Development of a Search Task Using Immersive Virtual Reality: Proof-of-Concept Study

Abstract

Background: Serious games are gaining increasing importance in neurorehabilitation since they increase motivation and adherence to therapy, thereby potentially improving its outcome. The benefits of serious games, such as the possibility to implement adaptive feedback and the calculation of comparable performance measures, can be even further improved by using immersive virtual reality (iVR), allowing a more intuitive interaction with training devices and higher ecological validity.

Objective: This study aimed to develop a visual search task embedded in a serious game setting for iVR, including self-adapting difficulty scaling, thus being able to adjust to the needs and ability levels of different groups of individuals.

Methods: In a two-step process, a serious game in iVR (bird search task) was developed and tested in healthy young (n=21) and elderly (n=23) participants and in a group of patients with impaired visual exploration behavior (ie, patients with hemispatial neglect after right-hemispheric stroke; n=11). Usability, side effects, game experience, immersion, and presence of the iVR serious game were assessed by validated questionnaires. Moreover, in the group of stroke patients, the performance in the iVR serious game was also considered with respect to hemispatial neglect severity, as assessed by established objective hemispatial neglect measures.

Results: In all 3 groups, reported usability of the iVR serious game was above 4.5 (on a Likert scale with scores ranging from 1 to 5) and reported side effects were infrequent and of low intensity (below 1.5 on a Likert scale with scores ranging from 1 to 4). All 3 groups equally judged the iVR serious game as highly motivating and entertaining. Performance in the game (in terms of mean search time) showed a lateralized increase in search time in patients with hemispatial neglect that varied strongly as a function of objective hemispatial neglect severity.

Conclusions: The developed iVR serious game, “bird search task,” was a motivating, entertaining, and immersive task, which can, due to its adaptive difficulty scaling, adjust and be played by different populations with different levels of skills, including individuals with cognitive impairments. As a complementary finding, it seems that performance in the game is able to capture typical patterns of impaired visual exploration behavior in hemispatial neglect, as there is a high correlation between performance and neglect severity as assessed with a cancellation task.

JMIR Serious Games 2021;9(3):e29182 doi: 10.2196/29182

KEYWORDS

virtual reality; serious game; search task; stroke; neglect; usability; development; immersion; concept; gaming
Introduction

A crucial component of neurorehabilitation and its success is therapy adherence and repetition. One promising possibility to improve patients’ therapy adherence, and thereby its outcome, is to enhance training motivation by means of serious games [1] and immersive virtual reality (iVR) [2,3]. The combination of serious games and iVR allows for the investigation of visual exploration behavior, which is highly relevant in activities of daily living and therefore frequently used in the diagnosis and rehabilitation of patients suffering from a neurodegenerative disease [4-7].

The primary purpose of serious games is not to be fun, but to teach, train, or assess skills in an entertaining way [8]. Due to technological achievements in recent years, their importance in education [9] rehabilitation [10], and medical training [11] is growing, as they have several advantages. First, serious games are standardized, which means that each user will experience the same task. Second, different game settings offer the possibility to automatically adapt the task difficulty to the user’s individual skills and performance level. Finally, serious games offer the possibility to track the user’s game performance (e.g., achieved scores) and game behavior as a measure for activities of daily living (e.g., reaction times) [12].

The different settings and measured performance values have been used to implement adaptive difficulty scaling, which has shown to be an important element of serious games leading to 2 key elements [13-16]. First, a balance point (i.e., when game difficulty is still challenging, yet does not exceed the player’s abilities) can be achieved. Second, changes in game mechanics should allow adapting the difficulty throughout the game, to continuously and optimally match the increasing skill level of the player [17]. Importantly, both elements are essential to keep the player motivated [18].

The benefits of serious games can be further improved by the use of iVR [19-21]. iVR presents computer-generated artificial, but interactable (i.e., hand-held controllers), 360° environments or pre-rendered 360° videos inside a head-mounted display (HMD). With technological improvements over the last decade, iVR is now also increasingly used in clinical applications such as in motor rehabilitation for gait and balance [22,23], surgery training [24-26], or anxiety treatment [27]. Particularly in tasks that involve any kind of motor activities, iVR has some advantages, such as the possibility to objectively measure progressive improvement in trained skills, perform task-oriented repetitive training, and apply multisensory feedback and task variation [28,29]. Rizzo et al [12] summarized evidence showing that skills gained in the iVR environment can be transferred to activities of daily living (e.g., crossing the street [30]) reflecting the ecological validity of tasks in iVR. The high ecological validity can be explained due to fewer distractions from external stimuli and the intuitive interaction with the virtual environment, whereas the interaction and thus behavior in the virtual environment can be tracked by recording head and hand movements. Conclusively, tasks in iVR tend to feel more naturalistic, and several studies have shown that the naturalistic feeling of a task correlates with higher enjoyment, better performance, and better motivation [3,31,32]. This naturalistic feeling is created by the so-called immersion [33-35] (i.e., a situation in which the real world is ignored in favor of the virtual environment [36]).

