Utrecht University

The feasibility, user experience and preference of immersive and non-immersive user interfaces and the influence of time post stroke

Master thesis Neuropsychology

Utrecht University

Name: Neeltje Op ’t Hoog
Studentnumber: 4184564

Supervisors:

Utrecht University	T. C. W. Nijboer, Dr.
De Hoogstraat Revalidatie	L. A. S. Spreij, MSc.
De Hoogstraat Revalidatie	I. K. Gosselt, MSc.

Abstract

Background: After a stroke cognitive impairments are currently assessed with a paper-and-pencil neuropsychological assessment. Virtual Reality (VR) may probably be more ecologically valid. It is not yet clear how stroke patients (often prone to overstimulation) tolerate and experience VR.

Objective: The aim of this study is to compare the feasibility, user experience and preference of immersive (HMD) and non-immersive (desktop) user interfaces in a stroke population and to explore the effect of time post stroke.

Methods: Fifty-nine stroke patients shopped in a virtual supermarket, once with each user interface. Feasibility measures (i.e., ability to complete the task, total duration, total number of correctly detected products) were derived. Afterwards patients filled in a questionnaire regarding their user experience with each user interface and a questionnaire regarding their preference.

Results: There was no significant difference in number of patients completing the task, total duration and number of products found between the user interfaces. The feeling of presence, transportation, flow, negative effects and the overall experience of patients were higher with the immersive compared to the non-immersive interface. No significant interaction was found between time post stroke onset and the user interfaces, except for the number of products found. The majority of the sample did not prefer either interface.

Conclusion: VR (immersive as well as non-immersive) can be used in clinical practice, at different times post stroke. The user experience is in general better with an immersive interface. Future research should therefore establish the ecological validity of (particularly immersive) VR in neuropsychological assessment.

Keywords: Virtual reality – user interface – immersive – non-immersive – feasibility – user experience – stroke – time post stroke

Introduction

A stroke may cause physical-, cognitive-, emotional-, or behavioral disorders (Consortium Cognitieve Revalidatie, 2007). Cognitive impairment, in particular, is one of the main causes of disability in daily functioning (Consortium Cognitieve Revalidatie, 2007; Rasquin et al., 2004) and occurs in 50-70% of the stroke patients (Jokinen et al., 2015; Linden, Skoog, Fagerberg, Steen, & Blomstrand, 2004; Wall, Isaacs, Copland, & Cumming, 2015). To assess cognitive impairment, a neuropsychological assessment (NPA) is used (Hendriks, Kessels, Gorissen, Schmand, & Duits, 2014). The NPA currently consists of different paper-and-pencil tests which are usually conducted in a quiet environment with little external distraction and little time pressure. The nature of this test situation is not comparable to the real-life situation (Spreij et al., 2017). The ecological validity (i.e., to which extent performance on a paper-and-pencil test corresponds to performance in the real world (Chaytor & Schmitter-Edgecombe, 2003)) of paper-and-pencil tests is therefore insufficient (Parsons, 2015). As a result, people sometimes encounter problems in everyday life that were not objectified with paper-and-pencil tests (Kang et al., 2008). Hence, more dynamic tests and more sensitive outcome measures are needed.

The introduction of Virtual Reality (VR) may offer solutions (Kang et al., 2008). With VR, people can be tested in a virtual environment that resembles a real-life situation (Lee et al., 2003). VR is therefore probably more ecologically valid than standard paper-and-pencil tests (Parsons, 2015). Other advantages of VR concern an adjustable level of difficulty (e.g., tasks can be adjusted to individual patients and their abilities) and the availability of new outcome measures derived from new technology (e.g., eye tracking enables data on point of gaze, which may be informative for measuring attention) (Lee et al., 2003; Spreij et al., 2017). Furthermore, through the game component in VR people may be more motivated to perform cognitive tasks (Tieri, Morone, Paolucci, & Iosa, 2018; Verheul et al., 2016).

Virtual environments can be presented with various VR user interfaces. VR user interfaces differ in the extent to which people can interact with the environment (Rizzo & Koenig, 2017). The least interactive implementation of VR user interfaces is non-immersive VR, in which an environment is displayed on a screen and interaction is managed by using a keyboard or a mouse. A more interactive implementation is immersive VR, in which people are fully immersed by using a Head Mounted Display (HMD) (Shahrbanian et al., 2012). Santos et al. (2009) found that global performance is better with a non-immersive than with an immersive user interface in healthy participants, because this interface was familiar to many people. Besides the performance, the user experience could also differ between non-immersive and immersive interfaces. An important concept to study user experience is 'presence', described as the subjective sensation of being in a virtual environment (Barfield, Zeltzer, Sheridan, & Slater, 1995). A sense of presence enables the same reactions in the virtual environment as in the real world (Schuemie, Van Der Straaten, Krijn, & Van Der Mast, 2001), as a consequence the experience in a virtual environment may be engaging and relevant (De Leo, Diggs, Radici, & Mastaglio, 2014). Previous studies reported a higher feeling of presence with immersive than with non-immersive user interfaces in healthy people (Gorini, Capideville, De Leo, Mantovani, & Riva, 2011; Lo Priore, Castelnuovo, Liccione, & Liccione, 2003). However, negative side effects (e.g. cybersickness and/or headache) are reported to be higher with immersive user interfaces (Lo Priore et al., 2003; Rand et al., 2005; Rizzo & Kim, 2005).

This research will further explore the feasibility, user experience and preference of immersive VR, compared to non-immersive VR, in stroke patients. Up to now, there has been very little research directly investigating the feasibility, user experience and preference in a stroke population (Ouellet, Boller, Corriveau-Lecavalier, Cloutier, & Belleville, 2018). This investigation is especially important since external noise (e.g. light, sound, crowds and/or touch) can be difficult to tolerate for stroke patients (Scheydt et al., 2017; Zedlitz & Fasotti, 2010). More environmental distractors are present in VR (i.e., it is a representation of the real world) compared to paper-and-pencil tests, so it may be less well tolerated. In particular immersive VR may cause overstimulation since the screen is close, the light is fairly bright and a heavy device is placed on the head of a patient. In a study of Kang et al. (2008) for example, stroke patients who were tested with an immersive user interface stated that they experienced problems with using the equipment and according to them the use was uncomfortable. Since the assessment of cognitive functions may be necessary on different times post stroke, it is important to explore the feasibility and user experience of VR on different times post stroke. Spontaneous recovery mostly takes place in the first three months post stroke (Horgan, O'regan, Cunningham, & Finn, 2009; Skilbeck, Wade, Hewer, & Wood, 1983) and after 6 months, only little recovery is expected (Langhorne, Bernhardt, & Kwakkel, 2011). We will therefore divide the group of patients in this study in three groups: (1) < 3 months post stroke; (2) 3-6 months post stroke; and (3) > 6 months post stroke.

