Looking Inside - Predicting physiological effects of fatigue using positional tracking and performance in a VR cancellation task for diagnosing egocentric neglect.

Hamzah Ziadeh
hziade18@student.aau.dk

Steffen Lehmann
slehma15@student.aau.dk

ABSTRACT

Research: Virtual reality diagnosis tests for neglect have higher sensitivity to neglect symptoms compared to traditional pen and paper tasks. However, neglect patients fatigue quickly in VR interactions, so predicting fatigue during diagnosis can create a motivating environment for patients as therapists can adjust task difficulty. This study aimed to predict physiological fatigue using positional tracking and performance data.

Method: 20 healthy young adults participated in single factor between subject experiment that compared their fatigue in two VR environments: 1) an empty white environment or 2) an environment with visual and auditory noise. Both conditions employed a whack a mole cancellation task with distractors for 10 minutes.

Results: Positional tracking successfully predicted fatigue using pupil diameter, head, hand, and eye movement. Environmental noise also effected participants' fatigue but to a small degree.

Conclusions: Future studies should employ narrative based VR environments and use tracking measures or performance over time as proxies for fatigue.

INTRODUCTION

Stroke survivors suffer from a number of cognitive disorders after their stroke, with 50% of survivors suffering from a disorder named Neglect. This paper focuses on egocentric neglect (referred to as neglect from this point), which causes patients to unintentionally omit one side of their their perception, causing them to e.g., collide with objects or miss cars incoming from one side of the road [40, 26, 27, 25].

Traditionally, therapists diagnose patients based on tasks done on pen and paper that assess neglect through the patients' performance [25, 10, 38]. However, pen and paper tasks have low ecological validity and do not accurately predict patients' behaviour in real life [3, 7, 40]. Simulating rich environments in Virtual Reality (VR) to test patients in real world contexts such as grocery shopping, street crossing, or dining yields more valid assessments of neglect compared to pen and paper tasks [16, 40, 28]. For example, the number of collisions with virtual objects in VR reflected patients' likelihood to collide with real world objects more accurately than pen and paper tasks. [3, 7].

Studies hypothesise that VR games achieve higher sensitivity towards neglect symptoms by requiring higher mental loads which increases the bias they have to one side of their perception [7, 30, 28, 39]. While high mental loads help diagnose neglect, prolonged periods may fatigue patients which may reduce their motivation to continue training [7, 40]. Therefore, studies noted the need to quantify patients' fatigue to help therapists adjust task difficultly according to patients' needs [3, 7, 28].

Many recent studies investigated mental effort by quantifying task demands (i.e. the number of concurrent tasks) [5, 23, 40] or investigated fatigue by observing patients' performance over time (i.e. a decrease in reaction time means higher fatigue) [18, 17, 3]. However, no studies on neglect inspect mental loads through physiological measures which strongly correlate with users' fatigue compared to most measures used in recent neglect studies [23, 13]. By incorporating concepts from physiological measures of fatigue into designing VR tasks for diagnosing neglect, this paper investigates users' mental states and the possibility of using machine learning methods to predict them.

BACKGROUND

We collected papers for this review by searching for keywords in research aggregators such as Google Scholar or the AUB search engine. In our first search, we used keywords such as "neglect", "VR", "Diagnostics", "Mental Effort", etc. We then selected papers based on their titles and abstracts. This yielded >50 studies which we narrowed down based on their relevance to our problem. This selected 11 studies which we considered as our core studies. After finding the gap regarding physiological measures, we conducted another search using the keywords "Physiological measures", "Heart rate variability", "fatigue", "Predicting Mental Load", etc. Using a similar process, we extracted 5 core papers from >10. Note that one of the writers wrote a portion of this synthesis as part of the "Foundations in Medialogy" exam. The text in italics (*such as this*) identifies content taken directly from that exam paper.

Most Pen and paper tasks as well as VR tasks require patients to mark, draw, or navigate to targets scattered across a paper or environment [10, 22, 8, 18]. Compared to those without neglect, neglect patients seeking targets took twice as long to spot targets and missed up to 60% compared to the 15% average of non neglect patients [8, 28, 19, 10, 22]. Many tasks added distractors that users avoid when finding targets to increase the mental load required from the task, increasing the number of omissions by 25% on the neglected side compared to conditions without distractors [18, 28, 8, 3, 10]. Some

VR tasks added non-task related environmental noise either visually (i.e. pedestrians walking or grocery shelves stocked with products) or auditory (i.e. dogs barking or cars honking) which studies hypothesise also raises the mental load required from a task [7, 40].

The majority of studies on neglect defined mental load as the amount of resources users commit from their working memory to find targets, whereas they defined fatigue as users' ability to find targets [7, 15, 19, 28]. Studies reported on mental load through task difficulty (i.e. the number of targets, distractors, or concurrent tasks) and quantified fatigue through performance measures (i.e. number of omitted targets or reaction time to flashing targets) [5, 28, 18]. See Table 1 for examples of mental load and fatigue measures. These studies hypothesised that neglect patients adapted to their fatigue by omitting one side to save mental resources, thus damaging their task performance [39].

Most studies investigating mental effort outside of neglect defined fatigue as the reduction of working memory resources over time resulting from high mental loads [23, 13, 24]. They quantified fatigue by comparing performance across different intervals during the task (i.e. performance in the first minute vs. last minute) [23]. Looking at task performance over time shows how users fatigued thus revealing the cost of mental load, i.e. what mental resources users sacrificed to accommodate their fatigue [23]. For healthy people, the cost may be slowing down their task performance [13] while costs for neglect patients can be increased omissions on the effected side over time [39].

Neglect patients performing cancellation tasks, where they located a number of targets hidden among distractors, took twice as long to find targets than healthy people [10, 18]. Studies hypothesised that neglect patients used more time likely because they employed different search strategies rather than fatigue [10, 22, 18]. Specifically, neglect patients look for targets away from the omitted side and only search within it when they can no longer find more targets, contributing to their longer search times [18, 10]. Recent cancellation studies using tablets and VR get around search strategies by having a static number of distractors on screen where one distractor temporarily transforms into a target for users to hit before going back to being a distractor, similar to a whack-a-mole (WAM) game [17, 19]. By prompting users to search in specific locations at specific times, we can observe the effects of mental load on their reaction time and errors without the confounds of search strategies [18, 17].

Overall, studies investigating the design of VR tasks to diagnose neglect recognised that current methods for quantifying fatigue do not yield accurate data to describe patients' mental state [18, 28, 17, 40]. Moreover, the lack of methods also means that studies could not accurately evaluate their tasks' effect on users [11]. This highlights a need to investigate how VR tasks fatigue patients and attempt to develop an understanding of how we can predict fatigue both to evaluate design and to provide data to therapists.

Virtual Reality

The majority of VR research focuses on navigation or steering tasks where users avoided distractors in their way while approaching a target [3, 40, 28, 38]. Few investigated exploratory scenes where patients freely explored an environment without a goal [15, 28]. However, recent studies investigated VR cancellation tasks as they provide a simple way of modulating difficulty through increasing distractors and/or targets [18, 17]. However, these studies have not investigated adding noise to their virtual environments but hypothesised that visual noise between targets and distractors requires higher mental loads compared to empty environments [19, 17].

Having a goal to complete requires higher mental load from patients as they constantly assess the current state of the task and their next actions [15, 7, 38, 28]. Exploratory scenes where patients freely looked at targets in a virtual museum did not see a larger bias to one side from neglect patients compared to those seen in paper tasks [28, 7]. Additionally, tasks where patients actively participated in the environment demanded higher mental loads than tasks where patients reacted to stimuli around them [7]. For example, a navigation task where patients stopped at road signs increased the number of missed signs on one side when patients steered their wheelchair through the environment compared to when the wheelchair steered itself [7].

