2025 SICE Festival with Annual Conference
September 9-12, 2025, Chiang Mai, Thailand

Adaptive Difficulty Setting using GPDM-Based Skill Model for Dynamic Difficulty Adjustment in VR Exercise

Yusuke Goutsu[1†] and Tetsunari Inamura[1]

[1]Brain Science Institute, Tamagawa University, Tokyo, Japan
(Tel: +81-42-739-8609; E-mail: goutsu, inamura@lab.tamagawa.ac.jp)

Abstract: Dynamic difficulty adjustment (DDA) systems provide users with an optimal level of challenge. Previously, some studies developed a DDA system that simply adjusts the difficulty up or down according to the number of successes, only balancing the difficulty to a success rate of 50%. However, setting the difficulty only to a moderate level of 50% is insufficient since the appropriate difficulty varies depending on the individual and the situation. It is preferable to set stepwise difficulty levels according to the user's skill to adjust the difficulty between them and to evaluate the difficulty setting while considering the psychological aspects of the user. For this purpose, we propose an adaptive difficulty setting that consists of stepwise difficulty levels (e.g., $hard$, $normal$, and $easy$) adaptive for each user and evaluate it using self-efficacy with successful experiences. In the experiment, we employ a $Kendama$ game in a VR space where difficulty can be easily adjusted to objectively and subjectively evaluate the adaptive difficulty setting with one in which difficulty levels are uniformly fixed for all users. The comparison results for objective evaluations indicate that adaptive difficulty settings reduce the average error between the target and actual success rates more than fixed difficulty settings, confirming that the difficulty levels in the former can be set according to the user's skill. Subjective evaluations demonstrate that stepwise difficulty levels may be consistent with subjective difficulty levels in the adaptive difficulty setting. Moreover, we clarify that, in adaptive difficulty settings, there is a strong correlation between successful experiences in imagination and self-efficacy, which is not observed in fixed difficulty settings. This means that adaptive difficulty settings can be modified to accurately reflect subjective difficulty levels of the user using a feedback system that utilizes this correlation.

Keywords: Difficulty adjustment, Virtual reality, User model, Self-efficacy

1. INTRODUCTION

Virtual reality (VR) technology has been widely applied in the field of games and sports. The exergames integrating with VR technology have shown promising results in enhancing user motivation [1]. A key element that effectively entertains users in some exergames is difficulty. According to flow theory, one loses track of time and sensations and concentrates on a task when in the flow state [2]. This state is likely to occur when both the difficulty of the challenging task and the one's skill are balanced at a high level. For this reason, many works presented dynamic difficulty adjustment (DDA) systems to adjust the balance between difficulty and performance [3][4].

Regarding the DDA approach, many of these systems were limited to a specific range of difficulties, only balancing the difficulty to a success rate of 50% [5]. They followed a strategy that lowers the difficulty when the user fails and raises it when the user succeeds. However, some works show that a slightly higher success rate than 50% can increase user motivation and avoid discouraging users by ensuring that they succeed more often than they fail [6]. In addition, setting the difficulty only to a moderate level of 50% is insufficient. Some users expect high difficulty to significantly improve their skills, while others want to have fun and hope to succeed easily with low difficulty. In such situations, the DDA system that can target any difficulty level is significant. Thus, it is preferable to set stepwise difficulty levels according to the user's skill to adjust the difficulty between them. When considering the setting of stepwise difficulty levels that adapt to each user's skill to realize specific success rates, we call it adaptive difficulty setting.

If stepwise difficulty levels corresponding to success rates of 75, 50, and 25% are defined as easy, normal, and hard, respectively, checking whether the success rate matches when performing tasks at those difficulty levels is an effective straightforward method to evaluate the adaptive difficulty setting. However, even if the predicted success rate matches the actual success rate, it does not imply that the related difficulty level corresponds to the difficulty level experienced by the individual. Thus, difficulty settings that support not only physical performance achievement, but also psychological aspects will be a significant component [7].

When considering the subjective evaluation of difficulty setting, we focus on self-efficacy and discuss the effectiveness of adaptive difficulty setting. Self-efficacy is a psychological term that refers to the perceived confidence that one will be able to perform a given behavior in the future. Bandura pointed out that self-efficacy has been consistently linked to successful experiences [8]. This indicates that self-efficacy increases if the difficulty level decreases appropriately at each lower level in the adaptive difficulty setting. For this reason, self-efficacy can be used to evaluate users' subjective experiences with difficulty level changes in difficulty settings such as [9].