Visual exploration behavior is a crucial element of activities of daily living (e.g., crossing the street, grocery shopping) [37] and corresponds to purposefully looking around in the present environment (i.e., actively acquiring visual information through coordinated movements of the eyes and head [38]). Therefore, an impaired visual exploration behavior could lead to a reduction of performance in activities of daily living and thus in quality of life [39,40]. Impairment of the visual search behavior is also one landmark of patients suffering from hemispatial neglect [37,41]. Hemispatial neglect is a visuospatial attention disorder that frequently occurs after a right hemispheric stroke. Its characteristic is the inability to attend or respond to stimuli presented within the left contralesional space [42].

Therefore, the aim of this proof-of-concept study was to develop a serious game using iVR in which participants perform a visual search task that encourages the exploration of their environment. We hypothesized that the given task has high usability and limited or no side effects and that the performance can be adapted dynamically to the skills of the participants.

Methods

The main goal of this study was to develop a gamified search task that encourages players to explore their visual environment. We named it the bird search task, and development and evaluation were divided into 2 steps.

First, the 2D game “Crazy Chicken” (ak tronic Software & Services gmbh, Saerbeck, Germany) was used as inspiration, and the game mechanics were transferred to a 3D iVR environment and then modified and tested with healthy young and elderly participants. Based on the findings of the first step, in the second step, the task was further adapted and tested with patients with hemispatial neglect after a right hemispheric stroke.

Game Development and Apparatus

Bird Search Task for Healthy Participants

The 2D game “Crazy Chicken” was identified as a suitable gamified task because it encourages players to explore their visual environment. In the original “Crazy Chicken” game, the chickens (visual targets) appear at random locations on the computer screen and fly at a constant velocity in random directions. The player, by constantly scrolling left and right on a scrollable, 2D screen, has to search and tag the targets before they disappear after a constant time delay.

To transfer the task into 3D, a simple virtual environment was designed using the gaming development platform Unity3D [43]. The environment was built as a wide circular area surrounded by trees and mountains, as shown in Figure 1 and Figure 2. A mobile gaming laptop was used to render the virtual environment (HP-Omen, graphic-card NVIDIA GTX1050 and CPU Intel i7).
Figure 1. Immersive virtual reality game, with (A) the participant’s field of view (yellow) within the head-mounted display, which moved if the player (blue) turned his or her head, and the area where the target could appear (dashed line; spawn area), which was locked to the midsagittal plane; and (B) a schematic representation of a participant wearing the head-mounted display and performing the task.

Figure 2. Exemplary scene of the gameplay, (A) as implemented in step 1 (healthy individuals) and (B) after the modifications performed in step 2 (patients with right hemispheric stroke and hemispatial neglect).

The iVR hardware consisted of an HMD and a hand-held controller (HTC Vive, High Tech Computer Corporation, Taoyuan, Taiwan; resolution of 2160 x 1200 pixels; full HD; horizontal and vertical field of view of 110 degrees [yellow area in Figure 1A]; frame rate of 90 Hz). The x-y position of the handheld controller was continuously recorded. Using iVR hardware, 360° of the virtual environment could be explored by moving the eyes and turning the head. The targets (eg, birds) could be tagged by aligning the handheld controller with the target and simultaneously pressing a trigger on the controller. Target appearance was set to randomly take place within a restricted area in front of the player (±60° horizontally and ±50° vertically, as defined with respect to the trunk’s midsagittal plane; see dashed lines in Figure 1A). In order to promote exploration of the whole extent of the virtual environment, the horizontal area in which the targets could appear was 10° larger than the player's field of view (ie, additional 5° on either side; Figure 1A).

To alert the player of a new appearing target, a short (1 second) auditory signal (chicken cackle) was presented binaurally via headphones (XQISIT oE400, Strax Americas Inc, Miami, FL). Then, the target appeared at a position randomly determined within the spawn area and flew horizontally towards the right or left (direction randomly selected) with a constant velocity of 2°/second. If the target was detected and successfully tagged, it fell vertically and disappeared. In case the target was not tagged within the maximum presentation time of 15 seconds, it disappeared. In either way, after a fixed interstimulus interval of 2 seconds, a new target appeared, again with an alerting cackle.

The game was played in rounds. Each of the 10 played rounds consisted of 30 targets. For each target, the time until it was found was measured, and the percentage of found targets was calculated for each round.

We modified the task further by integrating several difficulty levels, to be able to adapt the task to participants with different needs and skills.

