Exergame Experience of Young and Old Individuals Under Different Difficulty Adjustment Methods

Oral Kaplan 1,* , Goshiro Yamamoto 2 , Takafumi Taketomi 1 , Alexander Plopski 1 , Christian Sandor 1 and Hirokazu Kato 1

1 Division of Information Science, Nara Institute of Science and Technology, Nara 630-0192, Japan; takafumi-t@is.naist.jp (T.T.); plopski@is.naist.jp (A.P.); sandor@is.naist.jp (C.S.); kato@is.naist.jp (H.K.)
2 Division of Medical Information Technology and Administration Planning, Kyoto University, Kyoto 606-8507, Japan; goshiro@kuhp.kyoto-u.ac.jp

* Correspondence: oral.kaplan.nv4@is.naist.jp; Tel.: +81-743-72-5336

Received: 4 October 2018; Accepted: 5 November 2018; Published: 7 November 2018

Abstract: In this work, we compare the exergaming experience of young and old individuals under four difficulty adjustment methods. Physical inactivity is a leading cause of numerous health conditions including heart diseases, diabetes, cancer, and reduced life expectancy. Committing to regular physical exercise is a simple non-pharmaceutical preventive measure for maintaining good health and sustaining quality of life. Incorporating exercise into games, studies frequently used exergames as an intervention tool over the last decades to improve physical functions and to increase adherence to exercise. While task difficulty optimization is crucial to exergame design, researchers consistently overlooked age as an element which can significantly influence the nature of end results. We use the Flow State Scale to analyze the mental state of young and old individuals to compare constant difficulty with ramping, performance-based, and biofeedback-based difficulty adjustments. Our results indicate that old individuals are less likely to experience flow compared to young under the same difficulty adjustment methods. Further investigation revealed that old individuals are likely to experience flow under ramping and biofeedback-based difficulty adjustments whereas performance-based adjustments were only feasible for young.

Keywords: human-computer interaction; exergames; difficulty adjustments; user experiences

1. Introduction

Health problems associated with sedentary lifestyles are undeniably critical issues of our time. According to the World Health Organization, 23% of the world population aged over 18 was insufficiently active in 2010 [1]. This number grows to a striking 81% among adolescents aged 11 to 17 years old [1]. Insufficient physical activity is one of the major causes of high mortality rates around the globe. Every year, lack of regular physical exercise and sedentary living lead to approximately 3.2 million deaths all around the globe [2]. People who exercise on a regular basis have a 20% to 30% lower risk of death from various diseases compared to people who do not [3]. Chronic conditions associated with physical inactivity such as heart diseases, cancer, and diabetes are responsible for 1.7 million deaths every year in United States alone [4]. Medical costs associated with physical inactivity reach up to a massive $67.5 billion on a worldwide scale [5].

Conventional approaches used to tackle this health crisis typically concentrate on reducing personal and environmental factors associated with physical inactivity such as obesity, stress, fatigue, associated cost, lack of time, and irregular sleep cycles [6–10]. Educational institutions consistently use physical training classes to teach children about the benefits of healthy living and an active lifestyle [11]. The World Health Organization regularly publishes recommendations on regular physical activity, but despite these efforts, many people still fail to meet the recommended levels of physical activity.
activity for health [12]. A growing body of knowledge on high-intensity interval training (HIIT) protocols aims to resolve problems associated with time constraints through low volume high intensity endurance exercise interventions [13]. Possibly the most well-known of them all to general public, video games that require physical activity, has been gaining significant attention among research communities [14,15].

A portmanteau of exercise and gaming, exergames are frequently used to provide means of engagement in physical and social activities that are also a form of exercise [16–18]. Although the health benefits of exergames are still subject to debate [19–22], many studies confirm their potential as a motivator in performing regular physical exercise [23–26]. Commercially successful exergames demonstrated how physical movements can be embedded in games to discourage sedentary lifestyles among players on a global scale. Konami’s Dance Dance Revolution was one of the first success stories of this genre, selling over three million copies after being ported from arcade to PlayStation [27]. Nintendo Wii Fit’s sales reached a stunning 21 million after the second year of its release in 2007 [28]. Researchers from a broad spectrum used both Wii Fit’s inertial measurement unit based controllers and its balance board for therapeutic interventions such as stroke or balance rehabilitation, functional improvements in individuals with brain injuries, and children with pediatric disabilities [26,29–32]. Following this trend, Microsoft Kinect allowed players to use their body as a remote controller, opening the door to a new world for researchers. Similar to Wii, researchers mainly focused on improving muscle strength and balance, rehabilitation, and self-efficacy for physical activity and sports [33–36].

The ever-growing body of knowledge supports the positive impact of exergames in satisfying the need for entertainment, self-esteem, success, and achievement among sedentary individuals. Researchers reported improvement in hand-eye coordination, recovery of fine motor skills, and enhancements in attention [37–39]. Falls are the most common accident among older adults, and preventing them has been used as an implicit objective in numerous exergames [40–43]. Positive social effects between individuals has been a consistently cited benefit of multiplayer exergames; significantly contributing to a growth in intrinsic motivation for regular exercise [17,44,45].