The objective of this study is to verify the effectiveness of adaptive difficulty setting by proving the following two hypotheses: (1) Our method with adaptive difficulty set-

† Yusuke Goutsu is the presenter of this paper.

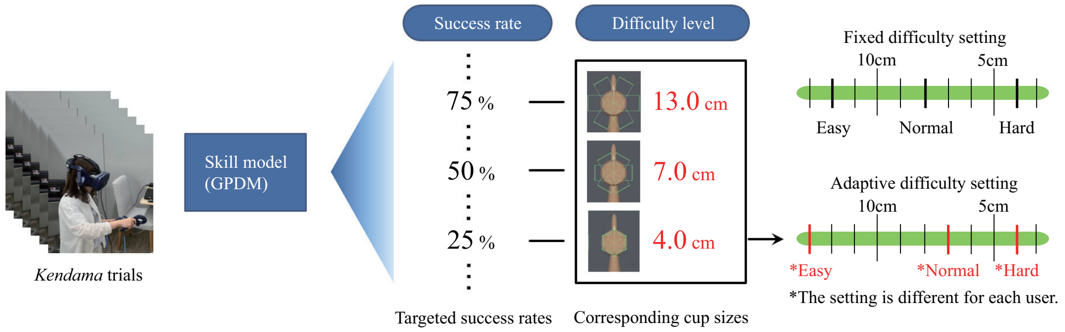

Fig. 1 Overall system flow to realize adaptive difficulty setting, which consists of stepwise difficulty levels, proposed for DDA in a VR *Kendama* game.

ting can accurately predict the success rate at a specific difficulty level. (2) Adjusting the difficulty level in the adaptive difficulty setting can affect self-efficacy based adequately on Bandura's theory. Figure 1 shows an overall system flow to realize adaptive difficulty setting. We use a stochastic temporal prediction method based on the Gaussian process dynamical model (GPDM) to construct the skill model that can capture the correspondence between difficulty and performance. The skill model can determine stepwise difficulty levels corresponding to target success rates for each user, and an adaptive difficulty setting is composed of them. As a sample game of this method, we employ a VR *Kendama* (a cup-and-ball game) task, in which the user tries to catch a ball with a cup that can change in size depending on the difficulty. In the experiments, the adaptive difficulty setting for each user is compared with the fixed difficulty setting in which the difficulty levels are uniformly fixed for all users. We use the validity of stepwise difficulty levels that realize targeted success rates for objective evaluation and the correlation coefficients between successful experiences (actual and imagined successes) and self-efficacy for subjective evaluation.

2. MATERIALS AND METHODS

2.1. VR Kendama System

This study deals with VR *Kendama* as in previous work [10]. Figure 2 shows the VR system. The VR system is based on an original open source project[1] developed using the Unity platform (version 2021.3.5f1). The user wears a head-mounted display called HTC Vive Pro Eye and holds hand controllers with both hands (left: virtual hand, right: the *Kendama*'s sword or "ken"). As shown in Fig. 2, the user plays *Kendama* while watching the projected first-person perspective scene. The positions and postures of the virtual left hand and the ken are correlated with those of the hand controllers. The ball, with a diameter of approximately 6.0cm, is linked to the ken by a 53cm length string. When the distance between the ken and the ball exceeds 53cm, a force drawing the ball toward the ken is generated with a significantly high spring constant, replicating the physical behavior of the string.

[1] https://github.com/y23586/vrchat-misc

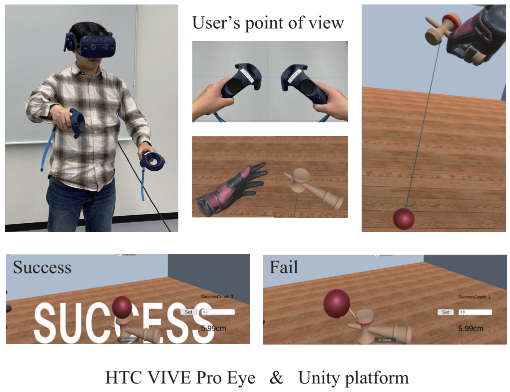

Fig. 2 Overview of VR *Kendama* system.