Neglect patients can fatigue almost twice as fast during VR tasks compared to pen and paper tasks [40, 28, 39]. The cost of fatigue for these patients can also be high after the VR interaction causing them to collide with more objects in the real world after training [3]. While the higher sensitivity from fatigue helps diagnose patients, therapists need information on the state of the patient during the VR task to balance the difficulty of the task with the accuracy of diagnosis [30, 38, 22]

Measuring Fatigue

We identified 3 categories of data for measuring Fatigue; Performance, tracking, and physiological measures. Performance measures quantify user' progress or proficiency at a task through e.g. accuracy, number of mistakes, and completion time [40, 3, 22]. Tracking measures in VR consist of extracting positional data from the VR controllers, headset, and eye trackers inside the headset to measure the speed, frequency, and distance of users' movements which studies hypothesised decrease when they feel fatigued [4, 34, 17, 15]. Finally, physiological measures include attaching sensors to users' bodies to measure physiological responses of fatigue such as Electrodermal levels (EDL), increased heart rate, or Brain activation measured through Electroencephalography (EEG) which studies used as proxies for fatigue [23, 13, 24]. See examples of these measurements in Table 1.

While performance (the most widely used measure) correlates with the mental demands of a task and generally decreases over time, studies noted that performance measures can not sustain their sensitivity to fatigue for an extended period of training [18, 28, 3, 23]. For example, both healthy users and neglect patients learned how to better perform VR tasks during training

They used performance over time as fatigue

Author	Media	Distractors (n)	Environmental noise	Auditory (A) Visual (V)	Measures	Mental load (M)/ Fatigue (F) measure
Baheux2004 [6]	VR	N/A	Revolving sushi table	V	- Eye-tracking - Eye-gaze position - Position - Head Movement	N/A
Blini2016 [5]	Screen	Condition 1: extra symbol Condition 2: Noise	N/A	N/A	- Accuracy in detection	M=Number of concurrent tasks F=Decrease in task performance
Buxbaum2012 [7]	VR	N/A	bushes, fountains, benches	A, V	- Targets found	N/A
Hougaard2021 [15]	VR	N/A	N/A	N/A	- Head orientation - Gaze - Saccades - Gaze time	N/A
Knobel2020 [18]	VR	Square (n = 100)	N/A	N/A	- Time taken - Accuracy	N/A
Knoche2016 [19]	Tablet	Red moles	N/A	N/A	- Fitt's a and b - Performance - Time taken	M=Number of targets
Liang2010 [22]	Screen	N/A	N/A	N/A	76 Variables: - Performance - Completion time - Time spent on left side	F=self-report
Ogourtsova2016 [28]	VR	Objects on shelves	N/A	N/A	- Time taken - Path taken - Success rate	M=Number of targets F=missed targets
Wagner2020 [40]	VR	Cars	City and bird noise, Pedestrians	A, V	- Time taken - Errors - Head direction	N/A
Wilson2018 [23]	Screen	Extra planes (n = unknown)	Music	A	- self-report - Passive BCI	F=performance F=self-report F=fNIRS (passive BCI)
Kim2021 [17]	VR	Spheres (n=96)	N/A	N/A	- Eye-tracking - Position - Performance	N

Table 1: The media column describes the artefact used for the task. The distractors column highlights the visual representation and amount of distractors. The environmental noise column explains how the noise was presented in the environment. The Auditory (A), visual (V) column refers to the modality of the noise. The Measure column provides an overview of the measures used. The Mental load / Fatigue measure column refers to the measures used to quantify mental load / fatigue.

VR Cancellation task

Demeyere & Gillebert (2019)

sessions exceeding 3 minutes and did not show a decrease in performance despite reporting fatigue after training [28, 39, 23]. Moreover, other users may have high tolerances, meaning they feel fatigue but maintain high task performance [23, 22].

Physiological measures such as HRV and EDL acted as accurate proxies for fatigue in healthy users during digital puzzle games similar to cancellation tasks [9, 13, 12]. HRV contains two measurements, sdnn and rmssd which describe the deviations in the time between each heartbeat which decreases during fatigue [33, 9]. EDL measures the electrical conductance from the sweat glands which decreases as users fatigue and experience stress under high mental loads [9, 24]. However, neither measure has been used to predict fatigue during VR training for neglect patients [13, 23]. While both measures receive noise from emotional responses (i.e. during failed input attempts), they still act as the most accurate proxies to describe fatigue levels in healthy subjects [13, 33, 20].

As seen in Table 1, most studies using tracking measures opted for eye tracking in VR. Fatigue lessens the amount of time users spend looking around their environment, leading to a fewer number of saccades; the small rapid eye movements users make to fixate on a point [9]. Similarly, users' gaze direction also remains stable, spending longer times with their head oriented in one direction [6, 7]. Neglect patients in particular shift their gaze away from the omitted side as they fatigue, increasing their bias to one side [15, 7]. However, no studies on VR tasks for diagnosing neglect used pupil diameter, which expands during fatigue [9]. Pupil diameter also correlates strongly with heart rate ($r=0.8$, $p<0.001$) meaning it can provide useful insights into patients' state during training [9].

According to physiological measures, a cancellation task performed with a mouse on a computer screen, healthy users begin to fatigue after about 5 minutes of interaction [13, 33]. However, most studies on neglect patients range between one to three minutes [18, 40, 3, 17] with the longest being 5 minutes [28]. These task lengths may be too short to induce fatigue in patients so any decrease in performance measures or change in physiological measures can likely be explained by an emotional response as opposed to fatigue as studies hypothesised [11]. Studies on physiological measures suggest collecting data over longer periods of time, ranging from 7 minutes and up to 30 minutes to collect data that shows fatigue [33]. While more mentally demanding tasks fatigue users more intensely [9, 24], no reliable measure of task difficulty indicates how long users need to perform a task before fatiguing [9]. As a result, extending task length up to 10 minutes increases the chances of collecting data that reflects the patients' fatigue [33].

RESEARCH GOAL

This study aimed to investigate the possibility of predicting users' fatigue through tracking and performance measures during a VR interaction designed to test for neglect. Additionally, the study also evaluated if the addition of environmental noise to the training game fatigued users more than an empty scene by requiring more mental load.

METHOD

To investigate fatigue during task performance, users played 10 minutes of a VR cancellation task in a single factor between subjects study. Heart rate and EDL sensors acted as proxies for users' fatigue during the interaction. while the VR headset and controllers provided the positions of their head, hands, and eyes to extract data about users' movements during the interaction. Finally, a program in Unity logged their task performance.

Design

Users performed the WAM cancellation task seated with the targets and distractors spread out in front of them. We modified a prototype of a VR WAM game [2] similar to a design by Knoche et al [19]. The task contained 96 grey moles visible at all times to the user. During the task, one mole turned green to signify becoming a target and another mole turned red to distract the user, see Figure 1 for examples. Users aimed using a VR controller and pushed its back trigger to shoot moles. Each target remained green for 5 seconds, which we refer to as the target's lifetime. If the user failed to shoot the target within its lifetime, the target turned back to grey and new target and disctractor moles appeared immediately. If the user shot a grey mole or into an empty space, a buzzer sound signified failure. Shooting a red distractor played the same buzzer sound and turned the red mole grey. In both cases, users could still shoot the target if its timer has not run out. Finally, if the users shot a target within its lifetime, a bell sound signified success and new target and distractor moles appeared immediately.