We implemented 15 difficulty levels, in which changes in difficulty were achieved by manipulating both target behavior and task mechanics. For each of the 15 levels, the values of the settings changed stepwise, according to the “Change Per Level” threshold presented in Table 1. Concerning target behavior, both the maximum lifetime and speed of the targets were manipulated across levels. For instance, in the easiest level (Level 1), the targets moved with a speed of 2°/second and were presented for 15 seconds before they disappeared; in the most difficult level (Level 15), the targets moved with a speed of 35°/second and were presented for 4 seconds before they disappeared. Hence, the more difficult the level, the faster the participants had to explore the visual environment in order to find the targets before they disappeared. Concerning task mechanics, the task adapted its difficulty (ie, changed the difficulty level) automatically, based on performance in the previous round. Each level had a level-up threshold (ie, if the threshold value was reached, the next round started in a higher difficulty level) and a level-down threshold (ie, if the threshold
value was not reached, the next round started in a lower difficulty level), defined according to the percentage of found targets. If the percentage of found targets in a particular round was not higher than the level-up threshold and not lower than the level-down threshold, the difficulty in the next round did not change.

For example, in a round at level 6, the targets move with a speed of 13.8 °/seconds, and the maximum lifetime of the targets is 11.07 seconds. Assuming that the participant would find 25 of the 30 targets in this particular round (ie, 83.3% found targets), then the next round would start at level 7 since the percentage of found targets is above the level-up threshold of level 6 (ie, 77.1%). If the player finds only 15 of the 30 targets (ie, 50% found targets), then the next round would start at level 5, since the percentage of found targets would be below the level-down threshold of level 6 (ie, 67.1%).

**Table 1. Change per level algorithm.**

<table>
<thead>
<tr>
<th>Task and target parameters</th>
<th>Level 1</th>
<th>Change per level$^a$</th>
<th>Level 15</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lifetime (seconds)</td>
<td>15</td>
<td>-0.786</td>
<td>4</td>
</tr>
<tr>
<td>Speed (°/second)</td>
<td>2</td>
<td>+2.36</td>
<td>35</td>
</tr>
<tr>
<td>Level-up threshold (%)</td>
<td>&gt;70</td>
<td>+1.42</td>
<td>&gt;90</td>
</tr>
<tr>
<td>Level-down threshold (%)</td>
<td>&lt;60</td>
<td>+1.42</td>
<td>&lt;80</td>
</tr>
</tbody>
</table>

$^a$The stepwise change per level between Level 1 and Level 15.

**Bird Search Task for Patients With Hemispatial Neglect**

In a second step, based on the findings collected in healthy participants and on recommendations of the clinicians, the version for patients was developed. As neglect's clinical picture is very heterogeneous across patients, the task needs to be easily adaptable to the patients' individual needs. The aim was that patients would not get overwhelmed and frustrated due to too great a difficulty but also not get bored by a task that was too easy. For this purpose, we implemented an easy-to-use setting file, in which the total number of targets, number of targets per round, and starting difficulty level could be set individually for each patient. Additionally, several adaptations in design and task mechanics were performed. First, as patients with neglect process visual information slower [44,45], the landscape of the gameplay would be too complex; therefore, in order to reduce the number of distractive elements, the trees in the background and the clouds in the sky were removed from the game scenery (see Figure 2). Second, based on the recommendations of clinicians, the minimal speed of the targets was lowered from 2 °/second to 0.1 °/second; this manipulation generally lowered the difficulty level as well as reduced the need for fast head movements that could promote side effects.

**Participants**

The study was approved by the Ethics Committee of the Cantons of Bern and the Ethics Committee of north-west and central Switzerland and was conducted in accordance with the latest version of the Declaration of Helsinki. All participants gave written informed consent before participation.

In step one, the feasibility of the task was assessed with 21 younger healthy participants recruited at the University of Bern (10 women; mean age, 28.1, SD 5.5 years) and with 23 older healthy participants recruited during a chess tournament for seniors (1 woman; mean age, 71.3, SD 6.3 years). All participants had no history of neurological nor psychiatric disorders. Previous VR experience was reported by 16 of the younger participants and 1 of the older participants.

In step two, 11 inpatients with hemispatial neglect (5 women; mean age, 69.6, SD 13.0 years) after right hemispheric, subacute stroke were recruited at the Neurorehabilitation Clinics of the Inselspital, Bern University Hospital (sites Bern and Riggisberg) and of the Kantonsspital Luzern, Switzerland. Demographic characteristics for each patient are presented in Table 2. The study was always conducted with a mobile setup in the place where the patient was currently hospitalized. All patients showed significant left-sided neglect in activities of daily living, as assessed with the Catherine Bergego Scale (CBS, Range 0-30, 0 = normal) [46], and had normal or corrected-to-normal vision. One of the patients reported previous VR experience.
Table 2. Individual demographical and neuropsychological data from step 2 for patients with hemispatial neglect.

