CHALMERS
UNIVERSITY OF TECHNOLOGY | UNIVERSITY OF GOTHENBURG

An evaluation of the feasibility of utilizing Virtual Reality as a tool to assess cognitive function

Master's thesis in Interaction Design and Technologies

EMMA VISKARI

Department of Computer Science and Engineering
CHALMERS UNIVERSITY OF TECHNOLOGY
UNIVERSITY OF GOTHENBURG
Gothenburg, Sweden 2020

Master's thesis 2020

An evaluation of the feasibility of utilizing Virtual Reality as a tool to assess cognitive function

EMMA VISKARI

UNIVERSITY OF GOTHENBURG

CHALMERS
UNIVERSITY OF TECHNOLOGY

Department of Computer Science and Engineering
Division of Interaction Design
Chalmers University of Technology
University of Gothenburg
Gothenburg, Sweden 2020

An evaluation of the feasibility of utilizing Virtual Reality as a tool to assess cognitive function.
EMMA VISKARI

Supervisor: Gordana Dodig-Crnkovic, Department of Computer Science and Engineering
Advisor: Magnus Kjellberg, Centrum för Digital Hälsa, Jacob Stålhammar, Minnesmottagningen, Sahlgrenska University Hospital
Examiner: Staffan Björk, Department of Computer Science and Engineering

Master's Thesis 2020
Department of Computer Science and Engineering
Division of Interaction Design
Chalmers University of Technology and University of Gothenburg
SE-412 96 Gothenburg
Telephone +46 31 772 1000

Cover: Screenshot from the grocery store task prototype.

Gothenburg, Sweden 2020

An evaluation of the feasibility of utilizing Virtual Reality as a tool to assess cognitive function
EMMA VISKARI
Department of Computer Science and Engineering
Chalmers University of Technology and University of Gothenburg

Abstract

Virtual reality (VR) is currently being studied to determine its' possible uses within healthcare, for example to assess cognitive impairment. This project investigates the feasibility of utilizing VR to assess cognitive function in a clinical setting. The report begins with a literature study investigating various aspects of VR and cognitive impairment, with a certain focus on dementia where VR could be particularly useful, as well as examples of VR being applied in this field. Furthermore, aspects pertaining to usability and user experience of these types of VR systems are discussed, and how VR assessment tasks could compare to more traditional methods in terms of efficiency, adaptability, and user comfort. The results of the literature study are summarized into identified problems and advantages, attempting to provide a holistic perspective of aspects that need to be considered when attempting to draw conclusions regarding the feasibility of utilizing VR to assess cognitive function.

Furthermore, the literature study, along with stakeholder interviews, served to inform the design of a low fidelity VR prototype. The prototype was designed in order to explore the various possibilities of designing a VR task for cognitive assessment. The prototype, which was based around a grocery store task combining memorization and navigation, was evaluated in order to identify possible issues that need to be considered when designing this type of assessment task.

Keywords: VR, virtual reality, interaction design, cognitive impairment, user acceptance, usability.

Acknowledgements

This study was carried out as a master's thesis at Chalmers University of Technology, at the program Interaction Design and Technologies, in collaboration with Centrum för Digital Hälsa, and Minnesmottagningen, at Sahlgrenska University Hospital.
I would like to thank Jacob Stålhammar, neuropsychologist at Minnesmottagningen, for his guidance and encouragement, and for providing valuable insight into the many aspects of cognitive assessment. I would also like the express my gratitude to Magnus Kjellberg, for his engagement in the project and support, and for the opportunity to collaborate with Centrum för Digital Hälsa. Furthermore, I would like to thank my Chalmers supervisor Gordana Dodig-Crnkovic for her supervision and valuable feedback. A special thank you also, to the participants who helped evaluate the prototype.

Emma Viskari, Gothenburg, May 2020

Abbreviations

AD = Alzheimer's disease
ADL = Activities of daily living
aMCI = Amnestic mild cognitive impairment
CSF = Cerebrospinal fluid
CT = Computed Tomography
GDPR = General Data Protection Regulation HCD = Human-centered design
HMD = Head mounted display
IADL = Instrumental activities of daily living
IPD = Interpupillary distance
MCI = Mild cognitive impairment
MMSE = Mini mental state examination
MRI = Magnetic resonance imaging
PET = Positron Emission Tomography
UEQ = User Experience Questionnaire
VR = Virtual reality

Contents

List of Figures

1 Introduction

The aim of this project was to investigate the feasibility of utilizing Virtual Reality (VR) to assess cognitive function or detect cognitive impairment in a clinical setting. In order to investigate the feasibility of utilizing VR within this field, literature studies and interviews with stakeholders were performed. Furthermore, a low fidelity prototype was created to explore how such a system could be designed, and to identify possible issues that may arise. Cognitive impairment can result from a number of conditions, one example is different types of dementia. Alzheimer's disease (AD) is an incurable neurodegenerative disease, and the most common cause of dementia. An early diagnosis of AD is imperative to be able to impede the progression of the disease, and ease symptoms. AD often causes a gradual degeneration, and is often preceded by lesser degrees of cognitive impairment which may progress to dementia. Mild cognitive impairment (MCI) can be a precursor of AD, but can also have other causes and does not necessarily have to progress to dementia. MCI may be difficult to detect, and often require extensive neuropsychological investigations, which may be time-consuming. VR could potentially be used as a tool to make the process of detecting signs of cognitive impairment more efficient, by for example being able to perform more screenings of patients, as well as making more aspects of cognition measurable. However, this type of system could possibly have a number of usability related issues, which needs to be investigated.

The primary stakeholders in this project are Centrum för Digital Hälsa, which is a unit at Sahlgrenska University Hospital that focuses on various aspects of digitization within healthcare, and Minnesmottagningen, a unit at Sahlgrenska University Hospital specialized on dementia and cognitive screening. Minnesmottagningen provided expertise on the topic of cognitive assessment, and are also potential users of such a system, if it was to be implemented in practice.

1.1 Aim

The goal of this project was to evaluate the feasibility of utilizing VR as a part of clinical assessment of cognition, for example in dementia assessment, using literature studies, interviews with stakeholders, and evaluate existing technologies within this field. The feasibility includes not only the technological aspects, but other aspects as well. Utilizing VR in healthcare could introduce numerous other concerns, such as concerns regarding hygiene, how the technology will be received by medical personnel, and how it will be received by the patients themselves. The second goal was to produce a design concept or early prototype to illustrate how this type of system could be designed. Since this project primarily focused on the theoretical aspects, the design was intended to illustrate the possible design considerations that need to be kept in mind when creating VR systems that aim to utilize VR as a part of clinical assessment of cognitive function. One critical issue is to attempt to determine which types of assessments are, not only feasible, but also actually valuable to perform using VR. Pen and paper tasks, and digitized versions of pen and paper tasks, can be very useful, since they are easy to perform and adapt according to what the clinician is looking for, and often uses tools patients may be familiar with. It would be useful to investigate whether VR could have a similar flexibility and ease of use. Furthermore, many of the studies of using VR for diagnosing or differentiating dementia are focused on aspects of navigation, which could indicate that this is one aspect where VR holds some particular value. However it would also be very relevant to investigate whether VR has some potential to assess other aspects of cognition.