While research into exergames mainly explored its application areas, design, and benefits, achieving optimal play experience through game difficulty balancing according to player skill continues to remain unclear. As a crucial component of the modern game design pipeline, difficulty balancing aims to minimize the effects of skill variations through challenge adjustments to optimize player experiences [46]. Commonly referred to as difficulty adjustments, recent work examined the effects of static and dynamic techniques in both exergame and non-exergame settings [47,48]. However, shortage in comparative empirical studies and use of homogeneous player groups aged between 18 to 30 fails to justify their applicability to broader audiences including old individuals [49,50].

To address these issues, and motivated by adequate exergame applications, we have compared constant, ramping, performance-based, and biofeedback-based difficulty adjustment methods while taking player age range into consideration. We conducted a within-subject repeated measures experiment through an exergame of our design (Figure 1a). Our sample consisted of young and old individuals and participants tried all four approaches to difficulty adjustments (Figure 1b,c). We employed Flow State Scale (FSS) to measure play experience of our sample group under each condition and examined the results based on difficulty adjustments, age groups, and their interaction. We found significant differences and statistical analysis revealed information in line with previous work when age is not taken into consideration. Additionally, overall FSS score was lower between older individuals and their difficulty adjustment preferences were different from young individuals. Finally, we deployed our system in a rehabilitation hospital for old individuals to observe its applicability in a real life scenario.

Our results contribute to the growing field of serious games by advancing our understanding of factors associated with appropriate task difficulty and optimal play experience. Taken together, we consider them substantially helpful in enabling designers to make informed decisions for
administering satisfactory exergaming experiences, and lead to a further increase in regular practice of physical exercise to bolster healthy lifestyles.

![Figure 1](image1.png)  
**Figure 1.** Exergame design and participants from both age groups. (a) Screenshot of our game where the user-controlled airplane and color coded target rings can be seen. (b) Facial expression of a participant during one of our experiment sessions with young individuals. (c) Snapshot taken from a video of an old individual which was used during post-experiment discussions with the same participant.

### 2. Related Work

In this section, we first review the attempts to use games as a form of health intervention. We divide our investigation into two by taking age range into consideration and focus on young and old individuals. We then briefly explain difficulty adjustments and present various factors used to optimize gaming experience of players. We conclude by introducing a few, highly-influential works that express the potential of applying these factors to exergaming scenarios.

#### 2.1. Purposing of Games for Health

Over the last decade we have witnessed a rapid rise in the use of novel digital concepts that combine exercise and games, such as Nintendo Wii and Microsoft Kinect. Labeled active games, exertainment, or exergames, Mueller et al. defined this genre as digital games where the outcome of the game is predominantly determined by physical effort [51]. We adhere to this definition and use the term **exergames** for referring to these games in our investigation of the vast amount of existing literature. We divide our investigation into two age based groups: young and old individuals.

Researchers have been widely investigating the physiologic and psychological effects of exergames on young individuals that can contribute to short or long-term weight maintenance. Straker and Abbott compared traditional and physically active gaming among 9 to 12 year-old children and found increased cardiovascular response and energy expenditure during the latter scenario [52]. Graves et al. compared physiological cost and enjoyment of exergaming on Nintendo Wii Fit with aerobic exercise and found it stimulating in light-to-moderate intensity exercise [53]. Staiano et al. examined physical and psychological effects of a 20-week exergame intervention on overweight and obese adolescents, and reported the effectiveness of cooperative exergames on weight loss among young individuals [25]. Althoff et al. studied location-aware multiplayer games that combine gameplay with the physical world and reported increased physical activity levels among players [54]. Douris et al. compared physiologic and psychological responses to 30-min of treadmill walking with Nintendo Wii Fit Free Run and described the latter as a potent alternative to traditional moderate-intensity exercise [55]. Staiano and Calvert outlined the literature on exergames used in physical education courses in schools and provided several physical, social, and academic benefits as related outcomes [23]. Staiano et al. also reported an increase in cognitive skills of adolescents as an outcome of competitive exergames and found a relationship between this increase and weight loss [24].

Old individuals have also been gaining much attention from exergame community due to their lack of motivation for physical activity, physical fitness, and mobility in extremities [56]. Billis et al.
developed a game platform using Nintendo Wii Balance Board in an attempt to increase exercise motivation and physical well being among older individual groups [57]. Pedram et al. addressed social, environmental, and physical issues of the elderly by designing an augmented dancing environment and provided design guidelines for creating similar interactive designs [44]. Gerling et al. evaluated the effects of age-related impairments on exergaming experience of frail old individuals and indicated their need to be considered in game design process [16]. Ijsselsteijn et al. identified augmented and virtual reality exergames as stimulating on intrinsic motivation for regular exercise [58]. Durick et al. explored and presented alternative approaches to aging myths in exergames for older individuals [59]. Skjaæret et al. carried out investigations into motion characteristics of older individuals in stepping exergames designed to increase movement quality through fall preventive exercises [60]. Gerling et al. examined exergame design for institutionalized older individuals and reported increased quality of life among players [61]. Romero et al. demonstrated how inclusion of family members or caretakers into exergame design could lead to persuasive solutions among older individuals, and increase motivation for social and physical activities [45]. Brox et al. focused on multiplayer exergames for older individuals and highlighted their power in preventing loneliness through social interaction in physical activity [17].