<table>
<thead>
<tr>
<th>Patient code</th>
<th>Age range (years)</th>
<th>Gender</th>
<th>Lesion type</th>
<th>Time since stroke (days)</th>
<th>CBSasts</th>
<th>CoCb (SNTc single)</th>
<th>Number of played chickens</th>
<th>Play duration (min)</th>
</tr>
</thead>
<tbody>
<tr>
<td>P_26</td>
<td>70-75</td>
<td>Male</td>
<td>Ischemic</td>
<td>85</td>
<td>6</td>
<td>0.184</td>
<td>195</td>
<td>16.5</td>
</tr>
<tr>
<td>P_27</td>
<td>55-60</td>
<td>Male</td>
<td>Hemorrhagic</td>
<td>145</td>
<td>2</td>
<td>0.208</td>
<td>100</td>
<td>10.4</td>
</tr>
<tr>
<td>P_28</td>
<td>80-85</td>
<td>Male</td>
<td>Bleeding</td>
<td>49</td>
<td>8</td>
<td>0.746</td>
<td>50</td>
<td>9.4</td>
</tr>
<tr>
<td>P_29</td>
<td>80-85</td>
<td>Female</td>
<td>Ischemic</td>
<td>52</td>
<td>9</td>
<td>0.998</td>
<td>44</td>
<td>8.1</td>
</tr>
<tr>
<td>P_30</td>
<td>75-80</td>
<td>Male</td>
<td>Ischemic</td>
<td>42</td>
<td>8</td>
<td>0.293</td>
<td>80</td>
<td>9.2</td>
</tr>
<tr>
<td>P_31</td>
<td>50-55</td>
<td>Female</td>
<td>Ischemic</td>
<td>106</td>
<td>3</td>
<td>0.414</td>
<td>80</td>
<td>13.2</td>
</tr>
<tr>
<td>P_32</td>
<td>65-70</td>
<td>Female</td>
<td>Ischemic</td>
<td>59</td>
<td>2</td>
<td>–0.067</td>
<td>80</td>
<td>10.2</td>
</tr>
<tr>
<td>P_33</td>
<td>70-75</td>
<td>Female</td>
<td>Ischemic</td>
<td>58</td>
<td>3</td>
<td>0.824</td>
<td>80</td>
<td>9.3</td>
</tr>
<tr>
<td>P_34</td>
<td>85-90</td>
<td>Male</td>
<td>Ischemic</td>
<td>29</td>
<td>18</td>
<td>0.998</td>
<td>80</td>
<td>8.4</td>
</tr>
<tr>
<td>P_35</td>
<td>50-55</td>
<td>Male</td>
<td>Hemorrhagic</td>
<td>39</td>
<td>10</td>
<td>0.046</td>
<td>150</td>
<td>13.0</td>
</tr>
<tr>
<td>P_36</td>
<td>60-65</td>
<td>Female</td>
<td>Ischemic</td>
<td>38</td>
<td>17</td>
<td>0.191</td>
<td>100</td>
<td>10.1</td>
</tr>
</tbody>
</table>

aCBS: Catherine Bergego Scale (0-30).
bCoC: Center of Cancellation (−1 to 1).
cSNT: Sensitive Neglect Test.

In addition to CBS, where a value ≥1 means the patient has a neglect, the objective neglect severity was assessed on the day of the task by means of the paper-and-pencil Sensitive Neglect Task (SNT), single version [47]. The SNT is a cancellation task in which patients are asked to mark 40 targets among 240 distractors. Based on the distribution of the marked targets, the Center of Cancellation (CoC) [48] was computed, representing an objective measure of neglect severity. The CoC reflects the mean deviation of the marked targets from the center and is normalized to values ranging from –1 to 1. Zero indicates no spatial bias, where a CoC ≥0.081 represents a significant rightward shift (ie, left-sided neglect).

Outcome Measures

Questionnaires

To evaluate the feasibility and usability of the newly implemented task, several questionnaires were used to assess the participants’ individual gaming experience.

To assess acceptance, usability, and participant’s perception of the visual search task and of the VR system, 3 questions from the System Usability Scale (SUS) [49] were used, as previously reported by Gerber et al [50] and Knobel et al [51]. The questions were answered using a 5-point Likert-scale, ranging from “fully disagree” to “fully agree.” The mean score across all questions was calculated for each participant.

The Simulator Sickness Questionnaire (SSQ) [52] was used to assess side effects such as cybersickness, oculomotor problems, and disorientation [53,54]. In order to reduce the workload of the study participants, a subset of 7 questions from the SSQ was used, as previously reported by Gerber et al [50] and Knobel et al [51]. The questions were answered using a 4-point Likert-scale (ie, “None,” “Slight,” “Moderate,” “Severe”). Again, the mean score across all questions was calculated for each participant.