1.2 Problem description

Certain signs of cognitive impairment may be difficult to detect during analog screenings, such as eye movements, subtle changes of motor abilities, speech impediments. Such screenings may be difficult to perform efficiently, and require a lot of resources. These screenings are often time consuming, and require a series of neuropsychological assessments, and physiological examinations. Furthermore, a current issue with dementia screening is that results can be affected by the patient's education level and native language, and thus make diagnosing the disease difficult. Neuropsychological assessment may become more accurate when using methods already familiar to the patient, for example pen and paper based tasks, alphabet based tasks or tasked based on verbal learning and previous knowledge. This may make neuropsychological assessment more reliable, they also depend on being familiar with language, pen and paper, and alphabets. As such, neuropsychological assessment is sensitive to whether a patient is tested in their native language or not, and the level of education. Computerized testing could allow assessment in more languages than spoken by a human assessor, and VR could potentially test performance on non-verbal cognitive tasks, for example navigational tasks, sorting tasks, or three dimensional puzzles. Utilizing VR in a clinical setting could potentially provide the possibility to assess multiple cognitive aspects, more efficiently, and aid in early detection of cognitive impairment, for example in AD.

1.3 Research question

In order to investigate the multiple aspects of utilizing VR within this field, the research question for this project was:
How feasible, both technologically and from a user perspective, is it to assess cognitive function utilizing Virtual Reality, and how could such a system be designed?.
In order to further investigate the potential advantages, or disadvantages, of utiliz-

ing VR in comparison to other methods of cognitive assessment, there was also a subquestion:
How does Virtual Reality compare to currently used methods of cognitive assessment?

In order to answer these questions, the results from the literature study were summarized to provide an overview of identified problems and advantages of utilizing VR in this context, in order to be able to draw conclusions regarding the feasibility, and how they relate to currently used methods. Furthermore, literature studies and interviews aimed to inform the design of a low fidelity prototype, to explore how such a system could be designed and elicit issues that need consideration.

2
Background

There were several challenges within this project. One challenge is the human aspects of use of this type of system, since VR is a relatively new technology, which many people may be unfamiliar with. This could affect how the technology is perceived and interacted with. Furthermore, there are important considerations to take into account when designing for VR, such as ergonomic aspects and the risk of eye-strain, motion sickness and disorientation (Jerald, 2016). These aspects could be especially important to keep in mind when designing for users with suspected cognitive impairment. It is possible that cognitive impairment could exacerbate the risk of adverse effect of using VR, which needs to be investigated.

2.1 Context

VR can be described as a computer-generated digital environment that can be interacted with, and experienced, as though it was a real environment (Jerald, 2016). VR is being applied in numerous aspects of the medical field today, and several studies have investigated VR used specifically for assessing various aspects of cognitive function. For example, Howett et al. (2019) conducted a study to differentiate MCI. The authors conducted an immersive path integration test, and uncovered that the VR navigation test could help distinguish patients with higher or lower risk levels of early stage Alzheimer' disease and could aid in early diagnosis. While the study shows promising results in discovering patients at risk of dementia, it lacks in discussion about actually applying the technology outside the controlled environment of the study, for example how this system would be received by the medical personnel or patients who would be using it, if it were to be used on a larger scale.

Cushman et al. (2008) compared navigation tests in a virtual environment, with the same navigation test in a physical environment, where the VR environment simulated the physical environment. The authors point out that real life navigational testing can be time-consuming, and that VR could be a good alternative to make the testing more efficient, and their results point to that virtual navigational tests could be used as a viable alternative. However, the authors also found that the tests scores for the VR navigation tests were slightly lower than the scores for real world navigation, for all test subjects. Furthermore, much like Howett et al. (2019), the article lacks a holistic perspective of actually using the system - while the results of using VR to assess cognitive function appear promising, there may be several usability-related issues both for patients and medical personnel.

Issues with currently used methods of assessment and diagnostic methods, such as neuropsychological assessment, neuroimaging and analysis of cerebrospinal fluid (CSF) biomarkers are that they may be time consuming, which is the case for neuropsychological assessments. Furthermore, neuroimaging may be expensive and time consuming, and the neuroimaging machines are not always available outside of specialist centers. Analysis of CSF is an invasive procedure which may be uncomfortable and cause side effects (Laske et al. 2015)

The primary research problem of this project was to uncover if it is feasible to utilize VR technology to assist in assessing cognitive function, in a clinical setting. This could be further divided into how this type of system could affect the users - both medical personnel and patients, and how a VR system for cognitive assessment could be designed.

3
Theory

3.1 Dementia and Alzheimer's disease

There are several different types of dementia, brain diseases which gradually affects cognitive function and daily life activities, to a greater degree than what is expected of normal aging (Funke & Willbold, 2011). The most common type of dementia is Alzheimer's disease, which makes up 60 to 80 percent of dementia cases, affecting 27 million people worldwide, the primary risk factor being aging (however AD is not considered a part of "normal aging") (Funke & Willbold, 2011). Dementia causes numerous symptoms, both behavioural, psychiatric, cognitive and physical. The physical symptoms become more common as the disease progresses. In the early stages of the disease, symptoms include difficulty remembering names and recent events, as well as apathy and depression. The later stages of the disease includes further cognitive symptoms such as disorientation, confusion, changes in behaviour, as well as difficulties swallowing, walking and speaking (Funke & Willbold, 2011). As the disease becomes more severe, the individuals' ability to function independently in daily life is progressively lost, with difficulties in washing oneself, going to the bathroom, or eating independently. This in turn can also lead to issues with physical health and make patients more susceptible to other illnesses (NICE-SCIE Guideline, 2007).