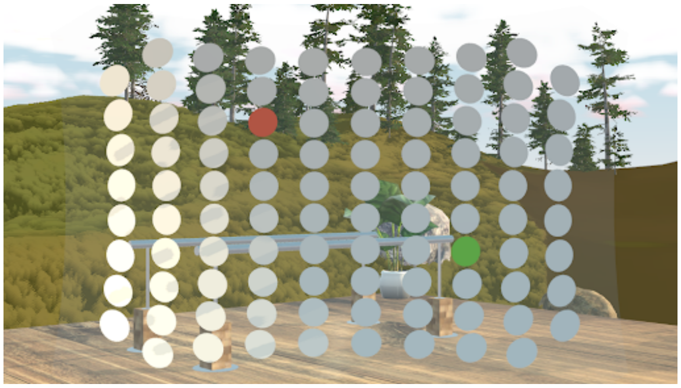
Users saw a blue circle indicating where their controller pointed to use for aiming. An invisible wall filled the gaps between the moles so users still saw the aiming circle even if they aimed at a gap in between moles. Users played in one of two possible virtual environments, an empty white world and an environment with visual and auditory noise as seen in sub-figure 1b. Users in the noisy environment completed the task in a room with two large windows overseeing a field with trees, grass, and clouds acting as visual noise as seen in sub-figure 1a. The trees, grass, and clouds continuously moved throughout the task to simulate being blown by the wind with matching audio of wind blowing adding auditory noise. The trees also cast light moving shadows on the moles. The mixture of visual and auditory noise aimed to distract the user during the task. In both conditions, users could see the environment around them and in between the moles as well.

Participants

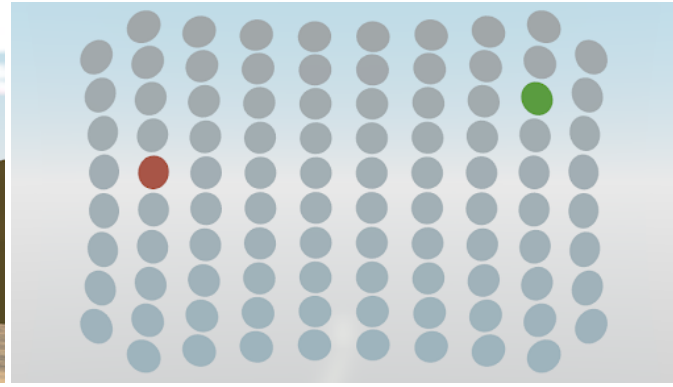
20 healthy subjects (8 female, 12 male) participated in the experiment ranging from 21 to 32 years old ($m=24$, $sd=2.6$). All but two had prior experience with VR and four had prior experience with the WAM game. Only one participant was left handed. Additionally, one stroke survivor (female, age = +60) with no diagnosed neglect participated in the study. She had prior VR and whack-a-mole experience and suffered from paralysis in her left arm and leg.

Pilot Study

Before the experiment, the aforementioned stroke patient participated in two pilot studies to provide input on our design.



(a) Participants' Point of view in the NE condition.



(b) Participants' Point of view in the ME condition.

Figure 1: Overview of the point of view from each condition in the experiment.

We conducted both pilot studies at her home and set up a camera to capture video and audio of the session to analyse it afterwards. For the first pilot study, she played the original prototype of the WAM game as seen in Figure 1b and provided her opinions in a debrief session. The patient played 3 sessions of the game with various board sizes to investigate which dimensions worked best for stationary interactions with patients who have little movement in their upper body. Each session lasted for 3 minutes with no break in between. The prototype in this pilot study did not summon new targets and distractors directly after the user shot the target. Instead, the game waited for the remainder of that target's lifetime before creating a new one. This allowed users to take a 1-2 second break between hits.

The patient reported feeling challenged but explained that it felt easy enough that she did not feel fatigued. The patient had little upper body movement and experienced difficulties reaching the bottom moles from her position in all sessions. Moreover, we found that the set up process of the WAM game did not suit a seated patient. The game required users to position themselves at a specific point in front of the board and did not account for users' height. The patient could not position herself correctly as she could not make fine movements from her wheelchair which also positioned her too high up to reach the bottom of the board. Before the second session, we addressed the setup problems by adding a step to our procedure to teleport and reorient users into the correct place in VR before the game started. This step also changed the users' heights to position their heads at level with the centre of the board. This ensured that all users played the game at the same coordinates and had similar abilities to see and reach all moles (not accounting for arm length). In the second pilot study, the patient played the finished WAM game with the environment seen in Figure 1a. The patient played the game once for a 7 minute session followed by another debrief. Even though both pilot studies had the same target lifetime (3 seconds), the patient could not hit most targets on time as a new one appeared right after a hit. The patient explained that that this change sped the game up too much. She expressed seeing the targets as they appeared but could not physically get there fast enough. We stopped the game after 3 minutes with mostly missed targets. We tried different lifetimes for the moles and the patients noted that a 5 second lifetime felt difficult but gave her enough time to move the controller across the board. Initially, the patient also found it difficult to see the aiming circle with the added visual noise but adjusted after 2 minutes. Finally, the patient stated not feeling fatigued from playing the game and that it could be more difficult or longer. We addressed this final point by extending the playtime to 10 minutes in our experiment.

Procedure

To start the experiment, participants read an instruction letter outlining the procedure of the experiment, the sensors we equipped on them, and a brief explanation of WAM and the different types of moles. Afterwards, they signed a consent form and filled out a mood questionnaire measuring their motivation and nervousness to play the game based off of Skola's mood questionnaire [35]. The participants then equipped and adjusted the VR headset (HTC vive) themselves and alerted us when they felt comfortable. Participants saw a live video of their eyes from inside the headset and adjusted the headset once more so that both pupils were in centre frame as seen in Figure 3. After verifying that the eye tracking program (Pupil Core headset) tracked both eyes, a facilitator asked participants to extend their non-dominant hand for the physiological sensors. The facilitator explained what sensor they attached as they equipped the heart rate sensor to the index finger and the EDL sensors to the middle and ring fingers. The participants then placed their hand on their lap or on a table next to them to keep it stationary.

Once participants expressed feeling comfortable, we ran an eye tracking calibration program where users tracked a teleporting ball with their eyes. Once calibration succeeded, we placed the users in the correct condition (noisy environment (NE) vs. muted environment (ME)) and asked them to sit in a position they felt comfortable before we teleported them into the correct position in VR. Afterwards, participants played 10 minutes and 30 seconds of the WAM game with their dominant hand. Targets had a lifetime of 5 seconds each and users played until time ran out. For the first 30 seconds, users kept their head and hands still while no targets or distractors appeared to

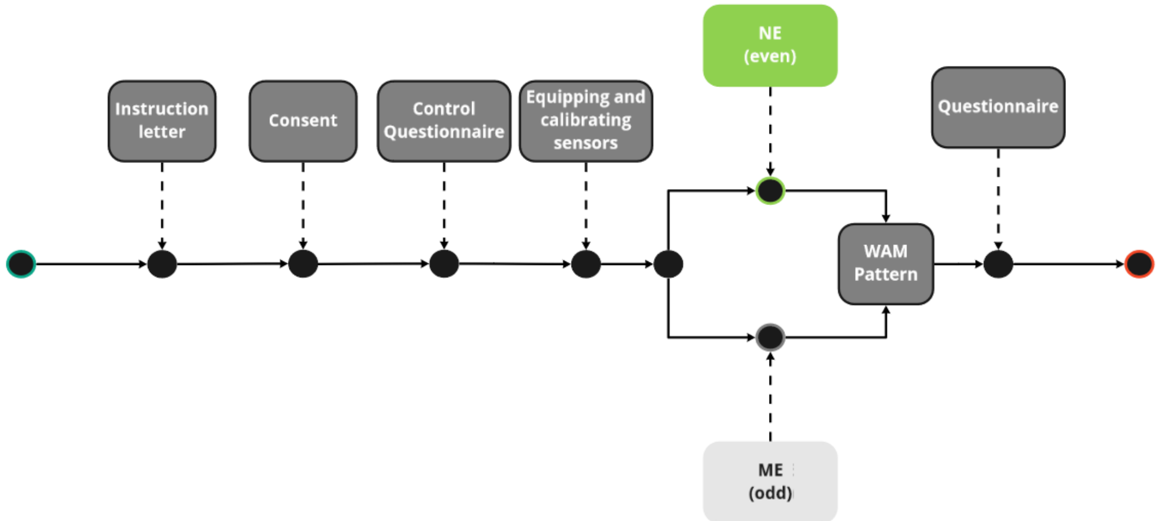


Figure 2: The flow of experimental procedure.