The first aim of the current study is to compare the differences in the feasibility, user experience and preference for an immersive and non-immersive user interface in a stroke population. The second aim is to explore whether an interaction exists between the feasibility, user experience and preference in the immersive and non-immersive user interface and the groups with different times post stroke (i.e. < 3 months, 3-6 months and > 6 months).

Materials and Methods:

Patients

We included stroke patients who were admitted for inpatient or outpatient rehabilitation. Outpatients were recruited at the University Medical Centre Utrecht (UMC Utrecht) (from June 2016 to September 2018) and at the Hoogstraat Rehabilitation Centre (from May 2018 to November 2018). Inpatients were recruited at the Hoogstraat Rehabilitation Centre (from February 2018 to November 2018). Inclusion criteria for this study were for all stroke patients: (1) clinically diagnosed stroke (confirmed by a MRI or CT scan); (2) > 18 years old; (3) mentally able to participate (evaluated by a neuropsychologist/ rehabilitation physician); (4) voluntary participation. Exclusion criteria were: (1) diagnosis of epilepsy; (2) diagnosis of neglect; (3) motor problems (preventing that a controller can be controlled); (4) comprehension- and communication problems (preventing the task from being properly understood and executed); (5) arousal problems. The experiment was performed in accordance with the Declaration of Helsinki. The Medical Research Ethics Committee of the UMC Utrecht approved the research protocol (number 15-761/C).

Procedure

Inpatients were first assessed by a neuropsychologist administering a visuo-spatial neglect screening. From there, patients with neglect were excluded and the mental ability to participate was evaluated. In case all other criteria were met, the researcher would inform the patient about the study and gave the patient the opportunity to ask questions. The patient was given a few days to consider participation. After confirmation, an appointment was scheduled.

Outpatients recruited in the UMC Utrecht were assessed by a rehabilitation physician with respect to the inclusion- and exclusion criteria. The physician discussed participation and there was the opportunity to ask questions. If patients confirmed their participation, an appointment was scheduled by phone with the researcher.

Outpatients at the Hoogstraat Rehabilitation Centre were assessed on the inclusion- and exclusion criteria in the electronic medical records by the researcher. The patients who met criteria were then assessed by their rehabilitation physician on the mental ability to participate. Eligible patients received an information letter and information folder. Participation was discussed by phone with the opportunity to ask questions. Appointments were scheduled and linked to an already existing appointment.

Experimental task

Hardware (Desktop and controller/Oculus Rift/HTC VIVE)

This study used two different VR user interfaces, a desktop pc with a controller (non-immersive) and a HMD (immersive).

The desktop (HP) used in the UMC Utrecht had a 24 inch monitor with a resolution of 1920 x 1200 pixels. The desktop (AOC) used in the Hoogstraat Rehabilitation Centre had a 23.6 inch monitor with a resolution of 1920 x 1080 pixels. To change position in the non-immersive virtual environment, a controller (Xbox 360) that was connected to the computer was used.

Two types of HMD’s (i.e. the Oculus Rift and the HTC VIVE) were used in this study. The first fourteen patients recruited in the UMC Utrecht (from June 2016 to October 2016) were tested with the Oculus Rift. The Oculus Rift has a refresh rate of 90 Hz and a 94 degree field of view. It uses an OLED panel for each eye, with a resolution of 1080 x 1200 (Borrego, Latorre, Alcaniz, & Llorens, 2018). The Xbox controller was used to navigate through the environment.

For the remaining patients (from February 2018 to September 2018) the HTC VIVE was used. The headset has a refresh rate of 90 Hz and a 110 degree field of view, the two screens each have a display resolution of 1080 x 1200 (Vive, z.d.). HTC VIVE comes with two controllers, the input methods are a track pad, grip buttons and a dual stage trigger. Finally, there are two base stations, necessary for tracking. These base stations create a 360 degree virtual space up to 15 x 15 foot radius. The tracking system provides its own personal GPS system in the room, down to the millimetre. Patients could move in the virtual environment through real time movement. Build in safety’s prevent collisions with walls or obstacles in the real room.

Software

The virtual environment used was a supermarket, based on a real-life Dutch supermarket. The supermarket contains over 12.000 products from well-known existing brands. There are 18 shelves, 8 cash registers and different departments among which were fruit and vegetables, bread and also a refrigerated section. The supermarkets surface is 50x30 virtual meters. Patients walked through the supermarket at a rate of 0.5 meters per second.

Procedure

For the outpatients, the Montreal Cognitive assessment (MoCA) was administered first (MoCA scores from the inpatients were extracted from the electronic medical record). Then, patients were instructed about the equipment and received a practical trial in an empty virtual supermarket to get familiar with the equipment. Patients were instructed to find three products in the supermarket, to mention these out loud when found and to pass the cash register to finish the task. Thereafter the shopping list was shown to the patient while the researcher read the products simultaneously out loud. After this the list had to be recited by the patient. This was repeated two times. The patient then started the task. All patients performed the task while sitting in a mobile chair with wheels. The task had a maximum time of 15 minutes, if patients didn't find all the products within this time, they were asked to finish the task. After finishing the task, patients were asked to recall the products they had to find. This was necessary to differentiate whether patients weren't able to remember (memory) or find (attention, executive function) a product in the supermarket. Next, a questionnaire was presented to the patient regarding their experience in the supermarket. The procedure was then repeated from the practical trial until the questionnaire for the other user interface. Two different shopping lists were used, one for each condition. Both the user interfaces and the shopping lists were offered to the patients in randomized order. Finally, patients completed a questionnaire regarding their preference for one of both user interfaces.

Outcome measures

Demographic and clinical characteristics

Age, gender and level of education were collected during the experiment. A Dutch classification system consisting of seven levels was used to assess level of education (Verhage, 1965). These seven levels were transformed into three levels for analysis: low (Verhage 1-4), average (Verhage 5) and high (Verhage 6-7). From the electronic medical records were furthermore extracted: days post stroke onset, lesion side, stroke type and MoCA scores for inpatients.

Feasibility measures

Three types of data collected in the supermarket were used to examine feasibility. First we examined the ability to complete the task (i.e., yes or no). When the task was aborted before finishing, this was graded as 'no' (not able to complete the task). Second, we examined the total time spent in the supermarket (i.e., in seconds). The maximum time in the supermarket was 15 minutes, patients who exceeded this limit were given the maximum of 15 minutes for analyses Finally, we examined the total number of correctly detected products (i.e., total hits, ranging from 1-3). Products that were found after 15 minutes were not counted.