Numerous diseases can cause dementia, Alzheimer's disease being the most common, followed by vascular dementia, dementia with Lewy-bodies which each make up about 15-20% of cases, and frontotemporal dementia, which makes up many cases of early onset dementia, second to Alzheimer's disease. Other diseases that can cause dementia include, among others, Parkinson's disease, Huntington's disease and other neurodegenerative diseases, as well as alcohol-related dementia. Mixed dementia also occur in older patients, with multiple types of dementia present at the same time, such as Alzheimer's disease and vascular dementia, for example (NICE-SCIE Guideline, 2007). Mild cognitive impairment (MCI) can be a preclinical stage of Alzheimer's disease, which often presents itself as more subtle cognitive impairment, affecting for example reasoning skills, memory, decision making, concentration or language, often before the onset of dementia. MCI presents as a decline from previous cognitive function, but not to the degree that it affects daily or social life. It is believed that several types of MCI exist, non-amnestic MCI, and amnestic MCI (aMCI) which affects memory and is believed to be most likely to develop into dementia (NICE-SCIE Guideline, 2007). It may be worth noting that MCI does not necessarily convert to dementia, and that cognitive symptoms can be caused by other issues, for example, vitamin B12 deficiency, psychiatric disorders, or lesions. Furthermore, it is possible that patients either return to normal cognition, or that the MCI remains stable without converting to dementia, however it can also progress further to dementia (Petersen et al., 2018)

Detection of the disease often occurs when the individual him- or herself recognizes changes in emotions, abilities or cognitive ability, or if somebody close to them, like a family member or care taker notices the changes. It can also be detected by medical personnel during check-ups, for example (NICE-SCIE Guideline, 2007).

3.2 Currently used assessment methods

There is a vast amount of assessment methods and neuropsychological screening tests used to diagnose Alzheimer's disease. This section will briefly discuss a few of the many tools and methods used.

Varying degrees of memory impairment is common as one ages, which can make it difficult to determine whether the impairment or forgetfulness is due to other, perhaps temporary, reasons, or due to Alzheimer's disease. It is necessary to rule out other possible causes of the impairment, such as psychiatric disorders (particularly depression), or, for example, vitamin B12 deficiency. In order to diagnose dementia, the patient often self-reports subjective experiences, often supported by a secondary person (for example, a relative), and other potential or diagnoses need to be excluded, and the individuals' cognitive abilities need to be assessed. Further differentiation of diagnosis often requires further cognitive assessment and possibly neuroimaging such as PET scans (NICE-SCIE Guideline, 2007).

One common screening test is the Mini Mental State Examination (MMSE). The MMSE assesses cognitive, language and memory abilities and provides a brief overview of cognitive function, and can help classifying the severity of impairment, but for diagnosis. The MMSE is a test with a maximum score of 30 points, and can take 5 to 10 minutes to administer. The test assesses several cognitive domains: time and place orientation, recall of words, language, attention and calculation, registration of words, and visual construction, with some variety. A score between 24-30 indicates normal cognition, while a score between 18-23 indicates MCI, and a score below 17 indicates severe cognitive impairment (Tombaugh & McIntyre, 1992).

Tombaugh and McIntyre (1992) also identified a number of issues to keep in mind regarding the MMSE: The test is sensitive to age, level of education, and cultural background, and the examination is highly verbal. Education appears to cause the highest variance compared to gender, cultural background or social class. Education level could, according to Tombaugh and McIntyre (1992) lead to a miss-classification, by introducing bias, for example increasing the risk of false positives for patients with lower education - classifying normal individuals as having a cognitive impairment. Similarly, individuals with a higher educational background could potentially mask

MCI. Furthermore, Tombaugh and McIntyre (1992) recommend not to administer the examination clinically unless the individual is fluent in English. In other words, native language should be kept in mind, and the test exist in multiple languages. According to Tombaugh and McIntyre (1992), the MMSE shows high specificity to moderate to higher degrees of cognitive impairment, however less specificity for lower degrees of impairment.

The Clock Drawing Test is another common screening test. It is relatively straightforward to administer - the patient is asked to draw a clock, enumerate the face, and place the hands of the clock at a certain time. It exists in multiple variants, for example, free-hand drawing, with a pre-drawn circle, or with a pre-drawn clock where the patient is only asked to draw the hands of the clock at a certain time. The test assesses visuo-spatial abilities, and requires verbal understanding and memory, and is normally used in conjunction with other assessments. It is able to fairly accurately classify patients with Alzheimer's disease, however, may be less reliable when it comes to MCI. Furthermore, the results may be affected by level of education, emotional state and age. The positive aspects of the test is that it is quick and easy to administer, and is easily acceptable by the patients (Agrell & Dehlin, 1998).

There are numerous neuropsychological screening tests to diagnose Alzheimer's disease and assess cognitive function. Screening test are useful to highlight abnormal cognitive function, but in order to fulfill criteria for diagnosis, a complimentary round of neuropsychological batteries consisting of several different tests are useful to outline impairment in different cognitive domains. Language, visuo-spatial abilities, executive function, attention, praxis, and naturally, memory are aspects that should be assessed (Picchi et al., 2011). However, clinical symptoms may arise late in the progression of the disease, and prodromal cases may be difficult to detect even in specialized centers. Detecting the disease before clinical symptoms arise is not impossible. Studies have shown that accumulation of neurofibrillary tangles and amyloid plaques in the brain tissue, may start as early as 10 to 20 years before symptoms arise, which can lead to extensive neuronal loss. (Funke & Willbold, 2011).

According to Jack et al. (2013) the first visible sign of possible Alzheimer's disease is the decrease of Amyloid Beta 42 (AB42), in cerebrospinal fluid (CSF). Another biomarker than can be used is the increase of Tau or phosphorated Tau. These biomarkers can aid in providing a more accurate, and earlier, diagnosis, since they often are present before the onset of cognitive symptoms (Jack et al. 2013). In other words, measuring cognition is only possible after the onset of cognitive symptoms, while biomarkers could detect the disease before cognitive symptoms arise. CSF is sampled through a procedure called lumbar puncture, which is done by using a needle to extract fluid from the spine, which can cause certain complications, most commonly headaches (Evans, 1998)

Aside from the neuropsychological screening tests and CSF biomarker analysis, it may also be useful to utilize neuroimaging, to be able to exclude other causes, that may be reversible. A computed tomography (CT) is a relatively fast and inexpensive option to detect possible reversible causes, however it cannot differentiate normal ageing from dementia. Magnetic Resonance Imaging (MRI) is more sensitive and able to provide a more exact image, however it also requires patients to lie still for a longer period of time, in an enclosed space, and is more expensive. Positron Emission Tomography (PET) can be helpful in detecting amyloid plaques and detect pre-clinical dementia (Picchi et al., 2011)

3.3 Digital technologies for cognitive assessment

This section briefly will discuss some digital technologies, aside from VR, that have been applied to cognitive assessment in Alzheimer's disease and mild cognitive impairment.

Computerized neurological assessment for touch screens and personal computer have been around since the 1980s and 1990s, for example Cambridge Neuropsychological Test Automated Battery and Examen Cognitif par Ordinateur (Astell et al. 2019), and numerous classic pen and paper assessments have been digitized to web or mobile applications. One example is the Clock Drawing Test (ibid).