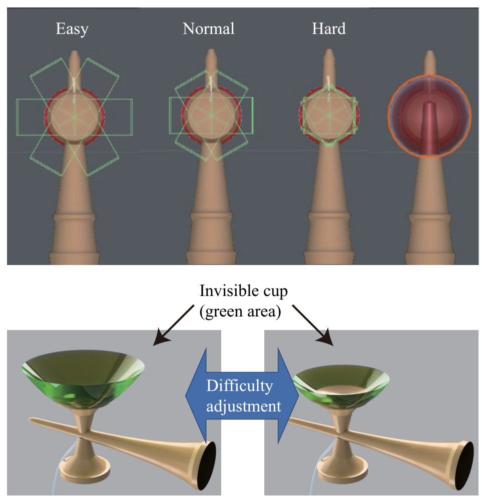

Fig. 3 The variety of cup sizes of VR *Kendama* according to the difficulty levels.

Figure 3 shows the relationship between the difficulty and cup size of VR *Kendama*. In the figure, the leftmost one is the easiest level (the largest diameter), and the second one from the right is the hardest level (the smallest diameter). The green frame is the collision detection area. As shown in the bottom of the figure, the green frame is concave like a bowl when viewed from the side. The green area is not visually indicated in the VR environment to prevent users from consciously realizing their adjustment. In other words, users always see the original cup size as shown on the rightmost side of the figure. The difficulty levels correspond to the cup's diameters, and the green area indicates the size of the enlarged cup (invisible to users). This invisible cup is enlarged only when the ball fell from above and maintained its original size of 4.0cm while the string pulled the ball up. This prevents the ball from colliding with the virtually enlarged cup during rising, ensuring that the users do not experience any strangeness.

2.2. Skill Model based on GPDM

Similarly to previous work [10], the personalized skill model for each user is constructed based on GPDM [11]. GPDM can predict sequences by learning latent representations (states) in a low-dimensional space and the dynamics of the latent states. In addition, the latent state is stochastically generated as it evolves over time since GPDM is a probabilistic method that considers the uncertainty of predictions. This state is expressed by two functions that follow a Gaussian distribution: a latent mapping f that projects observations into a latent space to convert latent states and a dynamical mapping g that makes the transition of latent states. GPLVM [12] is a non-linear dimensional compression method that transforms an observed feature $\boldsymbol{x}_t \in \mathbb{R}^{\mathbb{D}}$ in a high-dimensional observation space into a latent state $\boldsymbol{z}_t \in \mathbb{R}$ in a low-dimensional latent space. GPDM extends this structure to handle the dynamics of latent state $\boldsymbol{z}_t$ and is then defined by

$$\boldsymbol{x}_t = f(\boldsymbol{z}_t) + \boldsymbol{\epsilon}_{x,t} \quad (1)$$

$$\boldsymbol{z}_{t+1} - \boldsymbol{z}_t = g(\boldsymbol{z}_t) + \boldsymbol{\epsilon}_{z,t} \quad (2)$$

Here, $\boldsymbol{\epsilon}_{x,t}$ and $\boldsymbol{\epsilon}_{z,t}$ are zero-mean, isotropic, white Gaussian noise processes. Unlike the original GPDM, we consider $\boldsymbol{z}_{t+1} - \boldsymbol{z}_t$ to be the output of the dynamical mapping according to [13] to improve the smooth transitions in the latent space. We can model Eq.(1) and Eq.(2) as the latent map model and the dynamical map model using Gaussian processes (GPs), respectively. We use a maximum likelihood estimation (MLE) to optimize unknown hyper-parameters of these models under the assumption that prior distributions follow GPs. Regarding the maximization of the negative log-likelihood in MLE, we use a gradient method known as the Adam algorithm. Additionally, we choose the RBF kernel for the latent map model and the RBF+linear kernel for the dynamical map model, respectively. A learned model can be used to serve two distinct functions: (i) map a given new latent state $\boldsymbol{z}_t^*$ to the corresponding $\boldsymbol{x}_t^*$ in the observation space, (ii) predict the evolution of the latent state at the next time step $\boldsymbol{z}_{t+1}^*$ given $\boldsymbol{z}_t^*$. In this way, given an initial state in the latent space, temporal data in the observation space is generated by repeatedly mapping the state and transitioning to the next state according to a Gaussian distribution.