Additionally, in order to assess the enjoyment of the task, the Perception of Game Training Questionnaire (PGTQ) [55] was administered. The PGTQ consists of 4 questions that are answered using a 7-point Likert-scale, ranging from “fully disagree” to “fully agree.” Each question represented a different aspect of the perception of the task (ie, motivation, frustration, how challenging it is, entertainment). Therefore, no mean score was calculated across questions; instead, each score was considered independently.

Furthermore, to assess immersion and presence, questions from the Igroup Presence Questionnaire (IPQ) [56] were used. The question subset was already used in another iVR study by Gerber et al [50] and was answered on a 5-point Likert-scale.

The questions from each questionnaire are shown in Table 3.
Table 3. Exact formulations that were asked in the questionnaires.

<table>
<thead>
<tr>
<th>Questionnaire, number</th>
<th>Question</th>
<th>Domain</th>
</tr>
</thead>
<tbody>
<tr>
<td>SUS&lt;sup&gt;a,b&lt;/sup&gt;</td>
<td>I thought the system was easy to use</td>
<td>Usability</td>
</tr>
<tr>
<td>1</td>
<td>I think that I would like to use this system</td>
<td></td>
</tr>
<tr>
<td></td>
<td>frequently</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>I felt very confident using the system</td>
<td>Usability</td>
</tr>
<tr>
<td>SSQ&lt;sup&gt;c,d&lt;/sup&gt;</td>
<td>General discomfort</td>
<td>Sickness</td>
</tr>
<tr>
<td>4</td>
<td>Stomach awareness</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Sweating</td>
<td>Sickness</td>
</tr>
<tr>
<td>6</td>
<td>Nausea</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>Headache</td>
<td>Oculomotor problems</td>
</tr>
<tr>
<td>8</td>
<td>Eye strain</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>Dizziness</td>
<td>Disorientation</td>
</tr>
<tr>
<td>IPQ&lt;sup&gt;e,f&lt;/sup&gt;</td>
<td>In the virtual world, I had a sense of “being</td>
<td>Immersion</td>
</tr>
<tr>
<td>11</td>
<td>there.”&lt;sup&gt;g&lt;/sup&gt;</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>Somehow, I felt that the virtual world</td>
<td>Presence</td>
</tr>
<tr>
<td></td>
<td>surrounded me.</td>
<td></td>
</tr>
<tr>
<td>PGTQ&lt;sup&gt;i,j&lt;/sup&gt;</td>
<td>I was motivated for a good performance.</td>
<td>Motivation</td>
</tr>
<tr>
<td>13</td>
<td>The game was frustrating.</td>
<td>Frustration</td>
</tr>
<tr>
<td>14</td>
<td>The game was challenging.</td>
<td>Challenge</td>
</tr>
<tr>
<td>15</td>
<td>The game was entertaining.</td>
<td>Entertainment</td>
</tr>
<tr>
<td>16</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<sup>a</sup>SUS: System Usability Scale.
<sup>b</sup>Fully disagree to fully agree; score range, 1-5; midpoint, 3; scored as the mean of Q1-Q3.
<sup>c</sup>SSQ: Simulator Sickness Questionnaire.
<sup>d</sup>None to severe; score range, 1-4; midpoint, 2.5; scored as the mean of Q4-Q10.
<sup>e</sup>IPQ: Igroup Presence Questionnaire.
<sup>f</sup>Score range, 1-5; midpoint, 3; each question is scored individually.
<sup>g</sup>Not at all to very much.
<sup>h</sup>Fully disagree to fully agree.
<sup>i</sup>PGTQ: Perception of Game Training Questionnaire.
<sup>j</sup>Fully disagree to fully agree; score range, 1-7; midpoint, 4.5; each question is scored individually.

**Performance Indicators During the Task**

The presented task allowed us to measure several performance indicators, namely the changes in difficulty levels over time and the mean search time to detect the targets.

Based on the mean search time of the targets, the mean search time per participant was calculated, representing the mean time until a participant tagged a target (not-found targets were excluded). Moreover, the controller position of the VR setup was recorded over the entire task, which allowed us to track the participants’ hand positions over time, hence delivering information concerning their spatial search behavior.

**Evaluation of the Difficulty Scaling**

There are different possibilities to assess whether an adaptive difficulty scaling is successful. An easy indirect, but less objective, possibility is to simply evaluate the results of the questionnaires concerning entertainment and frustration. A more elaborate and objective approach is to consider the number of level changes over the task rounds with respect to the starting level. A population in which the starting level is much easier than the average balance point should show a greater increase in difficulty (ie, more upward level switches) in the initial phases of the task (ie, when they increase to their balance point) and then a smaller increase over time (ie, when the balance point is reached, but the participants still gradually get better at the task due to practice). The better the starting level matches the abilities of the group, the smaller the difficulty changes should be in the initial phases of the task. Nevertheless, due to practice (eg, better aiming, better search strategies), participants are expected to get better in the task, and there should thus be at least a small difficulty level increase over time.
A plateau reflects the balance point at which the level did not change anymore between rounds. This indicates that participants reached their optimal task difficulty level (i.e., their performance was not at the ceiling in a particular difficulty level so that the algorithm would increase it in the next round and not at the floor that the algorithm would decrease the difficulty level in the next round).