2.2. Difficulty Adjustments in Exergames

Balancing of game difficulty based on player skill has been widely recognized as a crucial element of game design to avoid boredom or frustration among players. Liu et al. point out the unsuitable and interruptive nature of static difficulty adjustment methods in realizing optimal play experience [50]. They designed a real-time, anxiety-based affective feedback mechanism for dynamic difficulty adjustment (DDA). Smedding et al. used adaptive difficulty parameters for speed, accuracy, and range of motion to design a challenging yet suitable exergame for Parkinson disease patients [47]. Baldwin et al. conducted a formal review of existing competitive multiplayer games and suggested a preliminary framework for classifying Multiplayer Dynamic Difficulty Adjustment (mDDA) instances [62]. Baldwin et al. also investigated the effects of multiplayer DDA awareness on play experience and found a negative relationship between the two [46].

Although research in game difficulty adjustments gained a significant momentum over the last years, to our knowledge, research on difficulty adjustment in exergames is scarce. Most studies follow a similar practice as [50] and focus on designing instruments or models for adjustment optimization. Hunicke described the design requirements of effective DDA systems that do not disrupt or degrade the play experience [63]. Jennings et al. suggested the use of level generation and machine learning techniques to achieve a DDA method that can change both structural and personalized difficulty elements [64]. Moreover, due to an inclination to adopt homogeneous and young sample groups, there is still considerable ambiguity with regard to impact of age related changes on difficulty adjustments mechanisms [49].

Much work has consistently been focusing on aging-related decline of physical or cognitive functions instead of game difficulty optimization. Interventions to prevent falls in older individuals is a common area of interest for researchers working on exergames. Lange et al. used a dancing game to administer step based exercises to reduce falls in older individuals [65]. De Groot et al. outlined motivating factors and barriers for older individuals to adhere to group exercises for fall prevention [42]. Uzor and Baillie described the use of tailored exergames to prevent falls and conducted a long-term study where observed better exercise adherence in participants who used exergames compared to standard care [66].

3. Materials and Methods

3.1. Hypotheses

There is still much controversy surrounding aging and difficulty adjustment mechanisms in exergames. In this study our goal was to answer the following questions:
Q1. Do traditional approaches fail to provide a satisfactory play experience?

Q2. Do flow experiences of young and old individuals differ from each other under similar exergaming conditions?

Q3. Is there a difference between difficulty adjustment preference of young and old individuals?

Based on these questions we formulated the following hypotheses:

H1. Flow refers to an equilibrium between challenge presented by the game and the respective skill set of players where maximum enjoyment is achieved. For example, in Wii Sports Tennis, players experience flow when the skill level of the non-playable characters (NPC) are on par with their own. Since controlling player skill is a mundane task, we achieve flow by adjusting game difficulty and modifying game elements such as NPC speed, health, gameplay duration, or power-up frequency. Traditional methods used for these adjustments follow a steady increase in difficulty, mostly in steps represented as game levels. Growth or amount of the adjustment depends on player's self-assessment of his perceived skill and selection of difficulty from a given set of choices such as easy, normal, and hard. This schema causes a decline in flow and leads to a diminished gaming experience [18, 46, 47, 50, 62]. Accordingly, we form our first hypothesis as follows: Constant difficulty will fail to provide a satisfactory exergaming experience compared to ramping, performance-based, and biofeedback-based difficulty adjustments.

H2. Aging-related changes have become a central topic in exergame design for older individuals. As an example, diminished motor control and neural functions affect various aspects of designing exergame experiences for this specific target group. However, effects of these changes on difficulty adjustment preference are still not clearly defined. We think that technology experience and gaming perception of young and old generations differ significantly, which is likely to have a substantial effect on preferred level of difficulty. Combined with varying game preferences, we believe that exergaming experience of young and old generations will demonstrate a considerable amount of dissimilarity under same conditions. Based on this background, our second hypothesis is: Measured exergaming experience of the two age groups will show a significant difference based on Jackson and Marsh's Flow State Scale (FSS) [67].

H3. Based upon the first and second hypotheses, we predict significant differences between difficulty adjustment preference of the two age groups. Building upon this interaction, we form our final hypothesis as follows: The two age groups will favor different difficulty adjustment methods.

3.2. Exergame Design

Throughout the design process, we followed the Waterfall model which is a sequential software engineering approach. First, we gathered requirements by working with Japan's only national sports college and observing traditional exercise classrooms for older individuals at multiple locations such as community halls and rehabilitation hospitals. Consequently, we opted for a competitive game design to address upper extremity exercises due to their widespread admission in physical therapy for older individuals. During our initial studies we developed three exergames and finally we chose the game we used in this study due to its conformity to entertainment criteria derived from the ElderGames project [68].

In our game, players control an aircraft by using predefined discrete arm gestures to pass through rings placed on its path. We used a metaphorical relationship between in-game gestures and their real-life representations to overcome the infamous chocolate covered broccoli problem [69]. As a result, we used four arm gestures that were based on aircraft marshalling movements as in-game controllers (Figure 2). We conducted preliminary studies with young and old individuals to determine a suitable range of motion (ROM) for each gesture.
Rings were color coded and randomly placed at four locations in correspondence with gestures used to represent movement direction. The distance between each ring was used as the game’s difficulty and was modified by the game in accordance to the selected difficulty adjustment method (Figure 3). Under constant difficulty mode the distance between each ring was set to a static value before beginning the game. Ramping adjustments reduced the distance over time regardless of player performance. Performance-based adjustments reduced the distance between each ring depending on the ratio between current and maximum score. Similar to Karvonen [70], biofeedback mode modified the distance based on fluctuations of the recorded data on player’s heart rate variability (HRV). We used Polar H7 output in the system directly (Figure A2). We used the ratio between current and maximum HRV measured at the beginning of each experiment. We increased maximum distance by 50% for older individuals to accommodate the aging related decline of physical and neurological functions. Table 1 represents the distance settings we used with each adjustment method in terms of airplane travel time.