2.3. Evaluation Method of Difficulty Setting

We assign different cup sizes for stepwise difficulty levels. Here, we consider three difficulty levels of $hard$, $normal$, $easy$ as a difficulty setting. In the adaptive difficulty setting, the cup sizes predicted to achieve 25%, 50%, and 75% success rates using the skill model are considered adaptive $hard$, adaptive $normal$, and adaptive $easy$, respectively. Here, the term "adaptive" means that three difficulty levels are individually set for each user.

One significant aspect regarding the evaluation metrics of difficulty setting is that three difficulty levels must be adaptive to the individual skill. This aspect is a significant factor to consider, as adjusting the difficulty level in an inappropriate difficulty setting could not affect the success rate. Obviously, difficulty adjustment in such a situation is not expected to be effective. For this reason, we use the validity of stepwise difficulty levels that realize target success rates ($hard$:25%, $normal$:50%, $easy$:75%) to evaluate the difficulty setting.

Furthermore, the psychological aspect is equally essential to subjectively assess the difficulty setting, and we consider the enhancement of self-efficacy when changing the difficulty level in the $Kendama$ game. We investigate the relationship between successful experiences and self-efficacy. Here, successful experiences refer to (i) actual successes, (ii) successes in memory, and (iii) successes in future expectations. The first and the rest are based on the Bandura and Maddux hypotheses, respectively [8][14].

2.4. Prediction of Cup Sizes Corresponding to Target Success Rates using Skill Model

We record various physical quantities related to the ball, cup, and both hands during the $Kendama$ trials. Here, the observational data for training is the ball's relative position seen from the cup, which are three-dimensional sequential data. We split the observational data by trial as input for training the skill model described in the previous section. Regarding the estimation of model parameters, we applied an MLE with Adam. Here, we set the learning rate to 0.001, the number of training iterations to 12,000 to ensure sufficient learning convergence, and the dimension of the latent space to d=3.

The sequential prediction of the skill model generates a ball trajectory by sampling data with stochastic processes. In other words, a simulation of the $Kendama$ trial is performed using this prediction. The trajectory of the ball is considered successful when the ball stayed within the cup. We derive the success rate by multiple random samplings based on the Monte Carlo method, which means that the success rate is calculated by the ratio of successes among the generated ball trajectories. Here, we set the number of random samplings to 40. Our method proposed in this study can derive the diameter of the cup corresponding to the target success rate by capturing the relationship between user performance and task difficulty, which is different from the method of [10].

2.5. Measurement of Subjective Sense around Self-efficacy

We employ a questionnaire to measure subjective senses. Specifically, the measurement targets are self-

Table 1 Questionnaire to measure subjective senses. Participants are asked to rate their level of self-efficacy on a 7-point Likert scale for Q1-4 and to write the number of successes for Q5 and Q7.

No.	Question
1	If you were asked to increase the success rate of $Kendama$, do you think you could achieve it?
2	Do you feel that you would be able to cope well with the increased difficulty of $Kendama$?
3	If you had to do a different kind of juggling action in VR, do you feel you could handle it well?
4	Compared to others, do you think you are more successful at $Kendama$?
5	How many times were you able to put the ball on the cup in the previous experiment? If your memory is not clear, a rough count is fine. (Please describe the number of times)
6	How certain do you seem to be of this memory?
7	How many times do you think you can get the ball onto the cup out of 10 tries next time? (Please describe the number of times)
8	How confident are you in that expectation?

efficacy, the number of successes in memory, and the number of successes in future expectation. Table 1 shows the content of the questionnaire. Here, users are asked to rate their level of self-efficacy on a 7-point Likert scale for Q1-4 and to write the number of successes for Q5 and Q7. Self-efficacy is evaluated by summing the points from Q1 to Q4, and therefore the maximum score is 7×4=28.

3. EXPERIMENT

3.1. Procedure

We asked 13 participants (male: 10/ female: 3, 19-45 years of age) who did not have physical problems with wearing VR devices to play the $Kendama$ game. In addition, we instructed participants who had no experience with VR on how to use VR devices and provided them with some practice time before actual VR $Kendama$ performance.

As described in the previous section, the cup sizes corresponding to the three difficulty levels at which the success rates are predicted to become 25, 50, and 75% are assigned by the skill model in the adaptive difficulty setting. We constructed the skill model with training data from 20 $Kendama$ performance observations, which are the sequences of relative position between the ball and the cup. Table 2 shows the diameters of the cup at the three difficulty levels for each participant. The mark '-' means the case in which the skill model could not determine the diameter of the cup, which occurred only in the case of adaptive $easy$. The reason for this indetermination is that the success rate did not increase to 75%, regardless of how large the cup diameter was in the prediction. Generally, the more the ball lands on the outside of the cup in predicting the ball trajectory, the less likely it is to stay on the cup (not judged as success). In that case, the successful ball trajectories concentrated on the center of the cup, which could lead to the indetermination issue.