Laske et al (2015) reviewed a number of different diagnostic tools for detecting AD in its' early stages. Their argument for the potential usefulness of new tools is that currently used methods of diagnosis can be either time-consuming, invasive, or expensive. One example of a potential method is Automatic Speech Analysis and Recognition (ASR), which can be used to analyze the aspects of verbal communication that can be affected by AD, such as fluency of speech and emotional responses. AD can affect verbal communication, such as causing difficulties with speaking and understanding, or recognizing and naming things. ASR can provide relatively quick and easy tools to analyze a patients' speech to detect difficulties that can be related to MCI. Analysis of spontaneous speech can also be useful to detect Alzheimer's disease, using Automatic Spontaneous Speech Analysis and Emotional Response Analysis, to detect additional aspects of speech. Verbal Fluency by category is a commonly used test where patients are given a category and asked to name as many words as possible from this category. This test has often been performed manually, but technology has enabled the Verbal Fluency test to be automated (Laske et al, 2015). Another potential tool, according to Laske et al. (2015), is gait analysis, which is used to detect gait impairment often associated with cognitive decline. Gait analysis can also be combined with performing a second task while walking. Uncovering gait impairment may be difficult to do with just the naked eye, but can be measured using quantitative gait analysis.

The Useful Field of View test is another example of a cognitive assessment tool that has been successfully digitized for a clinical setting. It can be performed using a regular desktop computer and mouse, or touch screen (Edwards et al. 2005). The UFOV test has been applied in multiple settings to assess processing speed, visual attention and search (ibid), and driving ability and predicting risk of car crashes, but can also be used to detect cognitive decline in AD (Edwards et al., 2006)

3.4 VR in healthcare and cognitive assessment

There have been several studies utilizing VR in healthcare. VR has been studied for training of medical personnel, for example surgical simulators used as 'serious games', where game-based learning is used to engage the users (de Ribaupierre et al., 2014), for social cognition training in youths with autism, such as training of conversational skills (Kandalaft et al., 2013), and to help healthcare personnel empathize more with elderly patients with age-related diseases (Dyer et al., 2018). VR has also been applied, in multiple different studies, to different areas of dementia related healthcare- for assessment and diagnosis, for training cognitive abilities and therapy, as well as for training of caregivers (García-Betances et al. 2015).

Howett et al. (2019) conducted a study to differentiate mild cognitive impairment, the pre-dementia stage of Alzheimer's disease using a path integration task in VR. Path integration is defined by the authors as the ability to keep track of a starting location, and being able to return to the same location, which is dependent on continuously integrating visual, proprioceptive (one's' own body position and its' movements) and vestibular cues to represent one's' current position and direction related to the location. Results of the study indicate that the VR navigation task could aid in diagnosing Alzheimer's disease at an early stage of the disease, and detect MCI patients with high risk of developing dementia.

During the trials, the participants navigated the real environment while wearing a VR headset showing virtual environments with different boundary and textural cues, navigating to three different cones, one at the time, where each cone disappears after being reached, and the next cone would present itself. Having located the third and final cone, the participants were prompted to return to the remembered location of the first cone utilizing textual and auditory instructions. When the participants felt that they had reached the location, they pressed a button on the hand controller, and the distance between estimated and actual distance of the first cone was measured - the absolute distance error. Howett et al. (2019) found that the path integration task showed a higher sensitivity and specificity for diagnosis, than the cognitive assessment tests, in differentiating the patients who showed positive for CSF biomarkers, and those who showed negative. The authors found that the results of the entorhinal cortex-based VR path navigation task had superior accuracy in classifying prodromal Alzheimer's disease, compared to commonly used cognitive tests that assess episodic memory, attention and processing speed.

In another study, Plancher et al. (2011) utilized VR to characterize episodic memory profiles in three different groups of participants - one group with aMCI (amnestic mild cognitive impairment), one with early to moderate Alzheimer's disease, as well as a group of healthy elderly people. The idea was to provide a more ecologically valid environment compared to classical neuropsychological assessments. The authors explored both active and passive explorations of the virtual environment - the participants were placed either as driving a car, or as a passenger of a car. After the sessions, the patients performed recall and recognition tests to evaluate episodic memories of the environment. Episodic memory concerns recalling everyday events and their contexts. In the study, the participants traveled through a virtual town, and were subsequently asked to recall for example what they saw, who they saw, when they saw something (at the beginning, middle or end of the town), as well as allocentric where (how the elements of the environment were situated relative to each other) and egocentric where (how they traveled the town). The recognition task consisted of yes and no questions regarding the presence of elements in the virtual environment, and the elements' spatial and temporal relationships to each other. Plancher et al. (2011) found that their VR tests appeared successful in differentiating between normal aging and pathological aging, the aMCI and AD participants were found to be impaired in spatial memory scores, both regarding immediate and delayed recall of both egocentric memory and allocentric spatial information, compared to the control group. Furthermore, the authors found that spatial allocentric memory assessment shows promise as a diagnostic cue for aMCI, since spatial allocentric memory was found to be particularly successful in differentiating healthy and aMCI participants.

Cushman et al. (2008) compared a real environment navigation test with the same test in VR, to discover how VR assessment of navigational skills corresponds to the real world. The study had two control groups, young normal participants and older normal participants, as well as two groups consisting of participants with MCI, and early Alzheimer's disease, respectively. Cushman et al. (2008) mean that since navigational impairment is unrelated to verbal memory impairment, and that issues with navigation, for example getting lost while walking or driving, often is an early indicator of AD, but that these navigational deficits rarely are measured, which may impede the detection of early stage AD, and affect the possibility of risk assessment for the patients' continued independent living, as well as driving abilities. The authors had previously conducted studies of real world navigation in participants with AD, where they identified visuospatial and verbal abilities related to navigation of the environment. The authors reason that VR could potentially be less time consuming and less difficult to perform. The real environment task was performed in a hospital lobby, where participants were taken on a tour in a wheelchair (for the sake of the elderly participants, as not to influence results because of discomfort or disability), following a certain path, after which they were subjected to subtests of navigational capacity. The VR test was performed in a virtual version of the same hospital lobby, from a wheelchair height perspective, after which the participants were subjected to the same subtests. The subtests consisted of route learning, free recall, self-orientation, route drawing, landmark recall, photograph recognition, photograph location, video location, as well as a neuropsychological test battery. The results indicated that the control group of younger adults showed the best navigational performance, both in real world and in VR environments, followed by the older control group and the MCI group, and lastly, the early AD participants. The VR environment resulted in lower navigational performance scores across all groups, compared to the real-world environment. Keeping in mind the slightly lower scores across all groups, Cushman et al. (2008) mean that their findings could indicate that VR could be useful to uncover navigational impairment, and found strong correlations indicating that navigational capacities of the real world, compared to VR, is not altered by aging or AD.