**Table 1.** Distance modification procedure based on difficulty adjustment method

<table>
<thead>
<tr>
<th>Modification</th>
<th>Distance (in Seconds)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>Fixed</td>
</tr>
<tr>
<td></td>
<td>1.0&quot;, 0.75&quot;, 0.5&quot;</td>
</tr>
<tr>
<td></td>
<td>1.5&quot;, 1.0&quot;, 0.5&quot;</td>
</tr>
<tr>
<td>Ramping</td>
<td>Reduced over time</td>
</tr>
<tr>
<td></td>
<td>1.0&quot;~0.5&quot;</td>
</tr>
<tr>
<td></td>
<td>1.5&quot;~0.5&quot;</td>
</tr>
<tr>
<td>Performance-based</td>
<td>Reduced on success</td>
</tr>
<tr>
<td></td>
<td>1.0&quot;~0.5&quot;</td>
</tr>
<tr>
<td></td>
<td>1.5&quot;~0.5&quot;</td>
</tr>
<tr>
<td>Biofeedback-based</td>
<td>Dynamic based on HRV</td>
</tr>
<tr>
<td></td>
<td>1.0&quot;~0.5&quot;</td>
</tr>
<tr>
<td></td>
<td>1.5&quot;~0.5&quot;</td>
</tr>
</tbody>
</table>

We set game’s success criteria as ratio of successful passes through rings. Each successful pass resulted in a one point increase in player’s total score. We used this information to calculate the ratio between accumulated points and total number of rings. This quantitative indicator of success was shown to players at the end of each game session and recorded along with point progression for evaluation purposes.
3.3. Measures

We used Jackson and Marsh’s Flow State Scale (FSS) to measure player’s game experience [67]. It consists of nine scales that represent dimensions of flow discussed by Csikszentmihalyi [71] and each scale incorporates four items. These scales are defined as challenge-skill balance, action-awareness merging, clear goals, unambiguous feedback, concentration on task at hand, sense of control, loss of self-consciousness, transformation of time, and autotelic experience. Each item measures a dimension of the player’s flow on a traditional 5-point Likert scale where 1 represented strongly disagree and 5 represented strongly agree.

In addition to FSS, we recorded in-game data; including point progression, percentage-based score, arm movements, heart rate, and heart rate variability. We also recorded gameplay videos of participants during each session.

3.4. Experiment Procedure

We started with a naive health screening process by using the Physical Activity Readiness Questionnaire (PAR-Q) with our young participants [72]. Health screening of old participants was conducted and guaranteed by the recruiting organization. After confirming participant eligibility, we explained the experiment protocol in detail and obtained informed consent in both age groups. Then, we installed the heart rate monitor and gesture armband on their respective body locations. Afterwards, we seated participants in a front facing posture and measured their maximum heart rate variability with a one minute breathing exercise protocol. We put each participant through a series of practice sessions using constant difficulty. We introduced the four difficulty adjustment methods in randomized order after participant demonstrated a sufficient understanding of the game interface and used in-game gestures (Figure 4). During the experiments, participants had to use arm gestures to steer the airplane through a series of rings, where the color of each ring corresponds to the gesture that will be needed to reach it. Participants played three consecutive games before filling out the FSS for the respective method. This task was repeated with each difficulty adjustment method and sufficiently long breaks were given in between sessions to eliminate residual effects. After completing all tasks, participants were debriefed and discussions were held for 10 to 20 min before their dismissal.
Our study was conducted in accordance with Declaration of Helsinki, and the protocol was approved by the Ethics Committee of Nara Institute of Science and Technology, Japan.

![Image](image_url)

**Figure 4.** Used upper extremity gestures. (a) Left. (b) Up. (c) Right. (d) Down.

### 3.5. Apparatus

The physical setup of our system included a Philips 43” 4K Ultra HD LCD display that was connected to a late 2013 Apple Mac Pro with 3.0GHz 8-Core Intel Xeon E5 processor, 16 GB 1866 MHz DDR3 ECC memory, and dual AMD FirePro D700 6GB GDDR5 VRAM graphics cards. We connected these two devices with a standard HDMI cable.

We used a Polar H7 to monitor participant heart rate. We installed the moistened sensor on participant’s chest as shown in its user guide. The sensor used Bluetooth Low Energy (BLE) to wirelessly transmit the heart rate and heart rate variability data.

We employed a Myo Gesture Control Armband to capture upper body extremity movements. We focused on orientation data provided by the armband’s inertial measurement unit (IMU) to track our four gestures.

We developed our exergame using Unity 2017.3.1f1 Personal 64-bit Edition considering its extensive platform support, ease of hardware integration, and our expertise with its development environment. We mainly relied on its standard asset package and modified it according to our requirements. We used C and Objective-C to create our own Bluetooth plugin for Unity due to its lack of native support.