In addition, we also used a fixed difficulty setting in which we empirically determined the cup diameters corresponding to the levels $hard$, $normal$, and $easy$, which are predicted to have success rates of 25, 50, and 75%, respectively, under the condition of equal intervals between difficulty levels. The three difficulty levels are fixed for all experimental participants. Here, the fixed $hard$, fixed $normal$, and fixed $easy$ were set to 4, 8, and 12cm, respectively.

Table 2 Cup sizes of VR $Kendama$ in adaptive difficulty setting. The cup diameters at three difficulty levels are obtained from the skill model to adapt to each participant's skill.

ID	Cup's diameter (cm)		
	Adaptive $hard$	Adaptive $normal$	Adaptive $easy$
1	4.6	11.4	-
2	7.4	10.2	13.5
3	4.8	10.6	-
4	3.6	7.5	12.6
5	6.7	8.1	9.8
6	7.7	15.8	-
7	5.8	8.9	18.2
8	3.6	10.2	12.3
9	6.3	8.5	11.2
10	7.9	11.7	14.2
11	6.4	9.4	12.9
12	14.8	16.1	18.6
13	7.7	10.0	-

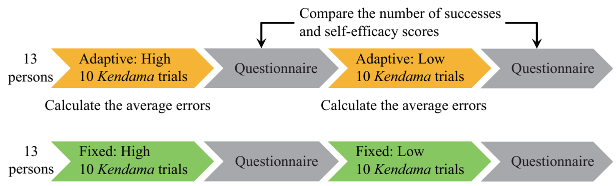

Fig. 4 Experimental protocol in the evaluation phase of difficulty settings. All participants play the $Kendama$ game in the same experimental conditions between adaptive and fixed difficulty settings.

Figure 4 shows the experimental protocol in the evaluation phase of the difficulty settings. As shown in this figure, we calculated the average errors between the actual and target success rates for each session of 10 $Kendama$ trials. We also compared the number of successes related to successful experiences and self-efficacy scores when decreasing the difficulty level from high to low in both adaptive and fixed difficulty settings. In other words, there were two types of difficulty changing: $hard$ to $normal$ and $normal$ to $easy$. Note that the order of three difficulty levels to be set in the $Kendama$ trial was randomly determined for each participant.

3.2. Result

We compared the adaptive difficulty setting with the fixed difficulty setting to evaluate their adaptiveness for the user's skill. Table 3 shows the actual success rates derived from 10 trials of the $Kendama$ game at the cup's

Table 3 Comparison between adaptive and fixed difficulty settings regarding the average error between actual and target ($hard$:25%, $normal$:50%, $easy$:75%) success rates at three difficulty levels for each participant.

ID	Actual success rate (%)			Error	Actual success rate (%)			Error
	Adaptive $hard$	Adaptive $normal$	Adaptive $easy$		Fixed $hard$	Fixed $normal$	Fixed $easy$	
1	0	70	-	22.5	0	20	-	27.5
2	30	50	70	3.3	0	50	70	10.0
3	10	60	-	12.5	10	30	-	17.5
4	10	50	80	6.7	20	70	80	10.0
5	20	60	70	6.7	0	30	70	16.7
6	20	50	-	2.5	0	30	-	22.5
7	20	40	100	13.3	0	40	70	13.3
8	20	60	70	6.7	20	50	70	3.3
9	30	50	70	3.3	0	30	60	20.0
10	10	70	90	16.7	0	0	70	26.7
11	20	50	60	6.7	10	50	80	6.7
12	30	60	60	10.0	0	0	50	33.3
13	30	50	-	2.5	20	70	-	12.5
All				**8.7**				16.9

Table 4 Comparison between adaptive and fixed difficulty settings regarding the self-efficacy scores at three difficulty levels for each participant.