Questionnaire user experience

The questionnaire consisted of 15 questions subdivided into five clusters, each cluster contained three questions (see Table 1). The clusters were: (1) *engagement*, defined as to what extent the task can fascinate and/or occupy the patient, (2) *transportation*, defined as transition in time and place, (3) *flow*, defined as engagement at the level of no room for secondary thoughts, (4) *presence*, defined as the sense of ‘being there’, and (5) *negative effects*, defined as possible discomfort patients may experience (Schuemie et al., 2001, Lessiter, Freeman, Keogh, & Davidoff, 2001). Answer options for the questions were based on a 6 point Likert scale, from strongly disagree (---) to strongly agree (+++).

Before we calculated the overall score (*total experience*), scores on the cluster negative effects were re-coded so that higher scores indicated less negative effects (score of - - - [1] was recoded into a + + + [6]. - - [2] was recoded into + + [5] and - [3] was recoded into + [4]). To obtain the overall score, all scores were summed into a total score. This questionnaire was compiled by the researchers themselves since no existing questionnaire was available including all the concepts described. The questionnaire was set up on the basis of existing literature regarding this concepts (De Leo et al., 2014; Lessiter et al., 2001; Price, Mehta, Tone, & Anderson, 2011; Schuemie et al., 2001). The face validity of the questionnaire has been explored in a sample of 32 people (gender (known from 25 participants): 16% male, mean age (known from 18 participants): 29.11 [10.55]). 81.3% of the sample thought the questions could be divided into 5 clusters (as intended), 15.6% thought that the fifteen questions could be divided into 4 clusters (according to them, transportation and presence were the same cluster) and one person thought the questions could be divided into 3 clusters (according to this person, transportation, presence and flow were the same cluster). The percentages of people categorizing the three questions to the clusters as intended were 59.3% (presence), 71.9% (transportation), 90.6% (engagement), 100% (negative effects) and 53.1% (flow). Based on these findings, engagement and negative effects seem quite solid clusters and flow seems a little less solid.

Table 1

Questions per cluster of the experience questionnaire

Transportation
During the assignment I had the feeling to step into a different world
I felt like I was being put into a different world
After the task I felt that I came back from a trip in another world
Presence
I had the feeling to be present in the virtual world
I felt absorbed in the environment
I was part of the environment
Engagement
I wanted to explore the area
The virtual world attracted my interest
I was curious about the environment
Flow
I only paid attention to the environment and less to other thoughts
I was hardly aware of the real world
I had forgotten the time
Negative effects
I got warm
I felt nauseous
I had a headache

Questionnaire Preference

The questionnaire consisted of five statements regarding patients' preference on the used user interfaces: (1) I was motivated to carry out the task; (2) I found the assignment enjoyable; (3) I would like to do virtual shopping again; (4) I would like to do virtual shopping at home; (5) I would like to do this task regularly. Answer options were: 'desktop', 'HMD' and 'both'. In the current study, only question two will be used to assess preference, because this question purely measures which user interface patients enjoyed the most.

Statistical analyses

Categorization and inclusion of patients

Patients were categorized in three groups on the basis of time post stroke. The first group consisted of patients who were tested in the period up to 3 months after stroke. The second group consisted of patients tested between 3 and 6 months after stroke and the third group consisted of patients that were tested more than 6 months after stroke.

Demographic and clinical characteristics

We compared the three groups ([1] <3 months post stroke; [2] 3-6 months post stroke; [3] >6 months post stroke) on demographic and clinical variables. Parametric tests (one way between groups ANOVA) were used for variables that met the assumptions. Non parametric tests (Kruskal-Wallis ANOVA for continuous variables and Chi-square tests for categorical variables) were used for data that was not normally distributed. In case the data for the Chi-Square did not meet the assumption of expected count, a Fisher's exact test was used. Because of multiple testing a Bonferroni correction was applied. The adjusted p-value for 7 tests is $p = .007$.

Differences between Head Mounted Displays (Oculus Rift vs. HTC VIVE)

To compare the results on the feasibility-, experience-, and preference variables between the Oculus Rift and the HTC VIVE, parametric tests (independent samples t-test) were used for variables that met the assumptions. We used Mann-Whitney U Tests for continuous variables and Chi-Square tests for categorical variables. Again, a Fisher's exact test was used when the data of the Chi-Square test did not meet the assumption of expected count. Since multiple tests were conducted on the same set of data, a Bonferroni correction was applied. The adjusted p-value for 10 tests is $p = .005$.

In case a significant difference exists between the Oculus Rift and the HTC VIVE on one or more of the outcome variables, the HMD's will be compared separately to the desktop user interface for these outcome variables.

Differences between user interfaces (Desktop vs. HMD)

To compare the results on the feasibility-, experience-, and preference variables between the desktop and the HMD, parametric tests (paired samples t-test) were used for variables that met assumptions. Variables that not met assumptions were analysed with non-parametric tests (Wilcoxon-Signed Rank Test for continuous and not normally distributed variables and McNemar test for categorical variables). Since multiple tests were conducted on the same set of data, a Bonferroni correction was applied. The adjusted p-value for 9 tests is $p = .006$. To describe the preference, percentages were given of how many times 'HMD', 'desktop' or 'both' was chosen as preferred user interface.

Interaction between user interfaces and time post stroke

For the analysis of the interaction effect between the two user interfaces and the three groups (based on different times post stroke), the only available statistical test is a parametric test (the Mixed Model ANOVA). Some of the outcome variables used did not meet the assumptions necessary to use a parametric test. However, since there is no non-parametric equivalent for the Mixed Model ANOVA, we used this test for all our outcome variables. To analyze the difference in preference for one of both user interfaces between the three groups, a Chi-Square was used. In case the data for the Chi-Square did not meet the assumption of expected count, Fisher's exact test was used. Since multiple tests were conducted on the same set of data, a Bonferroni correction was applied. The adjusted p-value for 10 tests is $p = .005$.

Results

A total of 59 stroke patients were recruited. 19 of these patients were tested < 3 months post stroke, 15 were tested between 3 and 6 months post stroke an 25 were tested > 6 months post stroke.

A total of 109 patients were in outpatient treatment in the Hoogstraat Rehabilitation Centre, 48 of these patients met criteria to participate in this study. 15 patients agreed on participation of which 2 cancelled their appointment. A total of 48 inpatients was approached for participation, 16 of them agreed on participation. Inclusion of outpatients in the UMC Utrecht was executed by the treating physician of patients. It is therefore not known how many patients were in outpatient treatment and how many of them met criteria.

Since exclusions differed per outcome variable, the exclusions and reasons for exclusion are described in the 'notes' section below each table providing data.

Demographic and clinical variables

Demographic and clinical characteristics are reported in Table 2. The groups did not differ significantly in gender, age, level of education, handedness, lesion side and stroke type. Global cognitive functioning, measured with the MoCA, was comparable between patients in the different groups.