### 3.6. Environmental Settings

We placed our display on a 150 × 60 × 70 cm desk. Participants were seated on a standard 50 cm height chair and positioned 150 cm away from the display in a front facing posture. We used two 90 × 180 cm room dividers on each side of the experiment space to avoid possible environmental distractions. The distance between left and right dividers was 200 cm. Figure 5 illustrates our configuration.

![Image](image_url)

**Figure 5.** Environmental settings.
The experimenter sat far behind participant throughout the experiment and no interruptions occurred during experiment phases. Video footage of each experiment was recorded with two video cameras. The first camera was located at the middle of the table facing the participant. The second camera was placed one meter behind the participant facing towards the experiment space.

3.7. Participant Profile

We recruited 29 participants for our experiments. Our sample consisted of 15 young and 14 old individuals with age 26.80 ± 2.62 and 72.36 ± 3.00 respectively (Table 2). Our young participants were students gathered from various labs at Nara Institute of Science and Technology, Japan. We used the services of a local organization that finds temporary or part-time jobs for senior citizens to recruit old participants. We paid a remuneration of 1000 JPY (approximately 9 USD) per hour to all individuals for their participation in our study.

Table 2. Descriptive statistics of our sample group.

<table>
<thead>
<tr>
<th></th>
<th>Male</th>
<th>Female</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>Mean</td>
<td>St.Dev.</td>
</tr>
<tr>
<td>Old</td>
<td>10</td>
<td>72.30</td>
<td>2.91</td>
</tr>
<tr>
<td>Young</td>
<td>10</td>
<td>27.10</td>
<td>3.03</td>
</tr>
<tr>
<td>Total</td>
<td>20</td>
<td>49.70</td>
<td>3.37</td>
</tr>
</tbody>
</table>

In both groups no participant identified himself as a gamer. Participants had no prior experience or knowledge about our exergame design before the scheduled experiment date. Additionally, they had no personal or professional level relationship with the experimenter that might effect the results.

4. Results

We used a fully randomized repeated measures design in order to investigate our hypotheses. We used Generalized Estimating Equations (GEE) with linear scale response model in our three step regression-analysis and preferred Wald test for Chi-square Statistics. We evaluated each FSS scale via Cronbach’s alpha and removed concentration on task at hand and transformation of time from our evaluation due to low level of internal consistency (Table 3). Additionally, one young male participant’s results were excluded from our final evaluation as a result of his low level of language proficiency and interruptions leading to differences in experiment procedure. We used difficulty adjustments and age group as factors. Each scale of FSS was evaluated according to three tests consisting of difficulty adjustment method comparison, age group comparison, and investigation of interaction effects between the two (Figure 6).

Table 3. Internal consistency of each scale according to Cronbach’s alpha.

<table>
<thead>
<tr>
<th>Scale</th>
<th>$\alpha_{const}$</th>
<th>$\alpha_{ramp}$</th>
<th>$\alpha_{perf}$</th>
<th>$\alpha_{bio}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Challenge-skill Balance</td>
<td>0.916</td>
<td>0.931</td>
<td>0.899</td>
<td>0.891</td>
</tr>
<tr>
<td>Action-awareness Merging</td>
<td>0.818</td>
<td>0.885</td>
<td>0.902</td>
<td>0.854</td>
</tr>
<tr>
<td>Clear Goals</td>
<td>0.943</td>
<td>0.947</td>
<td>0.938</td>
<td>0.874</td>
</tr>
<tr>
<td>Unambiguous Feedback</td>
<td>0.874</td>
<td>0.852</td>
<td>0.902</td>
<td>0.891</td>
</tr>
<tr>
<td>Concentration on Task at Hand</td>
<td>0.740</td>
<td>0.649</td>
<td>0.701</td>
<td>0.805</td>
</tr>
<tr>
<td>Sense of Control</td>
<td>0.883</td>
<td>0.877</td>
<td>0.869</td>
<td>0.923</td>
</tr>
<tr>
<td>Loss of Self-consciousness</td>
<td>0.828</td>
<td>0.798</td>
<td>0.825</td>
<td>0.730</td>
</tr>
<tr>
<td>Transformation of Time</td>
<td>0.506</td>
<td>0.642</td>
<td>0.728</td>
<td>0.827</td>
</tr>
<tr>
<td>Autotelic Experience</td>
<td>0.827</td>
<td>0.933</td>
<td>0.911</td>
<td>0.899</td>
</tr>
</tbody>
</table>
4.1. Challenge-Skill Balance

*Difficulty Adjustments:* The used method had a significant effect on challenge-skill balance, $p = 0.013, \chi^2 = 10.737$. Pairwise comparisons using Bonferroni correction revealed that mean scores...
were significantly lower in constant difficulty compared to performance-based and biofeedback-based difficulty adjustments ($M_{const-perf} = -0.446 \pm 0.137, p = 0.007; M_{const-bio} = 0.491 \pm 0.176, p = 0.032$).

**Age Groups:** Age group had a significant effect on challenge-skill balance, $p = 0.003, \chi^2 = 8.774$. Pairwise comparisons using Bonferroni correction revealed that mean scores were significantly lower between old individuals ($M_{old-young} = -0.732 \pm 0.247$).