ID	Self-efficacy score			Self-efficacy score		
	Adaptive $hard$	Adaptive $normal$	Adaptive $easy$	Fixed $hard$	Fixed $normal$	Fixed $easy$
1	9	13	-	10	12	-
2	10	13	14	6	14	15
3	13	14	-	14	17	-
4	10	15	19	14	17	20
5	14	14	15	13	14	14
6	12	20	-	8	15	-
7	10	12	14	7	12	14
8	18	20	20	8	13	22
9	13	14	16	16	16	18
10	9	15	17	9	12	18
11	14	17	20	16	20	23
12	11	13	14	12	16	22
13	14	16	-	9	15	-

Table 5 Comparison between adaptive and fixed difficulty settings regarding the correlation coefficients of successful experiences (actual successes, successes in memory, and successes in future expectations) with self-efficacy at three difficulty levels.

Factor	Correlation coefficients with self-efficacy			Correlation coefficients with self-efficacy		
	Adaptive $hard$	Adaptive $normal$	Adaptive $easy$	Fixed $hard$	Fixed $normal$	Fixed $easy$
actual successes	0.34	-0.02	-0.19	0.13	0.30	-0.02
successes in memory	**0.58**	0.23	0.03	0.13	0.16	-0.05
successes in the future expectation	**0.66**	**0.60**	**0.78**	0.14	0.20	**0.64**

diameters corresponding to the levels $hard$, $normal$, and $easy$ in both the adaptive and fixed difficulty settings. As the success rates are expected to be 25, 50, and 75% at the levels $hard$, $normal$, and $easy$, respectively, we checked the errors between the actual and target success rates. The smaller the error, the more adaptively the difficulty levels are set for the user's skill. Here, the mark '-' originates from the one corresponding to the same participant ID and difficulty level in Tab. 2. We confirmed that the average error in the adaptive difficulty setting, which was within 10pt, became much smaller than that of the fixed difficulty setting. This means that the GPDM-based skill model can set the three difficulty levels to become target success rates even though individuals have different skills.

We conducted a questionnaire survey after the participant tried $Kendama$ at each difficulty level and calculated the self-efficacy score using the questionnaire in both adaptive and fixed difficulty settings. Table 4 summarizes the results. The result demonstrated that almost all the transitions between difficulty levels from $hard$ to $normal$ and from $normal$ to $easy$ enhanced self-efficacy. When additionally considering the result of Tab. 3, increasing the self-efficacy scores and actual success rates co-occurred when lowering the difficulty level. It was unexpected that the self-efficacy score increased when the actual success rate increased in fixed difficulty settings. This means that three difficulty levels can be set in both adaptive and fixed difficulty settings with appropriate difficulty differences between them.

As factors associated with success experiences, there are successes in memory and successes in future expectations in addition to actual successes. Table 5 shows the correlation coefficients between the number of successes in those factors and the self-efficacy score in both adaptive and fixed difficulty settings. Here, the correlation coefficients of 0.4 or more are in bold. The result demonstrated that the number of successes in future expecta-

Table 6 Comparison between adaptive and fixed difficulty settings regarding the correlation coefficients of increase in successful experiences with the enhancement of self-efficacy when adjusting the difficulty level.

Factor		Correlation coefficients with the enhancement of self-efficacy
Increased actual successes	Fixed $hard$ to $normal$	0.37
	Fixed $normal$ to $easy$	0.23
	Adaptive $hard$ to $normal$	0.25
	Adaptive $normal$ to $easy$	0.36
Increased successes in memory	Fixed $hard$ to $normal$	0.35
	Fixed $normal$ to $easy$	0.22
	Adaptive $hard$ to $normal$	**0.41**
	Adaptive $normal$ to $easy$	**0.47**
Increased successes in future expectations	Fixed $hard$ to $normal$	0.10
	Fixed $normal$ to $easy$	**0.58**
	Adaptive $hard$ to $normal$	**0.46**
	Adaptive $normal$ to $easy$	**0.75**

tions was correlated with the self-efficacy score at all adaptive difficulty levels, which was not seen in the fixed difficulty setting. To investigate this correlation in more detail, Table 6 shows the correlation coefficients of the increase in successful experiments with the improvement in the self-efficacy score by reducing the difficulty level from $hard$ to $normal$ and from $normal$ to $easy$. Here, the correlation coefficients of 0.4 or more are in bold. In both adaptive and fixed difficulty settings, we verified that there are weak correlations between the increased number of actual successes and the enhancement of self-efficacy, which are previously discussed in Tab. 3 and 4. Additionally, the adaptive difficulty setting had strong correlations between the increase in successful experiences in imagination (successes in memory and successes in future expectations) and the improvement in the self-efficacy score compared to the fixed difficulty setting. This means that, in the adaptive difficulty setting, as the successful experiences in imagination increase when reducing the difficulty level, self-efficacy is also enhanced, which supports what Maddux has pointed out as Imagined Experience. We also confirmed that the correlations between increased successful experiences in imagination and the enhancement of self-efficacy were stronger than the correlations between increased actual successes and the enhancement of self-efficacy in the adaptive difficulty setting. This means that the increased successful experiences in imagination more influence the enhancement of self-efficacy than the increased actual successes in the adaptive difficulty setting.