Table 2

Demographic and clinical characteristics, means (SD), or percentiles split per group

	<3 months post stroke (*n* = 19)	***n***	**3 - 6 months post stroke (*n* = 15)**	***n***	**>6 months post stroke (*n* = 25)**	***n***	**Statistics**
Gender (%)		19		15		25	$\chi^2(2)=3.914$, $p = .141$
Man	73.3		53.3		44.0		
Women	26.3		46.7		56.0		
Age (years)	60.68 (14.74)	19	54.73 (17.01)	15	55.80 (8.87)	25	$F(2,56)=1.05$, $p = .356$
Education (%)		19		15		25	$H(2)=3.28$, $p = .194$
Low	21.1		20.0		28.0		
Average	5.3		33.3		32.0		
High	73.3		46.7		40.0		
Time post stroke (days)	48.79 (15.60)	19	145.60 (24.26)	15	689.00 (604.08)	22	
MoCA-score (1-30)	22.69 (5.10)	16	24.64 (3.59)	11	23.63 (4.56)	19	$F(2,43)=.603$, $p = .552$
Handedness (%)		17		14		18	Fisher's = 4.766, $p = .104$
Right	58.8		92.3		83.3		
Left	41.2		7.7		16.7		
Lesion side (%)		19		14		22	$H(2) = 1.46$, $p = .483$
Right	57.9		50.0		45.5		
Left	36.8		42.9		31.8		
Both	5.3		7.1		22.7		
Stroke type (%)		19		14		22	$H(2) = 6.90$, $p = .032$
Ischemic	94.7		57.1		63.6		
Hemorrhage	0.0		7.1		9.1		
SAH	5.3		35.7		27.3		

* $p < .007$

Notes. MoCA = Montreal Cognitive; Assessment; SAH = Subarachnoid Haemorrhage

Differences between HMD's: HTC VIVE and Oculus Rift

Overall, feasibility outcome measures (completion, total time and total hits), experience outcome measures (engagement, transportation, flow, presence, negative effects and total experience) and also the preference outcome measure were comparable between the two types of HMD user interfaces (see Table 3). As such, the data of the two types of HMD user interfaces will be aggregated when comparing the HMD user interface with the desktop user interface.

Table 3

Differences between the HMD's, means (SD), or percentages split per type of HMD

	Oculus Rift (n = 14)	*n*	HTC VIVE (n = 45)	*n*	Statistics
Feasibility					
Completion (% completed)	92.9%	14	86.0%	43	Fisher's $p = .669$
Total time (sec)	475.85 (64.71)	13	609.43 (39.35)	37	$U = 161.00$, $z = -1.76$, $p = .078$
Total hits (1-3)	2.69 (.18)	13	2.19 (.18)	37	$U = 183.00$, $z = -1.46$, $p = .143$
User experience					
Engagement (3-18)	11.86 (1.18)	14	13.44 (.52)	43	$U = 237.50$, $z = -1.18$, $p = .237$
Transportation (3-18)	12.57 (.83)	14	12.79 (.52)	43	$t(55) = -.214$, $p = .832$
Flow (3-18)	12.50 (.81)	14	11.98 (.53)	43	$t(55) = .505$, $p = .616$
Presence (3-18)	12.79 (.77)	14	14.17 (.41)	42	$U = 220.00$, $z = -1.41$, $p = .158$
Negative effects (3-18)	11.36 (1.21)	14	8.02 (.71)	43	$U = 185.00$, $z = -2.17$, $p = .030$
Total experience (15-90)	59.36 (2.50)	14	65.57 (1.56)	42	$t(54) = -2.03$, $p = .047$
Preference (%)		14		42	Fisher's = 3.765, $p = .172$
Desktop	51.7		31.0		
HMD	14.3		11.9		
Both	28.6		51.7		

$*p < .005$

Notes. For all analyses, $n = 2$ were excluded because they did not start the task. For the analyses of the total time and total hits $n = 7$ were excluded due to abortion before finishing the task. For the analysis of presence $n = 1$ was excluded due to missing data on one of the questions of this cluster. For the analysis of the total experience $n = 1$ was excluded due to missing data on the 'presence' cluster.

For the score on the cluster negative effects applies the higher the score, the more negative effects experienced. For the score on the cluster total experience the scores on the cluster negative effects are recoded: higher scores indicating less experience of negative effects.

Differences between user interfaces

There were no significant differences between the two user interfaces regarding the feasibility outcome measures (i.e., completion, total time and total hits). There were significant differences between the two user interfaces regarding the user experience. Significant effects were found for the feeling of transportation, flow and presence, the experience of negative effects and the total experience. Scores on all these measures were significantly higher for the HMD user interface than for the desktop user interface. There was no significant effect for the feeling of engagement (see Table 4 and Figure 2). Regarding the preference: 11.9% of the patients chose the option 'HMD', 35.6% chose for the option 'desktop' and 47.5% chose the option 'both'.

Table 4

Differences between the user- interfaces, means (SD), or percentages split per user interface

	Desktop (n = 59)	*n*	HMD (n = 59)	*n*	Statistics	Effect Size
Feasibility						
Completion (% completed)	96.6%	56	87.7%	56	McNemar, $p = .016$	
Total time (sec)	634.93 (232.40)	49	571.82 (244.14)	49	$T = 17$, $p = .045$	
Total hits (1-3)	2.16 (.97)	49	2.35 (.99)	49	$T = 10$, $p = .219$	
User experience						
Engagement (3-18)	11.93 (3.07)	55	12.93 (3.73)	55	$T = 13$, $p = .015$	
Transportation (3-18)	9.30 (3.59)	56	12.66 (3.29)	56	$T = 9$, $p < .001$*	$r = .68$
Flow (3-18)	10.23 (3.59)	56	12.09 (3.38)	56	$t(55) = -3.83$, $p < .001$*	$d = .53$
Presence (3-18)	10.62 (3.29)	52	13.92 (2.73)	52	$T = 7$, $p < .001$*	$r = .72$
Negative effects (3-18)	5.25 (3.27)	55	8.64 (4.78)	55	$T = 6$, $p < .001$*	$r = .66$
Total experience (15-90)	58.76 (10.65)	50	65.04 (9.41)	50	$t(49) = -4.71$, $p < .001$*	$d = .63$

* $p < .006$

Notes. For all analyses, $n = 3$ were excluded because they did not start the task. For the analyses of the total time and total hits $n = 7$ were excluded due to abortion before finishing the task. For the analysis of engagement $n = 1$ was excluded due to missing data on one of the questions of this cluster. For the analysis of presence $n = 4$ were excluded due to missing data on one of the questions of this cluster. For the analysis of negative effects $n = 1$ was excluded due to missing data on one of the questions of this cluster. For the analysis of the total experience $n = 9$ were excluded due to missing data on the ‘engagement’, ‘presence’ and ‘negative effects’ clusters.

For the score on the cluster negative effects applies the higher the score, the more negative effects experienced. For the score on the cluster total experience the scores on the cluster negative effects are recoded: higher scores indicating less experience of negative effects.