AD commonly affects a persons' ability to perform instrumental activities of daily living (IADL), activities that allow a person to function independently in daily life. Allain et al. (2014) created a non-immersive VR task to provide an assessment of IADL with high ecological validity, where participants were tasked to make a cup of coffee in a virtual environment. The environment consisted of a virtual kitchen with kitchen appliances and interiors in the background. In the foreground, on a countertop, were all items required to make a cup of coffee with milk and sugar, and three items unrelated to making coffee, acting as distractors (a bottle of wine, cocoa powder, and a fork). The task of making a cup of coffee was selected since it is a task many are familiar with, and requires multiple cognitive processes. The task in the study consisted of 14 steps that had to be performed in a sequential order. The study included 24 participants with AD, and 32 healthy controls, and measured time to completion of the task, accomplishment (what percentage of the 14 steps of the task was completed), total errors, omission errors (steps of the task omitted) and commission errors (the sequence of the steps of the task). Furthermore, the authors also compared with a real environment version of the coffee task. They found strong relations between scores of the real and virtual coffee tasks, as well as scores from an IADL scale, and found that AD participants' performance differed significantly in comparison to healthy controls. In addition, Allain et al. (2014) also found that scores between the real world task and the virtual task consistently differed between healthy controls and AD participants - the VR task resulted in lower scores for both groups, which may be important to keep in mind. One could also note that Allain et al. (2014) elected to use a non-immersive environment in order to reduce the risk of simulator sickness, which may cause particular assessment issues when used by elderly people.

Bellassen et al. (2012) utilized a non-verbal navigation test utilizing a VR maze. The study consisted of one group of participants with mild grade Alzheimer's, one group with aMCI, one group with frontotemporal lobar degeneration (FTLD) and one healthy control group, who had all undergone a neuropsychological examination consisting of a number of assessments. The virtual environment consisted of a maze with a joystick input device, with environmental cues surrounding the maze for orientation, and where participants were asked to find a goal placed at the same location in the maze. After a number of learning trials, the test phase began. The temporal memory tests consisted of sequential navigation, where the environmental cues were removed and participants asked to reproduce the sequence of turns they had performed in the last learning trail, and route tracing, where they had to trace the path from the last learning trial. Furthermore, participants were asked to name the environmental cues they had previously seen (the what-test), and place the cues they recalled on a map of the maze (spatial memory test/the where-test). The results showed temporal memory impairment in the AD group, and temporal memory deficit had high specificity and sensitivity for AD compared to FTDL patients and the healthy control group, which could be due to that the temporal memory test model hippocampal function well in comparison to other memory tests. Furthermore, the temporal memory test successfully and significantly discriminated aMCI patients from healthy controls - temporal sequence memory impairment of sequencing spatial locations has been reported for MCI. Of the 14 aMCI participants in the study, 7 were monitored over 3 years. 5 of the 7 participants converted to AD during this time, and these 5 participants had low temporal memory scores (below the cut off for MCI detection), which could indicate that the temporal memory test might help identify MCI patients who may convert to AD. In addition, the authors also assessed the healthy controls to detect memory degradation, to investigate if the temporal memory test could detect prodromal AD, and found that two participants of the control group showed signs of memory impairment, however not enough to be classified as aMCI. These participants had the lowest temporal memory test scores, which the authors claim indicates that temporal memory deficits could be tested to screen for very early stage AD.

Lesk et al. (2014) developed a virtual task where participants followed a path to a destination and asked to remember it, in order to assess visuospatial memory. The authors also claim that few assessment tests exist for detecting navigational issues in the early stages of the disease, and that VR has several benefits in comparison to pen and paper tasks and real-world navigation tasks. For example, VR could provide more control of the stimuli and complexity, as well as an accurate recording of data. Furthermore, Lesk et al. (2014) claim that pen and paper spatial navigation tasks lack ecological validity, since they cannot be generalized to real-environment navigation. The participants consisted of one group with MCI, and one healthy control group. The task consisted of a virtual path along which participants were asked to navigate, to reach a goal which they were required to remember. The participants were first asked to go through a learning phase, where they learned the controls (using the arrows on a keyboard to navigate), before either entering VR Park, a virtual park, or VR Games, a virtual environment with outdoor activities and games. Both VR modules had five difficulty levels, where the first was the easiest and the last was the most difficult, and the paths within both environments were mirrored. The study found significant correlations between scores from the VR task and neuropsychological tests, and successfully discriminated MCI participants from healthy elderly, specifically difficulty level 4 of the VR task. Lesk et al. (2014) are unable to explain precisely why difficulty level 4 showed the most correlation with the scores from the neuropsychological tests.

Tarnanas et al., (2013) employed a VR task based on a fire evacuation drill to assess executive function in participants with AD, aMCI and a healthy control group, respectively. The VR task was performed on a treadmill, with motion sensors and a Kinect which tracked the users, and the VR task consisted of a fire evacuation drill with six different, increasingly difficult, scenarios, in an apartment block. The task required multitasking and was selected with the motivation that it requires complex reasoning, because activities of daily living (ADL) that require complex reasoning are sensitive to cognitive impairment, as well as functional impairment. The VR task was intended to show how the participants reasoned, planned, initiated and executed a number of tasks to evacuate the apartment block. They found that the AD group was the most impaired, followed by the aMCI group, compared to healthy controls, and that the results correlated with other measures of cognitive function such as MMSE and ADL scores.

Yeh et al. (2012) conducted a study to evaluate a system to assess executive function and memory utilizing VR, and created a virtual convenience store. The tasks had multiple levels of difficulty, and consisted of memorizing a shopping list, searching for items, and checking out at the register. The study used a HMD embedded with posture sensors to track the participants' movements and line of sight. The participants interacted with the system with a joystick. The participant is first asked to memorize a shopping list with 1 to 10 items, before entering a store to find the items on the list. A shopping cart can be shown in the corner of the screen, and/or the shopping list, if the evaluator deems it appropriate. After gathering all items, the participant is able to check out and pay, by selecting the correct amount of money. The system was evaluated by two healthy participants, who generally rated the system positively.

To summarize, several studies have attempted to utilize VR to assess various aspects of cognition. Several of the concerned studies (Howett et al., (2019), Cushman et al.,(2008), Bellassen et al.,(2012), and Lesk et al. (2014)) have utilized different types of navigation or orientation tasks, with promising results regarding the detection of impairment related to MCI and AD. Plancher et al. (2011) utilized a VR driving task to characterize episodic memory as well as spatial memory. Tarnanas et al. (2013) and Yeh et al. (2012) focused on executive function, while the coffee task construted by Allain et al. (2014) aimed to assess everyday action deficits by creating a task which required multiple cognitive processes. The results from these studies all indicate that VR may be a useful tool to assess several aspects of cognition, however with certain considerations, such as the risk of lowered performance both in healthy and impaired participants.