**Interaction Effects:** Difficulty adjustments and age groups had a significant interaction effect on challenge-skill balance, $p < 0.001, \chi^2 = 52.292$. Pairwise comparisons using Bonferroni correction revealed that mean scores were significantly lower between old individuals under constant difficulty adjustments ($M_{const-old-const-young} = -1.18 \pm 0.307, p = 0.003$). Additionally, mean scores for constant difficulty adjustments were significantly lower than performance-based and biofeedback-based difficulty adjustments between old individuals ($M_{const-old-perf-old} = -0.661 \pm 0.178, p = 0.006; M_{const-old-bio-old} = -0.839 \pm 0.250, p = 0.022$).

### 4.2. Action-awareness Merging

**Difficulty Adjustments:** The used method had a significant effect on action-awareness merging, $p = 0.047, \chi^2 = 7.965$. Pairwise comparisons using Bonferroni correction revealed that mean scores were not significantly different.

**Age Groups:** Age group did not have a significant effect on action-awareness merging, $p = 0.304, \chi^2 = 1.059$.

**Interaction Effects:** Difficulty adjustments and age groups did not have a significant interaction effect on action-awareness merging, $p = 0.058, \chi^2 = 13.654$.

### 4.3. Clear Goals

**Difficulty Adjustments:** The used method did not have a significant effect on clear goals, $p = 0.409, \chi^2 = 2.888$.

**Age Groups:** Age group did not have a significant effect on clear goals, $p = 0.094, \chi^2 = 2.800$.

**Interaction Effects:** Difficulty adjustments and age groups had a significant interaction effect on clear goals, $p = 0.034, \chi^2 = 15.151$. Pairwise comparisons using Bonferroni correction revealed that mean scores were not significantly different.

### 4.4. Unambiguous Feedback

**Difficulty Adjustments:** The used method had a significant effect on unambiguous feedback, $p < 0.001, \chi^2 = 31.518$. Pairwise comparisons using Bonferroni correction revealed that mean scores were significantly lower in constant ($M_{const-ramp} = -0.232 \pm 0.084, p = 0.033; M_{const-perf} = -0.384 \pm 0.114, p = 0.005; M_{const-bio} = -0.670 \pm 0.122, p < 0.001$) and significantly higher in biofeedback-based difficulty adjustments ($M_{bio-ramp} = -0.438 \pm 0.092, p < 0.001; M_{bio-perf} = -0.286 \pm 0.098, p = 0.022$).

**Age Groups:** Age group had a significant effect on unambiguous feedback, $p = 0.002, \chi^2 = 9.506$. Pairwise comparisons using Bonferroni correction revealed that mean scores were significantly lower between old individuals ($M_{old-young} = -0.705 \pm 0.229$).

**Interaction Effects:** Difficulty adjustments and age groups had a significant interaction effect on unambiguous feedback, $p < 0.001, \chi^2 = 119.871$. Pairwise comparisons using Bonferroni correction revealed that mean scores were significantly lower between old individuals under constant difficulty adjustments ($M_{const-old-const-young} = -0.911 \pm 0.271, p = 0.021$). Additionally, mean scores for ramping and biofeedback-based were significantly lower than constant difficulty adjustments between older individuals ($M_{const-old-ramp-old} = -0.464 \pm 0.133, p = 0.014; M_{const-old-bio-old} = -0.768 \pm 0.215, p = 0.010$) while this changed to performance-based and biofeedback-based difficulty adjustments between young individuals ($M_{const-young-perf-young} = -0.307 \pm 0.081, p = 0.005; M_{const-young-bio-young} = -0.571 \pm 0.114, p < 0.001$).
4.5. Sense of Control

**Difficulty Adjustments:** The used method had a significant effect on sense of control, $p < 0.001, \chi^2 = 26.560$. Pairwise comparisons using Bonferroni correction revealed that mean scores were significantly lower in constant difficulty adjustments ($M_{\text{const}} - \text{ramp} = -0.393 \pm 0.123, p = 0.008; M_{\text{const}} - \text{perf} = -0.500 \pm 0.130, p = 0.001; M_{\text{const}} - \text{bio} = -0.714 \pm 0.160, p < 0.001$). Additionally mean scores were significantly lower in ramping compared to biofeedback-based difficulty adjustments ($M_{\text{ramp}} - \text{bio} = -0.321 \pm 0.080, p < 0.001$).

**Age Groups:** Age group did not have a significant effect on sense of control, $p = 0.635, \chi^2 = 0.225$.

**Interaction Effects:** Difficulty adjustments and age groups had a significant interaction effect on sense of control, $p < 0.001, \chi^2 = 90.647$. Pairwise comparisons using Bonferroni correction revealed that mean scores were significantly lower between young individuals under constant difficulty adjustments ($M_{\text{const}}_{\text{young}} - \text{perf}_{\text{young}} = -0.464 \pm 0.133, p = 0.014, M_{\text{const}}_{\text{young}} - \text{bio}_{\text{young}} = -0.786 \pm 0.128, p < 0.001; M_{\text{ramp}}_{\text{young}} - \text{bio}_{\text{young}} = -0.500 \pm 0.098, p < 0.001$).

4.6. Loss of Self-Consciousness

**Difficulty Adjustments:** The used method did not have a significant effect on loss of self-consciousness, $p = 0.435, \chi^2 = 2.732$.

**Age Groups:** Age group did not have a significant effect on loss of self-consciousness, $p = 0.723, \chi^2 = 0.125$.

**Interaction Effects:** Difficulty adjustments and age groups had a significant interaction effect on loss of self-consciousness, $p = 0.230, \chi^2 = 9.332$.