Interaction between user interfaces and time post stroke

No significant effects were found on any of the outcome measures for the interaction between the user interfaces and the groups with different times post stroke, except for the total amount of hits. There was a significant interaction between the user interfaces and the groups with different times post-stroke, $F(2,46) = 6.44$, $p = .003$. Examination of the means indicated that patients in group 1 (< 3 months post stroke) found more products with the desktop user interface than with the HMD user interface, while patients in group 2 (3-6 months post stroke) and group 3 (> 6 months post stroke) found more products with the HMD user interface than with the desktop user interface (see Table 5 and Figure 1).

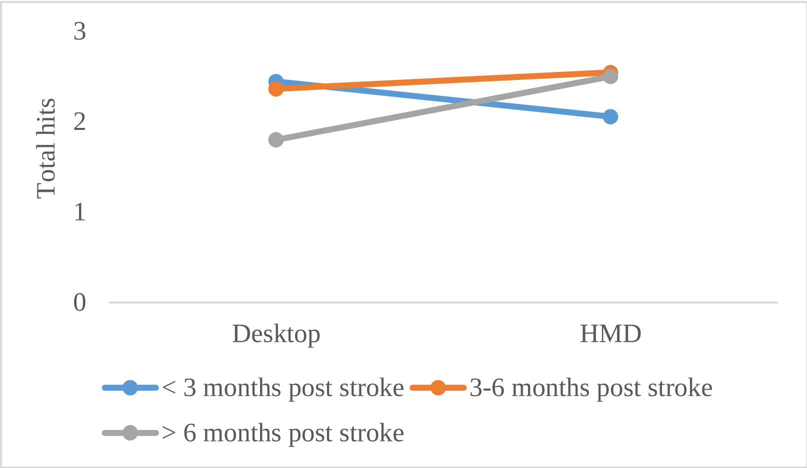

Figure 1. Interaction effect between time post stroke and the user interfaces on the total amount of hits (1-3).

Table 5

Interaction between the user interfaces and the groups with different times post stroke, means (SD), or percentages split per group and type of user interface

	< 3 months post stroke				**3-6 months post stroke**				**> 6 months post stroke**					
	Desktop (n = 19)	***n***	**HMD (n = 19)**	***n***	**Desktop (n = 15)**	***n***	**HMD (n = 15)**	***n***	**Desktop (n = 25)**	***n***	**HMD (n = 25)**	***n***	**Statistics**	**Effect size**
Feasibility														
Completion (% completed)	0,00 (.00)	19	0.05 (.23)	19	0.00 (.00)	12	0.08 (.29)	12	0.00 (.00)	25	0.20 (.33)	25	*F*(2,53) = .315, *p* = .043	
Total time (sec)	636.50 (248.16)	18	631.11 (256.16)	18	644.36 (230.98)	11	441.37 (242.48)	11	627.44 (230.54)	20	590.21 (218.45)	20	*F*(2,46) = 2.29, *p* = .113	
Total hits (1-3)	2.44 (.71)	18	2.06 (1.21)	18	2.36 (.81)	11	2.55 (.93)	11	1.80 (1.15)	20	2.50 (.76)	20	*F*(2,46) = 6.44, *p* = .003	partial $\eta^2 = .18$
User experience														
Engagement (3-18)	12.32 (2.34)	19	13.11 (3.57)	19	11.00 (3.84)	12	12.50 (4.23)	12	12.08 (3.19)	24	13.00 (3.74)	24	*F*(2,52) = .146, *p* = .865	
Transportation (3-18)	8.84 (3.88)	19	13.11 (2.87)	19	10.25 (3.96)	12	13.58 (2.58)	12	9.20 (3.23)	25	11.88 (3.79)	25	*F*(2,53) = .812, *p* = .450	
Flow (3-18)	10.68 (3.70)	19	12.42 (2.89)	19	10.17 (2.86)	12	13.17 (3.27)	12	9.92 (3.92)	25	11.32 (3.69)	25	*F*(2,53) = .797, *p* = .456	
Presence (3-18)	10.68 (3.27)	19	14.16 (2.34)	19	9.83 (3.59)	12	14.67 (2.57)	12	11.00 (3.21)	21	13.29 (3.10)	21	*F*(2,49) = 2.06, *p* = .138	
Negative effects (3-18)	4.21 (2.04)	19	6.21 (3.63)	19	6.00 (4.78)	11	10.09 (4.28)	11	5.72 (3.17)	25	9.84 (5.18)	25	*F*(2, 52) = 1.62, *p* = .208	
Total experience (15-90)	59.32 (10.84)	19	67.58 (8.95)	19	56.27 (11.00)	11	64.00 (9.84)	11	59.60 (10.63)	20	63.20 (9.53)	20	*F*(2, 47) = 1.38, *p* = .263	
Preference (%)	19.0	19	28.6	19	23.8	12	28.6	12	57.1	25	42.9	25	Fisher's=4.513, *p* = .343	

* $p < .005$

Notes. For all analyses, *n* = 3 were excluded because they did not start the task. For the analyses of the total time and total hits *n* = 7 were excluded due to abortion before finishing the task. For the analysis of engagement *n* = 1 was excluded due to missing data on one of the questions of this cluster. For the analysis of presence *n* = 4 were excluded due to missing data on one of the questions of this cluster. For the analysis of negative effects *n* = 1 was excluded due to missing data on one of the questions of this cluster. For the analysis of the total experience *n* = 9 were excluded due to missing data on the 'engagement', 'presence' and 'negative effects' clusters. For the score on the cluster negative effects applies the higher the score, the more negative effects experienced. For the score on the cluster total experience the scores on the cluster negative effects are recoded: higher scores indicating less experience of negative effects.