**Statistical Analyses**

The mean SUS and SSQ scores, reflecting the usability and number of side effects, respectively, were computed for each group (i.e., young, elderly, and stroke). The 3 and 7 items of the SUS and SSQ, respectively, were averaged.

The means of the groups for the PGTQ and the IPQ questions were calculated per question and were displayed as histograms.

The change in level for every participant was computed by subtracting the starting level from the levels they were in the consecutive rounds. This change of level was used as a performance measure and visualized as the level change over time relative to the starting level. The position over time of the hand was illustrated by plotting the controller x-y position over time.

The difference in search time between the young group and elderly group was calculated using a 2-sided $t$ test. For the neglect group, a Pearson correlation was calculated between neglect severity (CoC in stroke group) and the mean search time of targets in the VR game. For the statistical analysis, the alpha was set to 0.05.

All analyses and visualizations were performed with R [57] and MATLAB [58].

**Results**

**Questionnaires**

In step one, the feasibility and usability of the task were assessed in 2 groups: young and elderly healthy participants. The analysis of the SUS scores (mean of the 3 questions; see Table 3) revealed high usability in both groups (Table 4). Both young and elderly participants reported that they would even like to play the task frequently (mean scores: young, 3.95, SD 0.74; elderly, 4.04, SD 1.26).

In general, almost no side effects were reported, as assessed with the SSQ score (mean of the 7 domains; see Table 3). In the young group, the mean score reflected minimal side effects (Table 4). More precisely, only 3 young participants reported severe side effects in 1 of the 7 domains (2 cases of stomach awareness and 1 case of dizziness). In the elderly group, the mean score was similarly low (Table 4). Only 1 elderly participant reported a severe side effect (sweating).

The PGTQ consisted of 4 questions (for details, see Table 3) concerning motivation, frustration, challenge, and entertainment. The participants of both groups were very motivated (mean scores: young, 6.38, SD 1.12; elderly, 6.73, SD 0.46), entertained (mean scores: young, 6.28, SD 0.56; elderly, 5.83, SD 1.70), and not frustrated (mean scores: young, 2.57, SD 0.90; elderly, 1.83, SD 1.44). The extent to which the task was challenging was rated above midline in both groups (mean scores: young, 4.29, SD 0.90; elderly, 5.61, SD 1.23).

The question for immersion (mean scores: young, 3.81, SD 0.75; elderly, 3.96, SD 1.02) and presence (mean scores: young, 4.24, SD 0.70; elderly, 4.48, SD 0.59) were rated high in both healthy groups (Figure 3).

In step two, the feasibility of the task was assessed in patients with neglect. Patients with neglect rated the task as highly usable (Table 4), and almost no side effects were reported on the SSQ (Table 4). More precisely, none of the patients reported any severe side effect. Only 3 patients reported side effects. One patient reported moderate sweating, while 2 others reported mild sweating and mild headache.

According to the PGTQ scores (Figure 3), the patients were highly motivated (mean 6.18, SD 1.17), entertained (mean 6.27, SD 0.79), and not frustrated (mean 1.91, SD 0.70). Furthermore, the degree of challenging score was around the midline (mean score: 3.82, SD 1.33; i.e., most patients rated the task as neither too difficult nor too easy). The patients felt high immersion (mean 3.64, SD 1.43) and presence (mean 4.09, SD 0.83).

### Table 4. Results of the System Usability Scale (SUS), where 1 means “unusable” and 5 means “very usable,” and the Simulator Sickness Questionnaire (SSQ), where 1 means “None” and 4 means “Severe” side effects.

<table>
<thead>
<tr>
<th>Study group</th>
<th>SUS, mean (SD)</th>
<th>SSQ, mean (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Young</td>
<td>4.41 (0.49)</td>
<td>1.42 (0.45)</td>
</tr>
<tr>
<td>Elderly</td>
<td>4.46 (0.73)</td>
<td>1.27 (0.24)</td>
</tr>
<tr>
<td>Stroke</td>
<td>4.73 (0.44)</td>
<td>1.05 (0.10)</td>
</tr>
</tbody>
</table>
Adaptive Difficulty Scaling

In all 3 groups, the qualitative illustration of the increase in the level differences shows a plateau over time (Figure 4).