3.4.1 Considerations of using VR in a healthcare setting

There are numerous aspects one needs to consider when utilizing VR within healthcare. There may be a number of potential side effects of using VR, such as disorientation, simulator sickness and physical discomfort (Costello, 1997). Furthermore, Costello (1997) suggest that hygiene may be an issue that needs to be considered in public settings, since like mouses and keyboards, they may harbour pathogens, and since the Head-mounted display (HMD) may cause people to sweat and become warm, and which could cause bacteria to thrive. This may be especially important to keep in mind in healthcare settings, where people with a lowered immune system may be using the equipment. Jerald (2016) also brings up concerns regarding hygiene, such as make-up and oil residues, as well as issues such as lice, and brings up options to sanitize the HMD. Jerald (2016) points out that cleaning with alcohol may not be sufficient, since the parts of the HMD closest to the face often are made of a porous material, but that there are washable covers that can be used.

Costello (1997) also points out that VR that is to be used by many, such as in healthcare or other public settings, may be difficult to customize so that the HMD fits properly and does not cause discomfort. Another potential issue is the interpupillary distance (IPD) which can be one potential cause of discomfort, and requires calibration for each individual who uses the VR system (Jerald, 2016) Calibrating the IPD could potentially become problematic for the medical personnel who need to be able to ensure that it is properly calibrated, together with the patient (which in turn may be difficult for people who are not used to using VR).

One example of when VR may be beneficial in healthcare and cognitive assessment, is accessibility to people of different educational background and literacy, as well as people with a different cultural backgrounds and adaptable to people who speak different languages. According to Franzen et al. (2019), neuropsychological tests are known to be affected by these factors, which may be problematic, since many tests were developed for people from Western populations (Franzen et al., 2019). VR has the potential to be language-independent, or easily adaptable, which is extremely important in healthcare, since one needs to be able to detect cognitive impairment and diagnose it in order to be able to help people. Franzen et al. (2019) suggest that VR may be beneficial in assessing people with a lower educational background, since it can be created with a higher ecological validity.

3.4.2 Adverse effects of VR

VR can have a number of adverse effects that cause discomfort for the user. *VR Sickness* is a term used to encompass the various types of symptoms that may arise after using a VR system. Typical symptoms include nausea, disorientation, dizziness, eye strain, headaches and general feelings of discomfort. VR Sickness can be rooted in multiple different causes, ranging from hardware, software, and the individual user. Motion sickness, which is included in VR Sickness, can for example result from when vestibular and proprioceptive cues are mismatched to the visual cues - when the physical position and movement of the body does not match what they eye sees. Adverse effects from VR can also include factors related to uncomfortable positioning of the body (for example, holding the arms up for too long, resulting in so-called 'Gorilla Arms'), repetitive movements, or standing or walking for too long (Jerald, 2016).

Luckily, there are numerous methods to counteract VR Sickness through design. Jerald (2016) proposes utilizing a rest frame which is an object that is stable relative to the user, even if the user moves in the virtual environment. One example would be having a cockpit which the user is placed in, which can help stabilize perception. Other methods that can, or perhaps rather should, be considered when designing for VR includes allowing the user to be in control of motion, avoiding visual acceleration and excessive physical head motions. Furthermore, keeping the VR experience fairly short may lessen the risk of VR Sickness, as well as keeping the virtual world upright.

Individual user factors that may contribute to VR Sickness include having a history of motion sickness or migraines, having little prior experience of VR, poorly calibrated HMD (interpupillary distance, the distance between the eyes), a poorly fitting, or poorly placed, HMD which can cause distortions to the image as well as discomfort. Being in poor health can also cause VR Sickness, for example ear infections, respiratory symptoms, emotional stress or fatigue are among the conditions where using VR would be ill-advised.
In addition, the risk of VR Sickness increases with age, it is unclear as to why, but it could be caused by level of experience, sense of balance, lacking eye-accommodation to objects on screen, capability, and so on (Jerald, 2016).

3.4.3 User acceptance

Flynn et al. (2003) applied a user-centered method to study the experiences of a virtual environment by participants with dementia, in order to investigate numerous possible feasibility issues. The VR environment for the tasks was a park, and a joystick was used for input. The first exercise aimed to investigate the quality of the experience - experiencing presence, navigation using the joystick, whether the objects and their movements appeared realistic, and if the participants felt in control. The second exercise consisted of IADL tasks, such as making a phone call, mailing a letter, finding a trash can to throw away litter, and finding somewhere to sit down to rest, and where the participants were allowed to freely explore the virtual environment.

The participants' well-being was monitored by measuring heart rate and a questionnaire to evaluate simulator sickness. Six participants took part in the study, three male and three female, between the ages of 52 and 91, and only one of the participants had previous experiences with virtual environments. Flynn et al. (2003) found no significant results indicating a decrease in physical or psychological well-being, and participants generally felt in control of the environment, and were able to interact with the environment using the joystick. Furthermore, they did not find simulator sickness to hinder the use of virtual environments with participants with dementia, but recommend taking precautions both before, during and after exposure to virtual environments. The sample size of the study was small with only six participants, however the study found that it was indeed feasible to use virtual environments with people with dementia.

One important note is that Flynn et al. (2003) took measures to minimize the risk of adverse effects of VR, such as interviews before the experiment to establish psychological well-being, screening for susceptibility to motion sickness, vertigo and epilepsy, as well as limiting the time spent in VR and restricting traveling speed in the virtual environment to further reduce risk of adverse effects. The participants also had a carer or relative with them at all times that could help them feel more comfortable, and offer support if the participant became distressed.

Manera et al. (2016) conducted a study to investigate the feasibility to use highly realistic image-based VR with people with MCI and dementia, in terms of usability, user acceptance and interest. The study consisted of 57 participants, 28 with MCI and 29 with various types of dementia. The task consisted of a VR training task for selective and sustained attention, which was compared to a paper version of the same task. Participants performed both tasks in a single session, in a randomized order, after which they were asked which they preferred, in a questionnaire where they were asked to rate levels of satisfaction, interest, discomfort, anxiety, feelings of security and fatigue. Participants reported high levels of satisfaction, security and interest of the VR task, as well as low levels of discomfort, fatigue and anxiety. Furthermore, participants preferred the VR task over the paper task, and even though it was more difficult and yielded a lower performance score, 70% of the participants stated that they preferred VR. Even apathetic participants showed great interest in the VR task in comparison to the paper task, and a significantly higher preference compared to non-apathetic participants.