4. CONCLUSION

In this study, we verified the effectiveness of adaptive difficulty setting, which is proposed for DDA, in both objective and subjective aspects in the VR $Kendama$ system. The adaptive difficulty setting is stepwise difficulty levels that reflect each user's skill using the GPDM-based skill model, and we showed that they may be consistent with subjective difficulty levels. In addition, there was a strong correlation between successful experiences in imagination and the subjective indicator of self-efficacy in adaptive difficulty settings. This means that it is possible to create a feedback system that makes subjective modifications to the stepwise difficulty levels even if they do not match subjective difficulty levels for future work.

Acknowledgement: This work was supported by JST [Moonshot R&D][Grant Number JPMJMS2034].

REFERENCES

[1] C. Zhang and S. Yu, "Technology to enhance patient motivation in virtual reality rehabilitation: A review", *Games for Health Journal*, vol. 13, no. 4, pp. 215–233, 2024.

[2] M. Csikszentmihalyi, S. Abuhamdeh, and J. Nakamura, "Flow", *Flow and the Foundations of Positive Psychology*, pp. 227–238, 2014.

[3] F. Mortazavi, H. Moradi, and A.-H. Vahabie, "Dynamic difficulty adjustment approaches in video games: A systematic literature review", *Multimedia Tools and Applications*, pp. 1–48, 2024.

[4] P. D. Paraschos and D. E. Koulouriotis, "Game difficulty adaptation and experience personalization: A literature review", *International Journal of Human–Computer Interaction*, vol. 39, no. 1, pp. 1–22, 2023.

[5] W. R. Fernandes and G. Levieux, "δ-logit: Dynamic difficulty adjustment using few data points", in *Joint International Conference on Entertainment Computing and Serious Games*. Springer, 2019, pp. 158–171.

[6] A. Sch¨utt, T. Huber, I. Aslan, and E. Andr´e, "Fast dynamic difficulty adjustment for intelligent tutoring systems with small datasets", in *Proc. of the 16th International Conference on Educational Data Mining*, 2023, pp. 482–489.

[7] R. Mandryk and M. Klarkowski, "Physiological measures for game evaluation", in *Game Usability*. CRC Press, 2022, pp. 161–187.

[8] A. Bandura, "Self-efficacy: Toward a unifying theory of behavioral change", *Psychological Review*, vol. 84, no. 2, pp. 191–215, 1977.

[9] T. Inamura, N. Takahashi, and K. Nagata, "Enhancing self-efficacy in VR Kendama juggling task through intervention and adjustment of virtual success experience", *Journal of Robotics and Mechatronics*, vol. 37, no. 3, pp. 614-626, 2025.

[10] Y. Goutsu and T. Inamura, "Instant difficulty adjustment using user skill model based on GPDM in VR Kendama task", in *Proc. of the 2024 IEEE International Conference on Artificial Intelligence and eXtended and Virtual Reality*, ser. AIxVR '24. IEEE, 2024, pp. 138–146.

[11] J. Wang, A. Hertzmann, and D. J. Fleet, "Gaussian process dynamical models", in *Proc. of the 18th International Conference on Neural Information Processing Systems*, ser. NIPS '05. ACM, 2005, pp. 1441–1448.

[12] N. Lawrence, "Gaussian process latent variable models for visualisation of high dimensional data", in *Proc. of the 16th International Conference on Neural Information Processing Systems*, ser. NIPS '03. ACM, 2003, pp. 329–336.

[13] F. Amadio, J. A. Delgado-Guerrero, A. Colomé, and C. Torras, "Controlled Gaussian process dynamical models with application to robotic cloth manipulation", *International Journal of Dynamics and Control*, vol. 11, no. 6, pp. 3209-3219, 2023.

[14] J. E. Maddux, "Self-efficacy: The power of believing you can", *Handbook of Positive Psychology*, pp. 277–287, 2002.