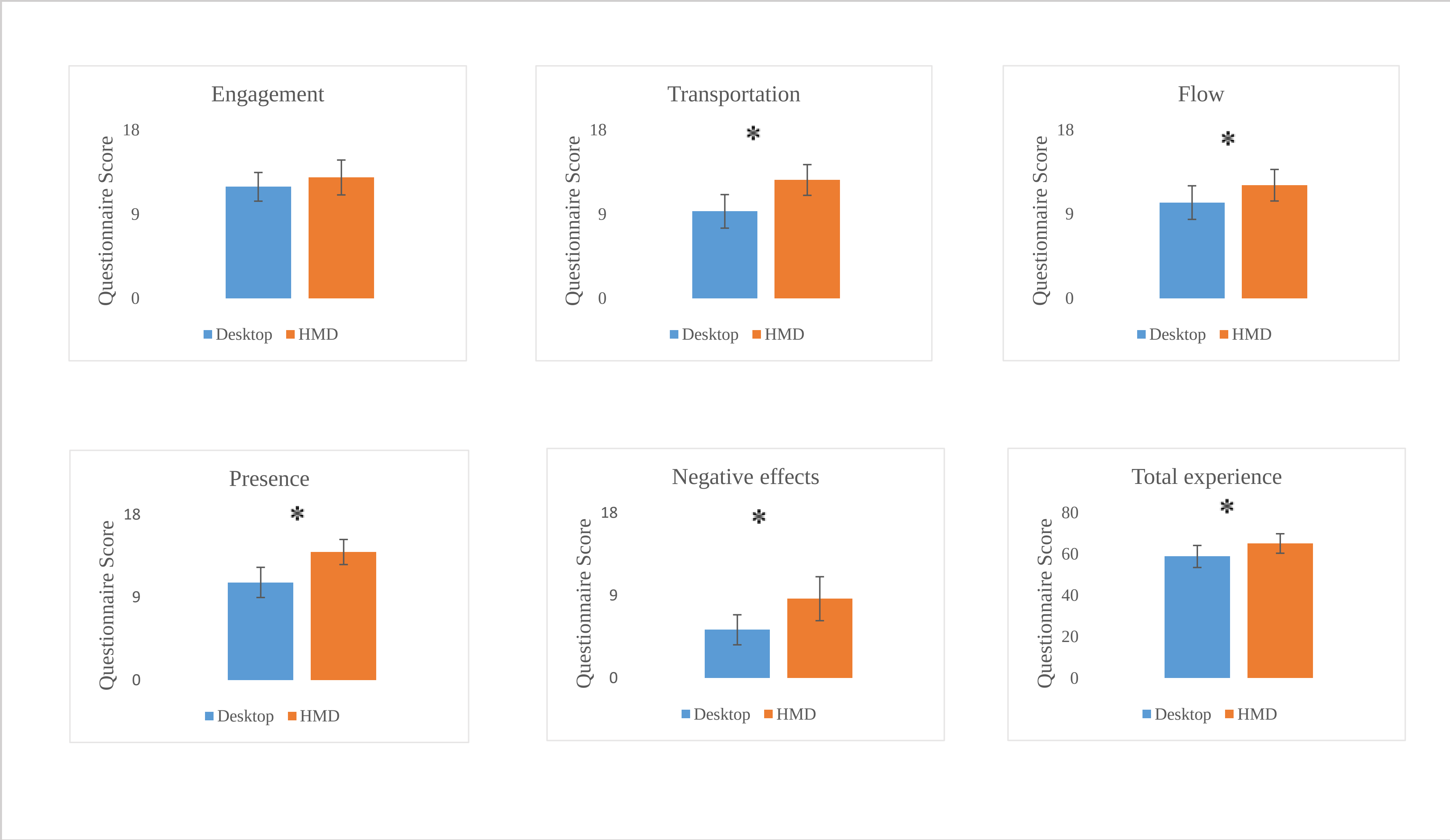

Figure 2. User interfaces (HMD and desktop) compared on the user experience questionnaire scores, split into the different clusters.

Discussion

The current study had two main aims. First, we compared the differences in the feasibility, user experience and preference for a non-immersive and immersive user interface in a stroke population. Second, we examined the influence of the user interfaces and the time post stroke on the feasibility, user experience and preference.

Regarding the feasibility no differences were found between the immersive and the non-immersive user interface concerning the ability to complete the task, the total time spent in the supermarket and the total amount of hits. When we look at the feasibility in general, we see that patients are reasonably able to complete the task with both user interfaces (96.6% with non-immersive and 87.7% with immersive interfaces), are able to find on average 2 out of 3 products and are on average able to complete the task within the predetermined time. This indicates that the use of VR is feasible, with both immersive and non-immersive interfaces. The type of hardware used for navigation could have contributed to the positive results found for both user interfaces. In a previous study, performance with an immersive user interface was poorer than with a non-immersive interface. In that study, a mouse was used to navigate through the virtual environment with an immersive interface (Santos et al., 2009). In the current study, navigation with the immersive interface was managed by real time movement in combination with a controller, which might have been more intuitive. Previous experience with a user interface could also have contributed to the absence of differences between interfaces in the current study. Having previously used a user interface, may lead to better performances with that particular interface (Santos et al.). Younger people are often more familiar with non-immersive interfaces than older people (Sayers, 2004). Since patients in our sample were relatively old compared to the healthy controls in the study of Santos et al., our sample was possibly not familiar with either user interface. As a result, the user interfaces may not have differed from each other.

When taking time post stroke into consideration, an interaction was found between the time post stroke and the user interfaces regarding the total amount of hits. Patients < 3 months post stroke were better able to find products in the supermarket with the non-immersive interface, while patients 3-6 months post stroke and patients > 6 months post stroke were better able to find products in the supermarket with the immersive interface. With regard to the number of patients able to complete the task and the total duration of the task, no effects of time were found. The effect of time post stroke on the total amount of hits could be explained by the relatively small sample size in each group. As a consequence, small variation in the data can lead to significant differences. When we look at the actual difference between the groups, we see this difference was less than one hit. The clinical relevance of this difference is questionable. Especially since the total amount of hits can also be a measure of cognitive performance. Yet, the primary aim of this study was not to investigate cognitive performance. We still examined this variable because, besides it being a measure of cognitive performance, it could also give insight into the general ability to find products in the supermarket. For a better understanding of the cognitive performance of patients in VR, studies are needed that compare performances to norm groups or to paper-and-pencil tests intended to measure the same construct.

Overall, positive results were found regarding the feasibility of VR on different times post stroke in the current study. It should nevertheless be taken into account that adjusting VR to individual patients may be necessary. This is now done with standard paper-and-pencil tests by means of testing the limits (Hendriks et al., 2014; Neuropsychological assessment, Lezak, 2004), and can also be applied with VR.

With respect to the user experience, differences were found for the constructs transportation, flow, presence, negative effects and the total experience between the user interfaces. The scores on all these constructs were higher with the immersive than with the non-immersive user interface, which is consistent with previous research (Gorini et al., 2011; Lo Priore et al., 2003; Rand et al., 2005; Rizzo & Kim, 2005). No differences were found between the user interfaces regarding engagement, which can be explained by the fact that patients were engaged regardless of the user interface. Since engagement is related to enjoyment and interest in the virtual environment (De Leo et al., 2014), engagement could be independent from the user interface, but be more related to the content of the virtual environment. Furthermore, no differences were found regarding engagement, transportation, flow, presence, negative effects and the total experience between the user interfaces on different times post stroke. An immersive user interface thus evoked a higher experience of immersion, which is expected to lead to performances that are more related to daily life functioning (Parsons, 2015). Immersive interfaces would therefore be preferred in neuropsychological testing. However, more negative effects were experienced with the immersive user interface which could affect its feasibility (Lampton et al., 1994). This is in line with the fact that patients who aborted the task with the immersive user interface did so because of the experience of nausea, dizziness or headache. This could counteract the use of immersive VR in clinical practice, but could partly be overcome by the provision of information about VR. Knowledge of VR can reduce the experience of negative effects (De Leo et al., 2014).