Mrakic-Sposta et al. (2018) attempted to evaluate the effects of a combination of physical activity and cognitive training utilizing a VR system in people with MCI, to investigate how it could be applied to mitigate impairment and oxidative stress, as well as user acceptance. 10 participants with MCI took part of the study, for three times a week during six weeks. The authors created three virtual environments - one related to physical exercise (riding a bike through a park), and two cognitive training tasks - grocery shopping and crossing a road where the participants had to avoid cars. The user acceptance was high for the VR tasks, participants reported high levels of engagement and motivation, and all participants rated the experience both acceptable and enjoyable. The participants reported that interacting with the technology was intuitive, and few reported sickness related to VR.

4
Methodology

This section discusses the various methodologies used within the study, as well as methodologies that were considered.

4.1 Research methods

4.1.1 Literature studies

The literature studies aimed to investigate various topics. First of all, they consisted of a further exploration of cognitive impairment, with some focus on AD, dementia and MCI, in order to attempt to gain a better picture of cognitive impairment, what actually happens to those affected, and how the it can progress. A thorough understanding of different types and causes of cognitive impairment was necessary in order to be able to investigate how VR could be beneficial. Furthermore, currently used diagnostic methods and tools needed to be explored, in order to understand how cognitive tests and neurpsychological assessment currently are performed, and which types of cognitive aspects could be valuable and feasible to attempt to assess utilizing VR. Previous studies of VR used to assess cognitive function or detect AD or MCI were also be expanded upon. Many of the previous studies have focused on navigation (for example, Cushman et al. (2008), Plancher et al. (2011) and Howett et al.(2019)), which may indicate that navigational tests have some merit to be performed using VR, especially for AD and MCI, but investigating whether other cognitive aspects could be applied in VR is also of great importance. Because of this, some other examples of how VR could be used in healthcare were explored, as well as other examples of technology mediated diagnostic tools. The literature studies also concern VR in general, such as ergonomic considerations, guidelines for design, and special considerations one may need to consider when using VR in a clinical context.

4.1.2 Interviews

The aim of the interviews was to investigate further into the process of assessment as it is currently - what are the positive aspects with using these methods? Could some of these aspects be beneficial to translate to VR? Which are the negative aspects? Which benefits are expected to be gained from utilizing VR? Furthermore, the interviews concerned some aspects of the expected usability of VR. Patients' reception of VR is difficult to predict, since no patients were involved in this study, however one could at least attempt to anticipate some potential issues that may arise when utilizing VR.

The interviews were decided to be semi-structured, with a combination of planned out questions, free conversation and follow-up questions, in order to obtain the information required, while still allowing the experts to freely associate and enlighten the topic. (Martin & Hanington, 2012.) The purpose of the interviews were not to gather vast amounts of data, but rather extract important information and insights into the domain, and used in both an exploratory and generative capacity. In order to cluster the interview data, and to understand the context and the most critical aspects, interview data could be analyzed in an affinity diagram. An affinity diagram could aid in finding common themes that need to be considered for design, as well as key concerns that may arise when using VR in this context (ibid).

4.2 Design, prototyping and evaluation

Designing the system benefits from a human-centered design approach, since the involvement of the people who would be using this system could be beneficial for creating a more useful and usable design. The design resulted in a lower fidelity prototype, at an early stage of development, which could benefit from expert Think-aloud evaluation (Martin & Hanington, 2012) to uncover possible issues.
As Gaver (2012) points out, it may be difficult to produce verifiable or falsifiable results through design, but it could serve as a starting point for further research. As such, the creation of a prototype could act as a starting point for further research into the topic of how VR could be designed and applied to assess cognitive function. According to Gaver (2012) research through design is generative in nature, asking questions of 'what might be'. Furthermore, by attempting to design a VR system based on theories both from design as well as theories of cognitive assessment, the resulting artefact could be viewed as an attempt to create an embodiment of these theories (Gaver, 2012). Since this project is explorative in nature, the 'what might be' has a lot of emphasis, it aims not only to answer the question of *if* VR could be applicable to assess cognitive impairment, but also *how* such a system could be designed, and act as a stepping stone for future projects.

4.2.1 Design process

This project followed the IDEO Human-centered design process (IDEO, 2015), with heavy focus on the research and understanding part, and less focus on the design and implementation phase. A human-centered design (HCD) approach may be beneficial in actually understanding how VR would be applied for real, and detect potential usability issues, both for the medical personnel and for the patients. Actually having one of the user groups, medical professionals, involved throughout the design process, and letting them have some decision making power, in order to produce a more usable design that supports the users in their work. The HCD process usually consists of three main phases: Inspiration, ideation, and implementation (IDEO, 2015). The inspiration phase consisted of understanding the people involved, in this project primarily through literature and interviews. Within the HCD process, the ideation phase consists of making sense of the new knowledge and generate and test ideas, and creating prototypes. The implementation phase usually consists of making the design a reality and bringing it to market, however in this project, the implementation phase will halted at the early prototype stage, and rather focus on inspiration and ideation. The HCD process is iterative, and within this project, the prototype was planned to run through two iterations, followed by expert evaluation, from one of the potential user groups. Since this project was more exploratory in nature, no users from the second user group, the patients, were involved.

4.2.2 Prototyping for VR

There are numerous methods to produce Virtual Reality prototypes. The highest fidelity prototype would likely be to utilize a game engine such as Unreal Engine (unrealengine.com/en-US/vr, retrieved 2020-05-21) or Unity (docs.unity3d.com/Manual/VROverview.html, retrived 2020-05-21) to produce a prototype. In producing a high fidelity prototype, one could truly get a sense of what a final product could look and feel like, and produce a good basis for thorough evaluation. However, this may also be time consuming and complex, and since the focus of this project was to investigate the feasibility of utilizing VR and how this type of system could potentially be designed, it may be far too early in the process to spend time designing and developing such a prototype. A high fidelity prototype could be the next step after this project is finalized. A lower fidelity prototype is also easier to iterate and redesign, compared to a high fidelity prototype in which one has invested a lot of time and effort.

Luckily, there are other options to create prototypes of a lower fidelity, which is a more appropriate level for this project. One very simplistic method is to take or create panorama images and view them with Cardboard viewer. However this also has very limited possibility of interaction. Another option is to use Google Blocks (arvr.google.com/blocks/, retrieved 2020-02-20) to create a rudimentary 3D prototype which can be viewed through a HTC Vive or Oculus Rift HMD. The downside is that Google Blocks is bound to Vive or Oculus Rift, but it can give a good first sense of look and feel of the prototype.