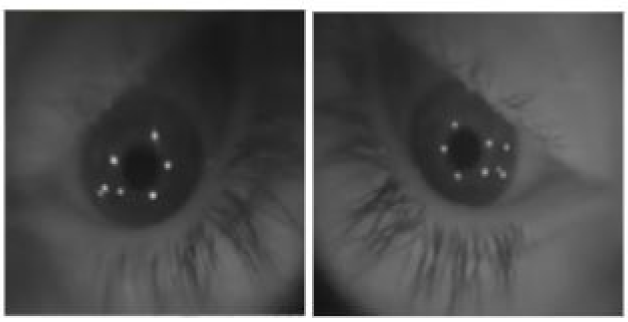


Figure 3: The live feed of each eye shown in the unity scene during the eye-tracking setup. The feed disappeared when the calibration started.

capture a baseline for their physiological data and the noise in the tracking measures caused by jitter. After 30 seconds, targets and distractors appeared for participants to shoot. All participants played the same pattern of moles, meaning the order and positioning of targets and distractors remained the same for all participants. When participants finished the 10 minutes of gameplay, they filled out a revised version of the iso 9241-11 questionnaire using a 7-point likert scale. The questions were:

- The smoothness of aiming was
- The force required for shooting was
- The mental effort required for aiming was
- The physical effort required for aiming was
- Aiming at the correct mole was
- The movement speed of the aiming cross was
- Finger fatigue
- Wrist fatigue
- Arm fatigue
- Shoulder fatigue
- Neck fatigue
- General comfort
- Overall, the game was

Measurements

Both EDL and heart rate senors sampled at 100Hz. The heart rate sensor measured the time between each detected heart beat, known as the Inter-beat Interval (IBI). The IBI data provides both the sdnn and rmssd. We also extracted the LF and HF band to which provides data about users' respiration during the interaction. The package hrv-analysis [1] for python provided all heart rate measures. Finally, we also extract the maximum and minimum heart rate values from the IBI data.

The x, y, and z values of users' head, hand, and eye movement sampled at 60Hz provided the distances between users' movement as well as any rotations they made to their head and controller. The saccades package [37] in R provided measurements for the amount of saccades as well as their duration and velocity. Finally, we extracted measurements of users' pupils in millimeters sampled at 350Hz from the eye tracking data.

DATA ANALYSIS

We used Wilcoxon signed rank tests on the questionnaire results to compare the two conditions. We also compared questionnaire results using control variables to investigate if previous experience impacted their questionnaire results. Finally, we used Pearson correlation coefficients to check for correlations in questionnaire answers.

A sliding average of 5 samples removed noise from the the EDL data. We filtered the IBI data by removing all samples below 400ms and above 1,200ms. A sliding window of 30 seconds extracted time domain features from the IBI data such as sdnn, rmssd, minimum, and maximum heart rate. A one minute sliding window extracted the frequency measures which yileded the LF/HF respiratory bands. The window length for the heart measures followed the recommendations of Shaffer et al. [33]. Afterwards, we resampled all data to 1 second bins, each aggregated to a sum, mean, minimum value, maximum value, and standard deviation. If two features correlated strongly (Pearson's r>0.9), we removed one at random to avoid overfitting our models.

We conducted a Principal Component Analysis (PCA) on the data after standardisation to reduce the dimensionality of the data and explore trends in the data set. The Lazypredict package [29] for Python ran multiple regression models on both the reduced data from the PCA and on the complete standardised data to find the best performing model. We used linear mixed models and t-tests for all other analysis to compare conditions.

To compare our conditions and inspect the effects of fatigue over time, we constructed one data set with the physiological measures and another one with the performance and tracking measures. Each data set yielded two subsets, one grouped by the condition which contains the condition label, and another data set where we extracted the data from only the first and last 60 seconds of gameplay and labeled them with what we call the time label. Note that for all results on heart rate measures excluded three participants (two from the NE condition) as a result of faulty data from moving the sensors during the interaction.

A linear Support Vector Machine (SVM) classifier using k-fold validation (k=8) trained and classified all the aforementioned subsets to test the differences that our conditions had on participants as well the effect of playing the game for 10 minutes. Finally, we used linear mixed models and t-tests on the features with the largest weights to investigate the differences. Note that all the analysis includes data from the stroke patient.

RESULTS

The Wilcoxon tests showed no differences between conditions from both control and mood questionnaires. Participants also rated their mental fatigue very low in the interaction (m=2, sd=1.02). Overall, the questionnaires contained no significant correlations or differences.

Regression on Fatigue Proxies

Out of 42 models, an Extra Trees regression model predicted our fatigue proxies most accurately. See Table 2 for the fatigue variables our model predicted with >50% accuracy. In all predicted variables, pupil diameter measures contributed the most to predictions in terms of coefficient size compared to other tracking measures of users' movements. All Performance measures coefficients remained at 0 and did not contribute to the model. Overall, using exclusively tracking measures we predicted 81% of EDL variance and 72% of the maximum heart rate variance. When trained on data from our healthy participants and tested on data collected from our stroke patient participant, the model achieved similar accuracy in all variables (EDA: R^2=0.8, RMSE=0.09, Max HR: R^2=0.65, RMSE=0.05).

Principal Component Analysis

Running a PCA on our tracking and performance measures (n=38) showed that the model explained >90% of the variance in our data using only 6 components as seen in Table 3, revealing many redundancies in the data. Isolating the features that contributed most to these components yielded the factor loading matrix shown in Figure 4. The figure depicts each feature and its correlation to the principal components, thus showing their contributions to the analysis. Similar to our regression analysis, the results show that tracking measures, specifically those related to pupil diameter, explain more of the variance in our data compared to performance measures. Isolating the data collected from the final minute of interaction showed that total head rotation per second contributed the most to one component that explained 91% of the variance. Finally, three principal components structured from the physiological data each explained 30% of the variance (sum=90%). Rmssd, IBI, and EDL respectively contributed the most to the aforementioned components.

Significant Predictors	**Predicted Variable**				
	EDL	IBI	Max HR	Min HR	RMSSD
Mean Pupil Diameter	0.18	0.14	0.13	0.17	0.11
SD Pupil Diameter	0.12	0.06	0.08	0.07	0.09
Max Pupil Diameter	0.08	0.04	0.04	0.04	0.06
Mean Head Movement	0.06	0.04	0.05	0.05	0.06
Mean Eye Movement	0.03	0.08	0.07	0.05	0.03
Sum Hand Movement	0.05	0.04	0.04	0.04	0.04
$\mathbf{R^2}$	0.81	0.66	0.72	0.7	0.6
RMSE	0.08	0.09	0.09	0.1	0.06

Table 2: Predicted Variables and all significant predictors according to the extra trees model. Values depict each predictor's contribution to the model.

Component	Explained Variance	Cumulative Sum
PC1	20%	20%
PC2	16%	36%
PC3	14%	50%
PC4	14%	64%
PC5	13%	78%
PC6	13%	91%

Table 3: Results from the PCA on data without physiological measures

Effects of Environmental Noise on Participants

An SVM classifier training on physiological measures predicted the condition label with an accuracy of only 47% (sd=0.13). In contrast, the classifier trained only on the last minute of physiological data achieved 68% average accuracy (sd=0.03) when predicting condition in contrast to the 54% average accuracy (sd=0.3) of training on the first minute data.