4.7. Autotelic Experience

**Difficulty Adjustments:** The used method did not have a significant effect on autotelic experience, $p = 0.109, \chi^2 = 6.048$.

**Age Groups:** Age group did not have a significant effect on autotelic experience, $p = 0.920, \chi^2 = 0.010$.

**Interaction Effects:** Difficulty adjustments and age groups did not have a significant interaction effect on autotelic experience, $p < 0.076, \chi^2 = 12.825$.

5. Discussion

5.1. General Examination of Results

Health problems associated with sedentary lifestyles are critical issues of our time. Exergames present an alternative solution to this health crisis through simultaneous introduction of entertainment and physical activity. However, effects of aging on game difficulty optimization remains unclear. Identification of optimization method preference between age groups will facilitate better game experiences and may improve adherence to regular physical exercise through well-designed exergames.

In this study, we compared the exergame experience of young and old individuals under four frequently used difficulty adjustment methods to better understand differences between their preferences. Through our analysis, we observed a significant decline in multiple aspects of our sample’s play experience under constant difficulty adjustments. On the other hand, biofeedback-based difficulty adjustments resulted in higher scores compared to other methods and led to better play experiences. Our second analysis showed that perceived balance between presented challenge and player skill was higher between young individuals compared to old. Additionally, ideas on quality of the received feedback differed significantly between age groups. This result provides necessary support for our second hypothesis. Finally, we found interaction effects between used difficulty adjustments methods and player age groups. Third analysis identified biofeedback-based difficulty adjustments as
preferable in both age groups in addition to performance-based between young and ramping difficulty adjustments between old individuals. As anticipated, this result proves our third hypothesis.

Our findings correlate favorably with Baldwin et al. [46,62] and further support the role of dynamic difficulty adjustments in delivering satisfactory play experiences. The evidence we found also correlates favorably with Liu et al. [50] and supports our first hypothesis.

Mean scores of old individuals were consistently lower compared to young individuals except action-awareness merging (Figure 6). Although the factual reasons behind this difference was not explored, we think this might be connected to aging related decline in sensory acuity between older individuals. Under constant difficulty adjustments the frequency of movements follow a steady pattern whereas other difficulty adjustment methods lead to changes that might be difficult to anticipate for older individuals.

The superiority of biofeedback-based difficulty adjustments over traditional methods is fairly clear. However, integration of biofeedback-based or another dynamic difficulty adjustment method into gameplay significantly increases the required engineering effort and introduces additional hardware requirements for consistent monitoring of player state. Despite its benefits, biofeedback-based difficulty adjustments might not be favorable for game designers or players alike and lead to utilization of ramping or performance-based difficulty adjustment mechanisms in exergames. Our study highlights the age based differences in player preferences on these mechanisms and suggests performance-based approaches when the target user group is young individuals. On the other hand, our results favor ramping difficulty adjustments in exergames designed for old individuals.

However, careful attention must also be paid to play and player characteristics. It is very likely that players with a high self-efficacy may enjoy consistently difficult play experiences. Additionally, despite the widespread applications of Csikszentmihalyi’s flow model [71] to exercise, consistent physical improvement may not be observed while challenge and skill consistently stay in complete balance and show no variation. This outcome may cause a decline in perceived enjoyment and benefits of exergames, eventually leading to reduced participation compared to regular physical exercise. We strongly believe that a combination of multiple difficulty adjustment methods based on administered exercise and player characteristics would lead to superior play experiences while simultaneously contributing to healthy lifestyles.

Given the age related changes in the cardiovascular system, our anticipation was to see more variation in heart rate measurements of older participants compared to young participants. This was the case in performance-based, ramping, and biofeedback-based methods, but we observed less variation in measurement values of old participants during scenarios with constant difficulty adjustments (Figure A1). A possible explanation for this result might be found in the affective state of players. Participants showed a tendency to stop performing gestures under performance-based and ramping difficulty adjustment methods when the difficulty level became significantly high. However, this was not the case in constant difficulty since the rhythmicity and predictability of movement patterns allowed them to perform for longer periods of time. This is an important factor since proper exercise routines are designed around reps and sets; referring to number of consecutive repetitions of an exercise and number of cycles where the repetitions are performed. Although dynamic difficulty adjustments may lead to superior gaming experiences, they could also demonstrate a decline in exercise gains and health benefits by introducing randomness to exercise routines performed by players. We think this is a fundamental issue for future research and requires careful investigations into sports science.

5.2. Post-Experiment Discussions

In addition to above findings, older participants also mentioned the metaphorical relationship between aircraft marshalling gestures and their in-game representation. This observation has serious implications in integrating logical design elements to exergames and eliminating the chocolate covered broccoli problem. One possible approach to further expand this relationship is to introduce narratives
to exergames. Ever increasing popularity of action role-playing video games (RPG) with exceptional narratives such as Witcher III and Elder Scrolls is a factual example of how good stories might increase motivation to play. Lu examined this idea to bridge the gap between these two important game design domains [73]. Yet, further studies on this topic are required in order to verify its effects and possible benefits.

Despite this, we observed several cases where upper extremity movements of participants deviated from their intended path. Most of our participants believed their movements were exact and as straight as possible, yet when we showed them the videos we captured the apparent error in their gestures was undeniable. It is probable that participants may have erroneously judged their own joint position which can be justified by poor proprioception. In our opinion, there is a strong probability that incorporation of proprioceptive feedback on extremity movements may lead to enhanced awareness of the body in space and increase the likelihood of experiencing flow in exergames.