Finally, this study investigated the preference of patients for one or both user interfaces. A majority (47.5%) of the patients indicated they did not prefer one of both user interfaces. A less substantial part (35.6%) indicated they preferred the non-immersive user interface and a minority (11.9%) preferred the immersive user interface. Patients felt more immersed with the immersive interface but also experienced more negative effects. As a result, patients probably preferred both user interfaces instead of only the immersive interface. Moreover, the preference for one or both user interfaces did not differ on different times post stroke.

A limitation of the present study that should be considered, is that a part of the stroke patients was excluded because their impairments made participation impossible (e.g., motor problems that prevent a controller from being controller or communication problems preventing the task from being properly understood). However, there is bias in every study due to exclusion (Wall et al., 2015) and our criteria were drawn up allowing patients who are affected to a certain extent to be included. This is reflected in the MoCA scores, which on average fall below the cut-off point (< 26) (Lees et al., 2014). We can therefore make statements about the feasibility and user experience in a stroke population that is cognitive impaired. We cannot comment on the patients that were excluded in our study due to impairments that made participation impossible. On the other hand, these patients would probably not be able to execute standard paper-and-pencil tests either. Therefore, they are not the subgroup we wanted to focus on in this study. Another limitation that could have affected the sample, is the voluntary participation. Possibly only motivated patients participated. However, we do not know the exact reasons for refusing participation. The patient characteristics and performance of this group were unknown because we did not test them. We therefore cannot comment on the feasibility of this subgroup and caution remains needed when generalizing results to the general stroke population.

Since this was one of the first studies examining the feasibility, user experience and preference of VR user interfaces in a stroke population, future studies could further explore which variables affect the feasibility and user experience. Previous experience with user interfaces was for example not investigated in the current study but led in healthy controls to a better performance with the user interface known to the participants (Santos et al., 2009). If this is also true in a stroke population, practice with a user interface could facilitate the use. Furthermore, an immersive user interface evoked a higher feeling of presence in the current study. This could probably enhance the ecological validity. However, this is still an assumption and future research should further establish the ecological validity of VR, in particular with immersive user interfaces.

In conclusion, the current study showed promising results regarding the feasibility and user experience of immersive and non-immersive user interfaces in a stroke population on different times post stroke. Patients included in this study were in different rehabilitation phases and in general cognitively affected. Since the feasibility and user experience did not differ on different times post stroke, we recommend that VR can be applied in different rehabilitation phases. Nevertheless, adjustment of the virtual environment may sometimes be necessary. It could also be helpful when physicians and neuropsychologists provide patients with information about VR, because this may make the experience more interesting and can reduce the negative effects. Rehabilitation physicians and neuropsychologists should always critically review for each individual patient whether VR is feasible. However, before VR can be actually implemented, more research has to be done into the usability and ecological validity of the technique. This may be especially interesting for immersive user interfaces since this user interface evoked higher feelings of presence, transportation and flow in the current study.

References

Barfield, W., Zeltzer, D., Sheridan, T., & Slater, M. (1995). Presence and performance within virtual environments. In W. Barfield, & T. A. Furness (Eds.), *Virtual environments and advanced interface design* (pp. 473-513). Oxford, United Kingdom: Oxford University Press.

Borrego, A., Latorre, J., Alcaniz, M., & Llorens, R. (2018). Comparison of Oculus Rift and HTC Vive: Feasibility for Virtual Reality-Based Exploration, Navigation, Exergaming, and Rehabilitation. *Games for health journal*, *7*(3), 151-156. https://doi.org/10.1089/g4h.2017.0114

Chaytor, N., & Schmitter-Edgecombe, M. (2003). The ecological validity of neuropsychological tests: A review of the literature on everyday cognitive skills. *Neuropsychology review*, *13*(4), 181-197. https://doi.org/10.1023/b:nerv.0000009483.91468.fb

Consortium Cognitieve Revalidatie (2007). Commissie CVA-Revalidatie: Richtlijn Cognitieve Revalidatie Niet-aangeboren Hersenletsel.

De Leo, G., Diggs, L. A., Radici, E., & Mastaglio, T. W. (2014). Measuring sense of presence and user characteristics to predict effective training in an online simulated virtual environment. *Simulation in Healthcare, 9*(1), 1-6. https://doi.org/10.1097/sih.0b013e3182a99dd9

Gorini, A., Capideville, C. S., De Leo, G., Mantovani, F., & Riva, G. (2011). The role of immersion and narrative in mediated presence: the virtual hospital experience. *Cyberpsychology, Behavior, and Social Networking*, *14*(3), 99-105. https://doi.org/10.1089/cyber.2010.0100

Hendriks, M. P., Kessels, R. P. C., Gorissen, M. E. E., Schmand, B. A., & Duits, A. A. (2014). *Neuropsychologische diagnostiek: De klinische praktijk*. Amsterdam, Nederland: Boom.

Horgan, N. F., O'regan, M., Cunningham, C. J., & Finn, A. M. (2009). Recovery after stroke: a 1-year profile. *Disability and rehabilitation*, *31*(10), 831-839. https://doi.org/10.1080/09638280802355072

Jokinen, H., Melkas, S., Ylikoski, R., Pohjasvaara, T., Kaste, M., Erkinjuntti, T., & Hietanen, M. (2015). Post-stroke cognitive impairment is common even after successful clinical recovery. *European Journal of Neurology*, *22*(9), 1288-1294. https://doi.org/10.1111/ene.12743

Kang, Y. J., Ku, J., Han, K., Kim, S. I., Yu, T. W., Lee, J. H., & Park, C. I. (2008). Development and clinical trial of virtual reality-based cognitive assessment in people with stroke: preliminary study. *CyberPsychology & Behavior*, *11*(3), 329-339. https://doi.org/10.1089/cpb.2007.0116

Lampton, D. R., Kolasinski, E. M., Knerr, B. W., Bliss, J. P., Bailey, J. H., & Witmer, B. G. (1994). Side effects and aftereffects of immersion in virtual environments. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *38*(18), 1154-1157. https://doi.org/10.1177/154193129403801802

Langhorne, P., Bernhardt, J., & Kwakkel, G. (2011). Stroke rehabilitation. *The Lancet*, *377*(9778), 1693-1702. https://doi.org/10.1016/S0140-6736(11)60325-5

Lee, J. H., Ku, J., Cho, W., Hahn, W. Y., Kim, I. Y., Lee, S. M., . . . Kim, S. I. (2003). A virtual reality system for the assessment and rehabilitation of the activities of daily living. *CyberPsychology & Behavior*, *6*(4), 383-388. https://doi.org/10.1089/109493103322278763