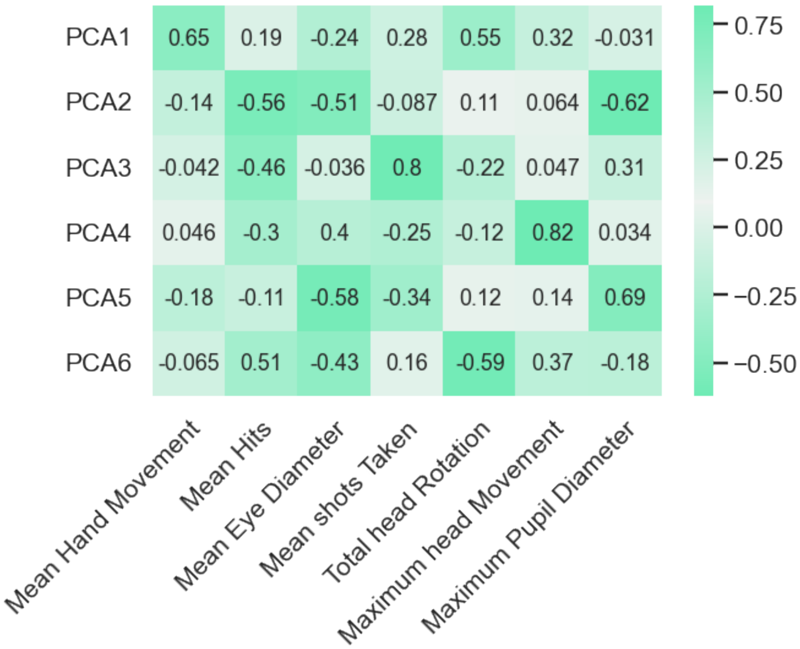


Figure 4: A factor loading matrix with the most significant variables and their correlations with the principal components.

In a multiple regression model, the condition ($\beta = -99.4$, p=0.02) predicted EDL (R^2=0.16, F(2, 39)=3.82, p=0.03) while the minute label did not (p=0.15). A t-test showed that participants had higher EDL (t=-1.93, df=33.84, p=0.06) in the ME condition (m=608.72, sd=148.38) compared to the NE condition (m=516.57, sd=138.62), indicating higher fatigue in the presence of Noise. Condition ($\beta = -23.3$, p=0.04) also predicted rmssd (R^2=0.26, F(4, 31)=2.79, p=0.04) but only when introducing an interaction effect ($\beta = -191.9$, p=0.05) between condition and mean target hits, which correlated with rmssd (r=-0.4, p=0.01). See Figure 5 for an illustration of this effect.

When trained on performance and tracking measures, the SVM predicted the condition label with on average 94% accuracy (sd=0.03). The accuracy dropped to 90% (sd=0.04) when trained only on the data from the last minute and down to 84% (sd=0.19) when trained on the first minute data. In all cases, the SVM assigned the highest weights to tracking measures in place of performance measures similar to the principal components.

A t-test confirmed that participants' smallest hand movements per second increased (t=6.17, df=27.02, p<0.0001) in the NE condition (m=0.38, sd=0.1) compared to without noise (m=0.1, sd=0.1). Likewise, a multiple linear regression showed that both condition ($\beta = -0.28$, p<0.001) and time labels ($\beta = -0.01$, p=0.002) predicted participants' hand movements (R^2=0.59, F(2, 39)=27.49, p<0.001). The smallest head rotations per second also increased (t=6.83, df=35.05, p<0.0001) in NE condition (m=0.04, sd=0.01) compared to the ME condition (m=0.01, sd=0.01) with the condition ($\beta = -0.02$, p<0.001) and time ($\beta = -0.002$, p<0.001) labels also predicting participants' head rotation (R^2=0.73, F(2, 39)=51.47, p<0.001). Finally, only the condition label ($\beta = -0.02$, p<0.001) predicted the minimum eye movements (R^2=0.27, F(2, 39)=7.4, p<0.001) which also increased (t=4.08, df=21.9, p<0.0001) in the NE condition (m=0.03, sd=0.003) compared to ME (m=0.004, sd=0.0004).

Overall, the results show no significant changes in pupil diameter across conditions, the feature that most strongly predicted our fatigue proxies. However, participants perform larger movements when presented with environmental noise.

Effects of Play Time on Participants

When trained to assign the time labels to features from the physiological data, the SVM classifier yielded an average accuracy of 70% (sd=0.18). Training and testing on data from the ME condition lowers the accuracy to 60% (sd=0.23) and down to 52% (sd=0.19) when trained and tested on the NE condition.

The time label predicted sdnn ($\beta = -1.48$, p=0.005) in a multiple regression model (R^2=0.21, F(2, 33)=4.45, p=0.02) while condition (p=0.81) did not. Time label ($\beta = -2$, p=0.01) also predicted 17% of the variance in rmssd (R^2=0.13, F(2, 33)=3.42, p=0.04). A t-test showed that rmssd decreased significantly (t=-2.21, df=27.67, p=0.04) from the first minute (m=42.4, sd=21) to the final minute (m=29.6, sd=12.4), showing that participants fatigued over time. An introduced interaction effect ($\beta = 83.76$, p=0.03) between the time label ($\beta = -12.58$, p=0.03) and mean target hits (p=0.4) predicted the minimum heart rate (R^2=0.24, F(4, 31)=2.45, p=0.06).

An SVM classifying performance and tracking measures using the time labels achieved 98% accuracy on average (sd=0.05) and achieved equal performance when training on data from each condition individually. In all models, performance measures contained the highest weights compared to tracking measures.

A t-test (t=-14.4, df=20.09, p<0.001) showed that participants hit moles longer into their lifetime in the last minute (m=1s, sd=0.01) compared to the first minute (m=0.3s, sd=0.21). The time label ($\beta = 0.07$, p<0.001) also predicted the target lifetime (R^2=0.84, F(2, 39)=101.1, p<0.001) while condition did not in a multiple regression model. The average targets hits per second was also significantly predicted (R^2=0.82, F(2, 39)=90, p<0.001) by the time label ($\beta = 0.01$, p<0.001) and increased (t=-13.52, df=21.8, p<0.001) from the first (m=0.05, sd=0.16) to last minute (m=0.15, sd=0.16), indicating a learning effect. Participants also took more shots (t=-13.12, df=38.84, p<0.001) on average per minute (first: m=0.28, sd= 0.04, last: m=1.04, sd=0.09) which the time label ($\beta = 0.01$, p<0.001) predicted significantly (R^2=0.9, F(2, 39)=90, p<0.001) in a multiple regression model. However, shooting more times also lead to a higher number of misses (t=-4.11, df=27.82, p<0.001) during the last minute (first: m=0.04, sd=0.04, last: m=0.1, sd=0.09), also predicted (R^2=0.3, F(2, 39)=8.36, p<0.001) by the time label ($\beta = 0.01$, p<0.001).