5.3. Deployment at a Rehabilitation Hospital

According to social cognitive theory, people observe others and evaluate the consequences of their behaviors, which can also lead to engagement of the viewer [74]. This theory has a strong association with multiplayer gaming where several players compete or cooperate depending on game objective and its structure. It also lays the groundwork for considerable amount of literature and makes social interaction an important factor within exergames to promote physical and psychological health of aging generations [17,44,45]. Systematic studies also report the positive influence of social interaction on increasing exercise adherence among old individuals [75,76].

Therefore, we implemented a multiplayer version of our game to further our investigations and examine a potential future research direction. We deployed it at a rehabilitation hospital for institutionalized older individuals where we tested it for four days (Figure 7). The majority of our participants during this period were older than 85 years old, maximum being 91. The overall response to our exergame was quite positive, but the difficulty adjustment methods worked properly with very few participants. It is crucial to note that the task difficulty in exergames does not fully depend on gesture frequency, but also required range of motion for their correct execution and player’s cognitive abilities to interpret displayed instructions. More than half of our participants had health issues with their shoulders that prevented them from raising theirs arms to a point where the angle between arms and trunk reaches 120 degrees, which was necessary for our system to separate different movements. This observation shows that difficulty adjustments should consider not only the frequency of actions, but also the range of motion involved in it, and we hope our research will serve as a base for future studies on this topic.

The oldest participant was also the healthiest in terms of resting heart rate, maximum heart rate variability, and range of motion. Our 91 year-old female participant voluntarily joined our every experiment and demonstrated a consistent increase in game performance with each day. This was the most surprising result to emerge from our efforts at the hospital since it was interesting for us to observe that the rate of physical and cognitive decline can greatly vary between individuals despite their identical age. Although getting older is inevitable, this has serious implications on importance of aging well to sustain a high quality of life.
Figure 7. We extended our game by adding multiplayer functionality to utilize perceived social interaction of the players. Furthermore, we deployed this version at a rehabilitation hospital for institutionalized older individuals.

5.4. Limitations

We are aware that our study has three limitations. The first is the homogeneous structure of our old individual group in terms of nationality. Although we managed to gather young individuals with different ethnic backgrounds, that was not the case for old individuals. Combined with their similar age bracket, same ethnicity may have an effect on their gaming perception and lead to an undesired bias in our results. We are planning to address this issue with a replication study outside of Japan in the future. The second limitation of our study is its short-term repeated measures design. It may be difficult for participants to realize and assess the differences in all dimensions of flow under various difficulty adjustment mechanisms with a study of this design. Yet, this limitation underlines the difficulty of collecting data on elderly. Given that, it is not inconceivable that different evaluations would have led to different conclusions. Finally, although the exergame we have designed for this study considerably adheres to recommendations of the ElderGAMES [68], one can argue that it is a mere simulation and games designed with a superior engineering effort may lead to different result. We acknowledge this limitation and encourage future studies with different games and game genres for further investigation.

6. Conclusions

In this study, we have compared multiple dimensions of flow between young and old individuals under four common approaches to exergame difficulty adjustments. Our work has led us to the conclusion that difficulty adjustment methods lead to a different exergaming experience based on the age of players. The evidence from this study points towards the idea that likelihood of experiencing flow differs between young and old individuals under same difficulty adjustment methods. Additionally, although biofeedback-based adjustments are preferable, old individuals are also expected to experience flow with ramping adjustments whereas performance-based methods are favorable for young individuals. Taken together, these findings implicate a role for age while choosing or designing difficulty adjustment methods in exergames. We think our research direction is promising in solving several problems attributed to low adherence rates to physical exercise through exergames and contribute to establishment of healthy lifestyles.

Despite its limitations, this work has many implications for research into difficulty adjustments in exergames. To further our research, we plan to have a long term study in an elderly care center and conduct several replication studies in other countries through the help of our collaborators. We are also currently in the process of investigating high intensity aerobic exercises and massively multiplayer online gaming applications as promising application areas.

Funding: This work was supported by the JSPS Program for Fostering Globally Talented Researchers and the MIC/SCOPE #162107006.

Acknowledgments: We thank Nara Rehabilitation Hospital and Nara City Senior Human Resource Center for their support and contributions.

Conflicts of Interest: The authors declare no conflict of interest. None of the authors of this paper has financial or personal relationship with other people organizations that could inappropriately influence or bias the contents of manuscript in discussion.

Abbreviations
The following abbreviations are used in this manuscript:

JSPS Japan Society for the Promotion of Science
MIC Ministry of Internal Affairs and Communications
SCOPE Strategic Information and Communications R&D Promotion Programme
DDA Dynamic Difficulty Adjustments
FSS Flow State Scale
HRV Heart Rate Variability
Const Constant
Ramp Ramping
Perf Performance-based
Bio Biofeedback-based

Appendix A

Figure A1. Histograms representing the captured heart rate and heart rate variability data through Polar H7.
Figure A2. Captured heart rate and heart rate variability data of an old and a young participant.

References


73. Lu, A.S. Narrative in exergames: Thoughts on procedure, mechanism, and others. Games Health J. 2015, 4, 19–24. [CrossRef] [PubMed]