In young participants, a steep increase in the level difference (i.e., the levels get more difficult) was found at the beginning of the task, reaching a plateau (optimum difficulty) at a level that was close to the possible maximum of 7 increases. There were only 7 increases possible because they started at level 8, meaning they could level up to 15 by increasing their level 7 times.

In elderly participants and stroke patients, the increase in the level difference revealed a more moderate and heterogeneous change over time. Both groups needed more rounds to reach their optimum; compared visually, this optimum was lower than in the young group.

Figure 4. Level difference (and standard error) in the number of levels over the round compared to the starting level. In the neglect-group for the bars, only the 6 patients that played for 10 rounds were included. The raw values are included for the other 3 patients. The level difference is the mean of the difference between the level in round x minus the starting level.

In-Game Measures

Controller Position Over Time

During the task, the controller position was continuously recorded and could be extracted for offline analysis. Figure 5 presents the exemplary data of the controller position (highlighted in blue) for one participant per group.

In the exemplary participants of the young group (Figure 5A) and the elderly group (Figure 5B), the controller movements were symmetrically distributed. In the exemplary participants of the stroke group (Figure 5C), a narrowing of the movement as well as a clear rightward shift can be observed. The patients’ hand movements mainly took place within the right hemispace due to left neglect.
Figure 5. The 3 panels show 3 exemplary controller movements (within a 2x2 m space), which represents the participant's hand movement around the participant's position (grey circle) over the total course of the task.

Search Time
The mean search time in young (mean 1.533, SD 0.342) and elderly (mean 1.694, SD 0.275) participants did not significantly differ ($t_{38.4}=1.71$, $P=.095$). There was a significant, strong correlation between the mean search time and neglect severity (ie, the CoC in the SNT; $r=0.70$, $P=.037$; [59]). An additional analysis investigating the correlation between search time per hemispace (ie, targets appearing within the left vs right hemispace) and neglect severity revealed a hemispace-dependent effect (Figure 6). A significant correlation was found for targets appearing within the left hemispace ($r=.809$, $P=.008$; strong correlation [59]). However, for targets appearing within the right hemispace, no significant correlation was found ($r=.353$, $P=.351$).

Figure 6. Correlation between the Center of Cancellation (CoC; as an objective measure of neglect severity) in the Sensitive Neglect Test (SNT) and the mean search time for the (A) left hemispace and (B) right hemispace in patients with neglect.

Discussion
For this study, we developed and evaluated a dynamic visual search task in iVR. As hypothesized, the task could easily be adapted to different skills, impairments, and experience levels. Further, the developed system has high usability and acceptance and resulted in only slight or no side effects in healthy participants and in right-hemispheric stroke patients with hemispatial neglect.

Questionnaires and Feedback
The ratings for usability and acceptance of the bird search task were high in both healthy participants and stroke patients with hemispatial neglect. This result is in line with previous findings from studies investigating the acceptance of HMD-VR in elderly [60,61], studies using VR for neglect diagnostics [51], and studies in which a similar VR setup was used to apply audio-visual stimulation in critically ill patients [62].

Regarding side effects, overall, only slight to none were reported. Furthermore, the nature and intensity of side effects
were in line with the literature [51,54,62]. In the group of patients with neglect, no severe side effects were reported (only 1 participant reported moderate sweating). This is potentially due to the game adaptations (eg, slower targets require slower head movements), simplification, and individual adjustments performed for the patients in step two. This low rate of side effects is also in line with the high values for immersion and presence, as those correlate negatively [63].

The game experience was assessed by means of a questionnaire with 4 parts (frustration, degree of challenge, entertainment, motivation). Indeed, motivation, entertainment, and low frustration levels are crucial aspects for any possible future application [64]. In the bird search task, all participants were highly motivated, entertained, and even though not all participants found all targets, they did not feel frustrated. In particular, the group of patients with neglect had high levels of motivation and entertainment and low frustration levels due to 2 possible reasons. First, this might be the result of the adaptive difficulty scaling; second, this could be the result of the ability to individualize the starting conditions to the patients’ needs and abilities. According to the literature, high motivation is also in line with the high values for immersion the participants reported [31,32,65,66].

Interestingly, the question as to whether the task was challenging was answered very differently among the 3 groups. While elderly healthy participants reported the task to be very challenging, the young healthy participants did find it rather easy. This difference between the elderly and young groups is not surprising, as most of the participants in the elderly group did not have any VR experience, whereas the young group was rather experienced in VR. The distribution of answers in patients with neglect was rather spread, indicating that, for some patients, the task was not particularly challenging, while for others, it was. A possible explanation for this result might be found in the individual settings adapted for the group of patients with neglect. As the individual performance level was estimated, some might have been under- or overestimated. Nevertheless, even if the performance level of some patients with neglect would have been under-or overestimated, according to their ratings, they were still highly entertained and motivated.