Lees, R., Selvarajah, J., Fenton, C., Pendlebury, S. T., Langhorne, P., Stott, D. J., & Quinn, T. J. (2014). Test accuracy of cognitive screening tests for diagnosis of dementia and multidomain cognitive impairment in stroke. *Stroke*, *45*(10), 3008-3018. https://doi.org/10.1161/STROKEAHA.114.005842

Lessiter, J., Freeman, J., Keogh, E., & Davidoff, J. (2001). A cross-media presence questionnaire: The ITC-Sense of Presence Inventory. *Presence: Teleoperators & Virtual Environments*, *10*(3), 282-297. https://doi.org/10.1162/105474601300343612

Linden, T., Skoog, I., Fagerberg, B., Steen, B., & Blomstrand, C. (2004). Cognitive impairment and dementia 20 months after stroke. *Neuroepidemiology*, *23*(1-2), 45-52. https://doi.org/10.1159/000073974

Lo Priore, C., Castelnuovo, G., Liccione, D., & Liccione, D. (2003). Experience with V-STORE: considerations on presence in virtual environments for effective neuropsychological rehabilitation of executive functions. *Cyberpsychology & behavior*, *6*(3), 281-287. https://doi.org/10.1089/109493103322011579

Ouellet, É., Boller, B., Corriveau-Lecavalier, N., Cloutier, S., & Belleville, S. (2018). The Virtual Shop: A new immersive virtual reality environment and scenario for the assessment of everyday memory. *Journal of neuroscience methods*, *303*, 126-135. https://doi.org/10.1016/j.jneumeth.2018.03.010

Parsons, T. D. (2015). Ecological validity in virtual reality-based neuropsychological assessment. In In Mehdi Khosrow-Pour (Eds.), *Encyclopedia of Information Science and Technology, Third Edition* (pp. 1006-1015). https://doi.org/10.4018/978-1-4666-5888-2.ch095

Price, M., Mehta, N., Tone, E. B., & Anderson, P. L. (2011). Does engagement with exposure yield better outcomes? Components of presence as a predictor of treatment response for virtual reality exposure therapy for social phobia. *Journal of anxiety disorders, 25*(6), 763-770. https://doi.org/10.1016/j.janxdis.2011.03.004

Rand, D., Kizony, R., Feintuch, U., Katz, N., Josman, N., Rizzo, A. S., & Weiss, P. L. (2005). Comparison of two VR platforms for rehabilitation: video capture versus HMD. *Presence: Teleoperators & Virtual Environments*, *14*(2), 147-160. https://doi.org /10.1162/1054746053967012

Rasquin, S. M., Lodder, J., Ponds, R. W., Winkens, I., Jolles, J., & Verhey, F. R. (2004). Cognitive functioning after stroke: a one-year follow-up study. *Dementia and geriatric cognitive disorders*, *18*(2), 138-144. https://doi.org/10.1159/000079193

Rizzo, A. S., & Kim, G. J. (2005). A SWOT analysis of the field of virtual reality rehabilitation and therapy. *Presence: Teleoperators & Virtual Environments*, *14*(2), 119-146. https://doi.org/10.1162/1054746053967094

Rizzo, A., & Koenig, S. T. (2017). Is clinical virtual reality ready for primetime?. *Neuropsychology*, *31*(8), 877-899. https://doi.org/10.1037/neu0000405

Santos, B. S., Dias, P., Pimentel, A., Baggerman, J. W., Ferreira, C., Silva, S., & Madeira, J. (2009). Head-mounted display versus desktop for 3D navigation in virtual reality: a user study. *Multimedia Tools and Applications*, *41*(1), 161-181.https://doi.org/10.1007/s11042-008-0223-2

Sayers, H. (2004). Desktop virtual environments: a study of navigation and age. *Interacting with computers*, *16*(5), 939-956. https://doi.org/10.1016/j.intcom.2004.05.003

Scheydt, S., Müller Staub, M., Frauenfelder, F., Nielsen, G. H., Behrens, J., & Needham, I. (2017). Sensory overload: a concept analysis. *International journal of mental health nursing*, *26*(2), 110-120. https://doi.org/10.1111/inm.12303

Schuemie, M. J., Van Der Straaten, P., Krijn, M., & Van Der Mast, C. A. (2001). Research on presence in virtual reality: A survey. *CyberPsychology & Behavior*, *4*(2), 183-201. https://doi.org/10.1089/109493101300117884

Shahrbanian, S., Ma, X., Aghaei, N., Korner-Bitensky, N., Moshiri, K., & Simmonds, M. J. (2012). Use of virtual reality (immersive vs. non immersive) for pain management in children and adults: A systematic review of evidence from randomized controlled trials. *European Journal of Experimental Biology*, *2*(5), 1408-1422.

Skilbeck, C. E., Wade, D. T., Hewer, R. L., & Wood, V. A. (1983). Recovery after stroke. *Journal of Neurology, Neurosurgery & Psychiatry*, *46*(1), 5-8. https://doi.org/10.1136/jnnp.46.1.5

[Specifics of HTC VIVE]. (z.d.). Consulted from: https://www.vive.com/us/product/vive-virtual-reality-system/

Spreij, L. A., Braaksma, S. W., Sluiter, D., Verheul, F. J. M., Visser-Meily, A., & Nijboer, T. C. W. (2017). Virtual Reality als potentiële aanvulling op de huidige neuropsychologische diagnostiek. *Tijdschrift voor neuropsychologie*, *12*(2), 73-96.

Tieri, G., Morone, G., Paolucci, S., & Iosa, M. (2018). Virtual reality in cognitive and motor rehabilitation: facts, fiction and fallacies. *Expert review of medical devices*, *15*(2), 107-117. https://doi.org/10.1080/17434440.2018.1425613

Verhage, F. (1965). Intelligence and age in a Dutch sample. *Human Development, 8*, 238–245. https://doi.org/10.1159/000270308

Verheul, F. J. M., Spreij, L. A., Rooij, N. D., Claessen, M. H. G., Visser-Meily, J. M. A., & Nijboer, T. C. W. (2016). Virtual Reality als behandeling in de cognitieve revalidatie. *Nederlands Tijdschrift voor Revalidatiegeneeskde, 2*, 47-53.

Wall, K. J., Isaacs, M. L., Copland, D. A., & Cumming, T. B. (2015). Assessing cognition after stroke. Who misses out? A systematic review. *International Journal of Stroke, 10*(5), 665-671. https://doi.org/10.1111/ijs.12506

Zedlitz, A. M. E. E., & Fasotti, L. (2010). Omgaan met beperkte belastbaarheid: Behandelprotocol voor ambulante hersenletselpatiënten met (ernstige) vermoeidheid. Amsterdam, Nederland: Boom