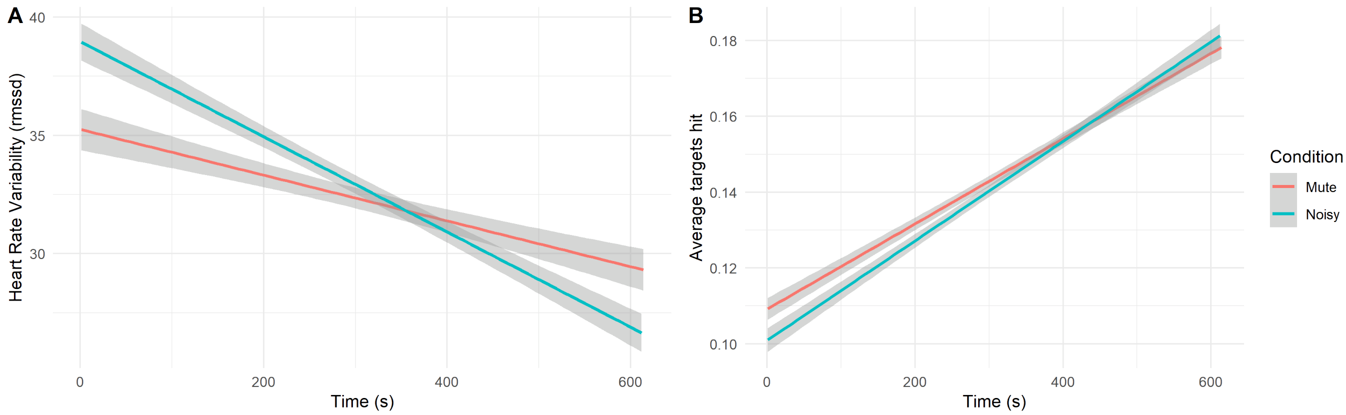


Figure 5: **A:** shows the linear trend of the average HRV over the 600 seconds of gamplay. **B:** shows the average amount of moles participants hit per second of the 600 seconds of gameplay.

Comparing stroke patient with healthy participants

On average, healthy participants EDL increased (first: m=550, sd= 124, last: m=618, sd=116) during the interaction indicating higher arousal whereas the patients' EDL decreased first: m=219, last: m=181), indicating a sign of fatigue. Both the patient and the healthy subjects experienced decreased HRV (rmssd) as a result of fatigue. However, the stroke patient experienced a much stronger decrease over time (first: m=48.4, last: m=17.7), likely meaning she felt more tired than the healthy subjects (first: m=42, sd=21.6, last: m=30.3, sd=12.5).

The patient started out with little eye movement, but eventually increased the size of the movements (first: m=81, last: m=12.1) closer to the amount of movements our healthy subjects made (first: m=8.29, sd=4.06, last: m=9.36, sd=4.12). We also observe the same effect in the total distance of hand movements during the interaction where both the patient (first: m=0.1, last: m=0.22) and healthy participants (first: m=0.17, sd=0.08, last: m=0.39, sd=0.12) increased the size of their movements.

The patient also shot at targets less times than (first: m=0.1, last: m=0.55) healthy participants (first: m=0.28, sd=0.18, last: m=1.05, sd=0.18), both of which increased the frequency of their shooting over time. The patient missed more shots (first: m=0.55, last: m=0.02) than healthy subjects (first: m=0.03, sd=0.18, last: m=0.13, sd=0.09) but the patient became more accurate over time while healthy subjects made more misses. Moreover, Both healthy participants (first: m=0.24, sd=0.17, last: m=0.92, sd=0.14) and the patient (first: m=0.02, last: m=0.53) hit more targets over time although the patient hit significantly less targets.

Qualitative Results

After the stroke patient finished the experiment, we conducted a short interview to unpack her experience of playing the game and participating in our project. She stated the she did not feel fatigued after the game and the she did not find it difficult. She explained that she could get to most targets before they disappeared but fell short if her initial shot missed.

She mentioned that she enjoyed the sounds and the trees moving with the wind. However, she explained that she found the Environment (in the NE condition) "black and white". As an example she suggested adding harsh colours to the pillows and either pieces of decoration in the room similar to those in her own home.

When asked about her motivation to participate in our project, she cited self improvement and interest in the research. Self improvement refers to her own mental improvements *"I enjoy it because I get to focus on the left side which I am very bad at. Often when I cannot find something, it is because I forgot to look over to the left side"*.

Moreover, she also wanted to help due to her disability and understood that she provided valuable feedback on the game as someone from our target group. When asked about what she though her impact on the project was, she responded: *"The influence I have is that I can tell you when something is not good, or when something is wrong or not wrong because I do not know anything about it, but difficult. I can only give the input, I can"*. Moreover, she expressed multiple times a desire to get feedback on the project after completion. She mentioned another projects she was apart of where she did not hear from them again which she described as *"boring"*.

DISCUSSION

Our study succeeded in predicting a number of fatigue proxies through tracking measures of users' head, hand, and eye movements during a VR cancellation task as seen in Table 2. The findings from our regression and PCA suggest that while performance can contribute to predicting fatigue, tracking measures provide more accurate data. Specifically, tracking pupil diameter and eye movements proved very useful as suggested by Charles et al. [9]. However, the interaction effect observed between target hits and rmssd suggests that studies can still use performance measures if the task demands higher mental loads by using noisy virtual environments. Although our study observes a performance-fatigue relationship that runs counter to most studies. Our setting showed that participants performed better as they fatigued (as indicated by HRV) as opposed to the hypothesis that they perform worse [28, 18, 17, 3].

Only two other studies using healthy subjects support this finding who also observed improved performance in puzzles such as the stroop task improving despite showing signs of fatigue (with EDL and HRV as proxies) [14, 36]. A possible explanation for this effect can be the learning effect, where users achieved higher performance despite their fatigue because they learned how to play better. The only other neglect study that investigated performance over time also found a learning effect that ran counter to assumptions that performance should worsen with fatigue [7]. Our results support this hypothesis by showing that participants moved more during the final minutes of interaction, showing that they reduced the mental effort of making fine movements to aim by making large movements and shooting more often to hit the moles at the cost of more misses and taking longer to hit moles. This suggests that future studies on cancellation tasks should use the users' accuracy while shooting targets as opposed to only how many targets they found. Moreover, the lifetime of targets proved a good indicator of fatigue which supports the WAM design of a cancellation task as opposed to the classical design that contains a static number of targets always present.

The only to other studies attempting to predict fatigue through performance measures achieved much lower performance than the models we presented in our papers. Gjoreski [12] labeled heart rate data as "high fatigue" or "low fatigue" according to whether or not participants rested or performed a task. Their model only achieved 60% accuracy when classifying which group the data belonged to. Heaton [14] achieved 41% using regression to predict EDL from performance, supporting our claim that performance measures may not be enough to predict fatigue.

Our study also showed that adding environmental noise leads to higher levels of fatigue during interaction, albeit with a small effect. A likely explanation for the small effect could be that our environment and task lacked narrative elements. Ogourstova [28] observed a much larger effect of adding environmental noise on performance and Wagner [40] found a much closer correlation of performance in their system to patients' real life behaviour compared to pen and paper tasks. Both studies contained a story to explain the task and the users' presence in the environment which included congruent noise and a target motivated by the story. Future studies should validate this hypothesis by comparing narrative tasks with unmotivated tasks such as our setup. However, our noise condition significantly increased the explorations and movements users made during gameplay. This finding may prove useful for programs aiming to rehabilitate neglect patients by training them to engage with the side effected by their condition as our stroke patient participant also stated. Future studies also need to validate these findings by investigating if neglect patients show a decreased bias in a longer task such as ours compared to those in previous studies [28, 40].

The stroke patient who participated in our study criticised others for including her in their research but never providing updates on the findings. An ethongraphic study on stroke patients who participated in HCI studies by Kögel [21] also found the same sentiments. Those patients describe feeling discarded by the researchers who never contacted them again. While these patients cited participating with the intent of self improvement just like ours, stroke patients participate in research to gain a sense of agency over their lives by contributing to society despite their disabilities [21]. Pink [31] also provided the same criticisms and considered it an ethical violation to not update participants on results especially those who belong to exposed groups like patients or those who made large time commitments. Therefor, after project completion, we intend to send our stroke participant a brief video explaining our project in more depth as well as summarising our findings. Additionally, we also plan on sending information regarding her performance in the game which she expressed a curiosity in.