**Difficulty Levels and Adaptive Difficulty Scaling**

The descriptive analysis of the change in difficulty levels over time evidenced that the participants in the young group showed a steep increase in the initial phases of the task, suggesting a higher balance point than the one they started with. In the elderly group, the initial increase was less steep, but it was still clearly observable and gradual over time. In the neglect group, the initial increase seems to be delayed, but over time, there is a level increase relative to the starting level. This might be due to the attentional impairments and other cognitive deficits, resulting in a need for more rounds to get better at the game. Nevertheless, based on this measure, we could show that the individualized task was able to adapt to the individual level of impairment of the patients over time and thus keep them motivated and not overwhelmed [18].

Therefore, as the feedback regarding motivation and frustration assessed by means of the questionnaires shows clearly, the adaptive difficulty scaling was able to address the issue of very different skill levels and different progression speeds across participants.

**In-Game Measures**

VR tasks offer several opportunities to individually evaluate participants’ task performance using in-game measures. One of these possibilities is the analysis of the hand position over time, reflecting the spatial search behavior of the participants. As patients with neglect typically fail to explore the contralesional space, the assessment of the individual hand movements may be a valuable parameter to characterize neglect manifestations and severity. Indeed, in exemplary data sets, we were able to show typical neglect patterns (ie, healthy participants move their hand in the peripersonal space symmetrically during the search, whereas for patients with neglect, these movements are limited within the left, contralesional side).

Another parameter to estimate the ecological validity of the task in patients with neglect is the average search time relative to the objective neglect severity. Our results revealed the typical neglect pattern [37,67]; the worse the neglect — as measured by the CoC — the more time the stroke patients needed to find targets appearing within their contralesional, left side.

**Limitations**

The main limitation of the present study is the small sample size of the neglect group and the fact that the patients were not assessed with a comprehensive test battery (ie, including other measures of neglect severity like behavior in free visual exploration [6], other cancellation tasks [68], or line bisection tasks [69]). Furthermore, even though the group was age-matched, there were mainly male participants in the group of healthy elderly participants, and as they were all chess players, the generalizability of the results in this group are limited. For this, further research with a more representative sample would be needed.

Due to the 2-step process, no direct comparison of the in-game performance between participants with normal and impaired visual exploration behavior was possible.

**Outlook and Conclusion**

The presented bird search task was shown to be entertaining, motivating, and even immersive. Due to the implemented difficulty levels, it adapted well to different populations with different skills and previous VR experience and even in patients with cognitive disturbance after stroke. In particular, the bird search task seems to be able to pick up on typical patterns of neglect and to correlate with the results of established instruments.

Future studies should investigate and evaluate these aspects for potential application in diagnosis or therapy. For the further evaluation of the potential diagnostic or rehabilitative value, the frequency and duration of playing the task should be investigated in a longitudinal randomized clinical trial. Furthermore, the tool should be compared with the standard care of patients with neglect.
Acknowledgments

T Nyffeler was supported by SNF grant 32003B_196915.

Authors’ Contributions

SEJK, SMG, DC, BCK, RMM, T Nef, and T Nyffeler designed the study. SEJK, SMG, and T Nef developed the tool and the setup. SEJK, BCK, and PU recruited the participants. SEJK, BCK, T Nef, and RMM analyzed the data. SEJK, DC, BCK, and PU wrote the manuscript. All authors approved the final manuscript.

Conflicts of Interest

None declared.

References


42. Unity Technologies. URL: https://unity.com/ [accessed 2021-06-13]


47. Reinhart S, Leonhard E, Kerkhoff G. Sensitive Neglect Test (SNT) single and dual task. Saarland University. URL: https://www.uni-saarland.de/fileadmin/upload/lehrstuhl/kerkhoff/Materialien_f%C3%BCr_Diagnostik_Therapie/Instructions_SNT_single_and_dual_task.pdf [accessed 2021-06-13]


57. R Core Team. R: A Language and Environment for Statistical Computing. 201. URL: https://www.r-project.org [accessed 2021-06-13]


Abbreviations

CBS: Catherine Bergego Scale
CoC: Center of Cancellation
HMD: head-mounted display
IPQ: Igroup Presence Questionnaire
iVR: immersive virtual reality
PGTQ: Perception of Game Training Questionnaire
SNT: Sensitive Neglect Test
SSQ: Simulator Sickness Questionnaire
SUS: System Usability Scale

©Samuel Elia Johannes Knobel, Brigitte Charlotte Kaufmann, Stephan Moreno Gerber, Prabitha Urwyler, Dario Cazzoli, René M Müri, Tobias Nef, Thomas Nyffeler. Originally published in JMIR Serious Games (https://games.jmir.org), 02.07.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Serious Games, is properly cited. The complete bibliographic information, a link to the original publication on https://games.jmir.org, as well as this copyright and license information must be included.