Our study did not include a measure for emotions, specifically arousal. According to Fairclough [11], this may be a potential limitation of our study as physiological measure can be confounded by users' emotional reactions. By not having measures of emotional arousal, we can not explain how much users' emotions and mood impacted the physiological measures. However, fatigue often correlates with low arousal, which leads to a decrease in EDL over time that we did observe in our study [9, 14]. Regardless, future studies should employ methods such as interviews or self reporting to account for any moments where users' arousal significantly changed. For example, in a more difficult system where users make many errors, participants may become frustrated which may impact physiological and tracking data [11].

Additionally, data from our stroke patient shows she had worse performance at the beginning of the game compared to our healthy subjects. However, the data from the final minute of interaction shows that her performance increased to levels much closer to the healthy subjects. This can be explained by the patient not having much VR experience and needing to train and learn whereas our healthy participants felt more comfortable. This suggests that studies with shorter tasks who observed worse performance from cognitively impaired patients compared to healthier subjects may see improvements after extending the task [40, 18, 3]. A study by Rodil [32] suggests that when introducing VR to a new context with inexperienced users, users should have the chance to train and familiarise themselves with the task to observe their behaviour seperate from their lack of experience. For example, as seen in our own results, the stroke patients performance started rather low but improved vastly over time whereas our experienced participants had decreased performance.

Overall, our findings suggest that studies aiming to assess neglect through VR diagnosis tasks should aim to employ tracking measures as well as performance measures to understand their participants' mental states. Moreover, we suggest extending task lengths and looking at changes over time and as opposed to only overall data. This paper also suggests that fatigue may not lower all performance measures due to learning effects during interactions. Finally, we also suggest that designers of VR diagnosis tasks should employ virtual environments with narrative elements even at the cost of fatiguing patients. As evidence from this study and others [28, 40] indicate, creating visually pleasing environments for patients can motivate them to engage with the system more strongly.

CONCLUSION

Using positional tracking measures as proxies for fatigue provide higher accuracy at predicting physiologically measured fatigue compared to task performance. Pupil diameter proved especially influential.

The duration of the task impacted fatigue but taking the learning effect into account provided a nuanced overview using performance. We observed participants' performance improve where the moles per second increased, but also degrade in other factors, e.g. the amount of misses and lifetime of targets also increased.

Employing environmental noise increased fatigue. However, the narrative component of the environment can further enlarge the impact. Future studies should investigate the impact of narrative based tasks with environmental noise on fatigue while taking emotions into account when designing tasks and using physiology.

ACKNOWLEDGEMENTS

The group is very grateful for the guidance by Hendrik Knoche and Bastian Ilsø Hougaard. Moreover, we would like to thank all the participants that volunteered their time for our study. Lastly, we are grateful for the cooperation with Quentin Roman, Vincent Simonetti-Diez, and Thomas Dupin who aided our development of the WAM game, logging programs, and feature extraction respectively.

REFERENCES

[1] 2021. hrv-analysis. (2021). https://github.com/Aura-healthcare/hrv-analysis

[2] 2021. Whack_A_Mole_VR. (2021). tps://github.com/med-material/Whack_A_Mole_VR

[3] Gayatri Aravind and Anouk Lamontagne. 2014. Perceptual and locomotor factors affect obstacle avoidance in persons with visuospatial neglect. *Journal of NeuroEngineering and Rehabilitation* 11, 1 (2014), 38. DOI:http://dx.doi.org/10.1186/1743-0003-11-38

[4] Kenji Baheux, Makoto Yoshizawa, Akira Tanaka, Kazunori Seki, and Yasunobu Handa. 2005. Diagnosis and rehabilitation of hemispatial neglect patients with virtual reality technology. *Technology and Health Care* 13, 4 (July 2005), 245–260. DOI: http://dx.doi.org/10.3233/THC-2005-13404

[5] Elvio Blini, Zaira Romeo, Chiara Spironelli, Marco Pitteri, Francesca Meneghello, Mario Bonato, and Marco Zorzi. 2016. Multi-tasking uncovers right spatial neglect and extinction in chronic left-hemisphere stroke patients. 92 (2016), 147–157. DOI: http://dx.doi.org/10.1016/j.neuropsychologia.2016.02.028

[6] L.J. Buxbaum, M.K. Ferraro, T. Veramonti, A. Farne, J. Whyte, E. Ladavas, F. Frassinetti, and H.B. Coslett. 2004. Hemispatial neglect: Subtypes, neuroanatomy, and disability. *Neurology* 62, 5 (March 2004), 749–756. DOI: http://dx.doi.org/10.1212/01.wnl.0000113730.73031.f4

[7] Laurel J. Buxbaum, Amanda M. Dawson, and Drew Linsley. 2012. Reliability and validity of the Virtual Reality Lateralized Attention Test in assessing hemispatial neglect in right-hemisphere stroke. *Neuropsychology* 26, 4 (2012), 430–441. DOI: http://dx.doi.org/10.1037/a0028674 Place: US Publisher: American Psychological Association.

[8] Laurel J. Buxbaum, Mary Ann Palermo, Dina Mastrogiovanni, Mary Schmidt Read, Ellen Rosenberg-Pitonyak, Albert A. Rizzo, and H. Branch Coslett. 2008. Assessment of spatial attention and neglect with a virtual wheelchair navigation task. *Journal of Clinical and Experimental Neuropsychology* 30, 6 (July 2008), 650–660. DOI: http://dx.doi.org/10.1080/13803390701625821

[9] Rebecca L. Charles and Jim Nixon. 2019. Measuring mental workload using physiological measures: A systematic review. *Applied Ergonomics* 74 (Jan. 2019), 221–232. DOI: http://dx.doi.org/10.1016/j.apergo.2018.08.028

[10] Nele Demeyere and Celine R. Gillebert. 2019. Ego- and allocentric visuospatial neglect: Dissociations, prevalence, and laterality in acute stroke. *Neuropsychology* 33, 4 (May 2019), 490–498. DOI: http://dx.doi.org/10.1037/neu0000527

[11] Stephen H. Fairclough. 2009. Fundamentals of physiological computing. *Interacting with Computers* 21, 1-2 (Jan. 2009), 133–145. DOI: http://dx.doi.org/10.1016/j.intcom.2008.10.011

[12] Martin Gjoreski, Bhargavi Mahesh, Tine Kolenik, Jens Uwe-Garbas, Dominik Seuss, Hristijan Gjoreski, Mitja Lustrek, Matjaz Gams, and Veljko Pejovic. 2021. Cognitive Load Monitoring With Wearables–Lessons Learned From a Machine Learning Challenge. *IEEE Access* 9 (2021), 103325–103336. DOI: http://dx.doi.org/10.1109/access.2021.3093216

[13] Eija Haapalainen, SeungJun Kim, Jodi F. Forlizzi, and Anind K. Dey. 2010. Psycho-physiological measures for assessing cognitive load. In *Proceedings of the 12th ACM international conference on Ubiquitous computing*. ACM. DOI:http://dx.doi.org/10.1145/1864349.1864395

[14] Kristin J. Heaton, James R. Williamson, Adam C. Lammert, Katherine R. Finkelstein, Caitlin C. Haven, Douglas Sturim, Christopher J. Smalt, and Thomas F. Quatieri. 2020. Predicting changes in performance due to cognitive fatigue: A multimodal approach based on speech motor coordination and electrodermal activity. *The Clinical Neuropsychologist* 34, 6 (July 2020), 1190–1214. DOI: http://dx.doi.org/10.1080/13854046.2020.1787522

[15] Bastian I. Hougaard, Hendrik Knoche, Jim Jensen, and Lars Evald. 2021. Spatial neglect midline diagnostics from virtual reality and eye tracking in a free-viewing environment. *unpublished* (2021).

[16] Georg Kerkhoff and Thomas Schenk. 2012. Rehabilitation of neglect: An update. *Neuropsychologia* 50, 6 (May 2012), 1072–1079. DOI: http://dx.doi.org/10.1016/j.neuropsychologia.2012.01.024

[17] Tae-Lim Kim, Kwanguk Kim, Changyeol Choi, Ji-Yeong Lee, and Joon-Ho Shin. 2021. FOPR test: a virtual reality-based technique to assess field of perception and field of regard in hemispatial neglect. *Journal of NeuroEngineering and Rehabilitation* 18, 1 (Feb. 2021). DOI: http://dx.doi.org/10.1186/s12984-021-00835-1

[18] Samuel E. J. Knobel, Brigitte C. Kaufmann, Stephan M. Gerber, Dario Cazzoli, René M. Müri, Thomas Nyffeler, and Tobias Nef. 2020. Immersive 3D Virtual Reality Cancellation Task for Visual Neglect Assessment: A Pilot Study. 14 (2020), 180. DOI: http://dx.doi.org/10.3389/fnhum.2020.00180

[19] Hendrik Knoche, Kasper Hald, Dorte Richter, and Helle Rovsing Møller Jørgensen. 2017. Self-rehabilitation of acquired brain injury patients including neglect and attention deficit disorder with a tablet game in a clinical setting. (2017). DOI: http://dx.doi.org/10.4108/eai.18-7-2017.152895 Publisher: European Alliance for Innovation (EAI) Section: Journal_Article.

[20] Johan Winther Kristensen, Lasse Lodberg Aafeldt, Peter Kejser Jensen, Rebecca Pipaluk Vinther, and Hendrik Knoche. 2020. Renoir in VR: Comparing the Relaxation from Artworks Inside and Outside of Virtual Reality. In *Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering*. Springer International Publishing, 217–228. DOI: http://dx.doi.org/10.1007/978-3-030-53294-9_15

[21] Johannes Kögel, Ralf J. Jox, and Orsolya Friedrich. 2020. What is it like to use a BCI?–insights from an interview study with brain-computerinterface users. (2020), 14. DOI: http://dx.doi.org/10.1186/s12910-019-0442-2

[22] Yiqing Liang, Richard M. Guest, Michael C. Fairhurst, and Jonathan M. Potter. 2010. A computer-based quantitative assessment of visuo-spatial neglect using regression and data transformation. 13, 4 (2010), 409–422. DOI: http://dx.doi.org/10.1007/s10044-009-0172-z

[23] Horia A. Maior, Max L. Wilson, and Sarah Sharples. 2018. Workload Alerts—Using Physiological Measures of Mental Workload to Provide Feedback During Tasks. *ACM Transactions on Computer-Human Interaction* 25, 2 (April 2018), 1–30. DOI: http://dx.doi.org/10.1145/3173380

[24] Michael Meehan, Brent Insko, Mary Whitton, and Frederick P. Brooks. 2002. Physiological measures of presence in stressful virtual environments. In *Proceedings of the 29th annual conference on Computer graphics and interactive techniques*. ACM Press. DOI: http://dx.doi.org/10.1145/566570.566630

[25] Mark Mennemeier. 2011. Neglect Syndrome. In *Encyclopedia of Clinical Neuropsychology*. Springer New York, 1722–1728. DOI: http://dx.doi.org/10.1007/978-0-387-79948-3_1384

[26] Tanja Nijboer, Ingrid van de Port, Vera Schepers, Marcel Post, and Anne Visser-Meily. 2013. Predicting Functional Outcome after Stroke: The Influence of Neglect on Basic Activities in Daily Living. *Frontiers in Human Neuroscience* 7 (2013). DOI: http://dx.doi.org/10.3389/fnhum.2013.00182

[27] T. C. W. Nijboer, A. F. ten Brink, M. Kouwenhoven, and J. M. A. Visser-Meily. 2014. Functional Assessment of Region-Specific Neglect: Are There Differential Behavioural Consequences of Peripersonal versus Extrapersonal Neglect? *Behavioural Neurology* 2014 (2014), 1–6. DOI: http://dx.doi.org/10.1155/2014/526407

[28] Tatiana Ogourtsova, Philippe Archambault, Samir Sangani, and Anouk Lamontagne. 2018. Ecological Virtual Reality Evaluation of Neglect Symptoms: Effects of Virtual Scene Complexity in the Assessment of Poststroke Unilateral Spatial Neglect. 32, 1 (2018), 46–61. DOI: http://dx.doi.org/10.1177/1545968317751677 Publisher: SAGE Publications Inc STM.

[29] Shankar Rao Pandala. 2021. Welcome to Lazy Predict's documentation! (2021). https://lazypredict.readthedocs.io/en/latest/

[30] Elisa Pedroli, Silvia Serino, Pietro Cipresso, Federica Pallavicini, and Giuseppe Riva. 2015. Assessment and rehabilitation of neglect using virtual reality: a systematic review. 9 (2015), 226. DOI: http://dx.doi.org/10.3389/fnbeh.2015.00226

[31] Sarah Pink. *Doing Visual Ethnography*. SAGE Publications, Ltd. DOI: http://dx.doi.org/10.4135/9780857025029

[32] Kasper Rodil, Donovan Maasz, and Heike Winschiers-Theophilus. 2020. Moving Virtual Reality out of its Comfort Zone and Into the African Kalahari Desert Field: Experiences From Technological Co-Exploration With an Indigenous San Community in Namibia. In *26th ACM Symposium on Virtual Reality Software and Technology*. ACM. DOI: http://dx.doi.org/10.1145/3385956.3418955

[33] Fred Shaffer and J. P. Ginsberg. 2017. An Overview of Heart Rate Variability Metrics and Norms. *Frontiers in Public Health* 5 (Sept. 2017). DOI: http://dx.doi.org/10.3389/fpubh.2017.00258

[34] Ludwig Sidenmark and Hans Gellersen. 2020. Eye, Head and Torso Coordination During Gaze Shifts in

Virtual Reality. *ACM Transactions on Computer-Human Interaction* 27, 1 (Jan. 2020), 1–40. DOI: http://dx.doi.org/10.1145/3361218

[35] Filip Skola and Fotis Liarokapis. 2018. Embodied VR environment facilitates motor imagery brain–computer interface training. *Computers & Graphics* 75 (Oct. 2018), 59–71. DOI: http://dx.doi.org/10.1016/j.cag.2018.05.024

[36] Mitchell R. Smith, Rifai Chai, Hung T. Nguyen, Samuele M. Marcora, and Aaron J. Coutts. 2019. Comparing the Effects of Three Cognitive Tasks on Indicators of Mental Fatigue. *The Journal of Psychology* 153, 8 (June 2019), 759–783. DOI: http://dx.doi.org/10.1080/00223980.2019.1611530

[37] Tmalsburg. 2021. Tmalsburg/saccades: Detection of fixations and saccades in eyetracking data. (2021). https://github.com/tmalsburg/saccades

[38] Inna Tsirlin, Eve Dupierrix, Sylvie Chokron, Sabine Coquillart, and Theophile Ohlmann. 2009. Uses of virtual reality for diagnosis, rehabilitation and study of unilateral spatial neglect: review and analysis. 12, 2 (2009), 175–181. DOI: http://dx.doi.org/10.1089/cpb.2008.0208

[39] Roza M. Umarova. 2017. Adapting the concepts of brain and cognitive reserve to post-stroke cognitive deficits: Implications for understanding neglect. *Cortex* 97 (Dec. 2017), 327–338. DOI: http://dx.doi.org/10.1016/j.cortex.2016.12.006

[40] Sebastian Wagner, Julia Belger, Fabian Joeres, Angelika Thöne-Otto, Christian Hansen, Bernhard Preim, and Patrick Saalfeld. 2021. iVRoad: Immersive virtual road crossing as an assessment tool for unilateral spatial neglect. *Computers & Graphics* 99 (2021), 70–82. DOI: http://dx.doi.org/10.1016/j.cag.2021.06.013