Marvel App is another alternative to create VR prototypes for Cardboard, with hotspots which actually allows for simulated interaction, by allowing the user to move between frames of the prototype (De Greve, 2018).

Another option was to utilize the web-based framework A-Frame, which can be used to create a prototype using HTML, CSS and Javascript. There is a lot of freedom in using A-Frame, which can used for multiple different platforms, and the level of fidelity is very flexible - one can create a simple prototype, or a more advanced design (aframe.io, retrieved 2020-01-24). Another A-Frame option is to utilized the Sketch-to-VR plugin (github.com/auxdesigner/Sketch-to-VR, retrieved 2020-01-24) which automatically creates an A-Frame website, allowing more focus on the design.

In other words, there are several options to create VR prototypes. Selecting which prototyping tools that were to be used needed careful consideration, since there are potential positive and negative aspects of each one, and they have different learning curves. The most complex aspect to prototype may be incorporating interactivity or simulated interactivity. However, interactivity may not be necessary, partially because of the low fidelity of the prototype, but it is also dependent on what is deemed appropriate considering the type of cognitive assessment the prototype would be attempting to perform, some may need interaction with the prototype, other types may be relying on other factors, such as timing or verbal feedback.

After some consideration, A-Frame was selected in order to obtain the possibility to create a more immersive virtual environment, and incorporating some form of interactivity. A-Frame was also deemed to be relatively easy to learn. Furthermore, selecting A-Frame allowed for great flexibility both in terms of fidelity of the prototype, and the ability to select a platform more freely.

4.2.3 Evaluation of prototype

Since this is a highly specialized field, the insight of the people who would actually be using the product is invaluable, which is why the input from medical professionals who are the potential users of the product was deemed to be important. Within this project, expert evaluation with medical personnel, who work with diagnosing cognitive impairment Alzheimer's is more relevant, since they have the experience and know-how of cognitive assessment to be able to detect issues with the design. Since the prototype was of a lower fidelity, the possibility to quickly redesign could allow for more exploration and evaluation of the design as issues occur. A Concurrent Think-aloud protocol, asks the participant to complete a task and is prompted to 'think aloud' and explain their thoughts while completing the task (Martin & Hanington, 2012). This could be a useful method to extract the experiences of the prototype, and detect potential issues in the moment they occur. This would be followed by a short interview to further discuss the positive and negative aspects of the prototype.

5
Planning

In this section, the planning of the project is discussed, and an overview of the time plan and its' steps is provided.
Below is a Gantt schedule (Figure 5.1) showing the different steps of the project and their estimated time frame.

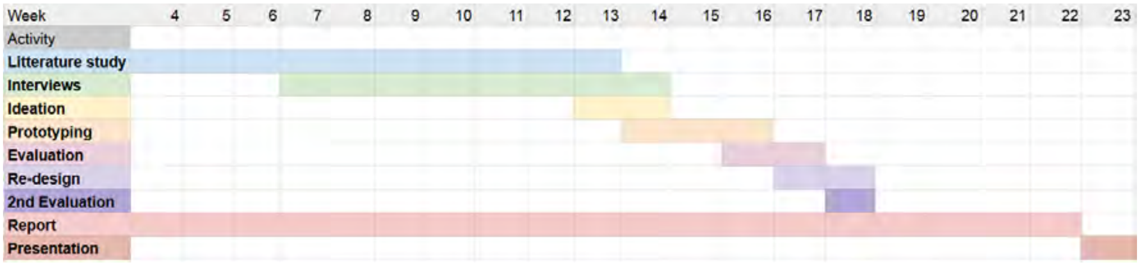

Figure 5.1: Gantt schedule visualizing the different activities of the project and their respective timing, by week.

5.1 Literature study

The literature studies made up a large part of the project, since the main focus lies on the theoretical aspects of evaluating the feasibility of utilizing VR as a tool for evaluating cognitive function in a clinical setting. This part of the project largely overlapped with the interviews, since they have a similar aim.

5.2 Interviews

The interviews with the stakeholders and experts aimed to lay a theoretical foundation regarding how cognitive function is assessed. Furthermore, to provide insights into currently used diagnostic tools and assessment methods, and their advantages and disadvantages, from the point of view of medical personnel. In addition, the interviews provided crucial insight regarding the design of the prototype, by addressing the various expected advantages and disadvantages of VR, which aspects of currently used assessment methods could be beneficial to attempt to include in the prototype, and the types of tasks that could be advantageous to perform in VR.

5.3 Ideation, prototyping and evaluation

The ideation, prototyping and evaluation phases of this project were relatively short in comparison to the literature study and interviews, since the main focus is not on the design of the prototype itself. Developing and evaluating the low-fidelity prototype was planned to go through two iterations, followed by expert evaluations to obtain in-depth insights and design considerations that need to be kept in mind.

5.4 Report and presentation

The master thesis report has been written throughout the entire project, since this project has a strong focus on theoretical aspects. The report was be finalized during four weeks towards the end of the project. The final week of the project will consist of preparations for the thesis seminar.

6
Process

6.1 Literature study

The literature study made up a large portion of the project, with the goal to examine the feasibility of utilizing VR to assess cognitive impairment. The literature study consisted of reading and summarizing literature divided into two main topics - *cognitive impairment* with a specific focus on AD and MCI, and *VR*. These topics were then further divided into sub-categories, which to some extent merged.

The cognitive impairment topic consisted of clinical aspects of cognitive impairment, AD and MCI, diagnostic methods, neuropsychological assessment and cognitve screening, as well as currently used tools, and proposed technologies for technology mediated assessment. These aspects were given a specific focus early in the project in order to gain an thorough understanding of the disease and cognitive impairment. Furthermore, to understand more specifically which types of cognitive abilities commonly are impaired during the early stages of dementia, and how these abilities currently are being assessed. Another important point was to find possible drawbacks of these currently used methods, and whether they possibly could be improved using VR.

The VR category consisted of the sub-categories design for VR, VR Sickness and ergonomics, VR used within healthcare in general (for example, training, treatment or diagnosis of other conditions), VR used with AD or MCI patients in general (for example, memory training, IADL assessment or training, or for therapeutic purposes), user acceptance and usability of VR among people with AD/MCI. In addition, to uncover specific considerations of using VR in a healthcare setting, and VR used specifically for neuropsychological assessment or cognitive screening.

Within the literature study, specific focus was given to the studies aiming to differentiate, diagnose or otherwise detect cognitive impairment related to AD or MCI, utilizing VR. The results from the literature study were compiled into a list of possible issues that may arise, as well as possible advantages, of using VR within this context.

The literature study also acted, partially, both as inspiration and understanding phase for the design of the prototype. By investigating previous studies of VR used for cognitive assessment, this previous research was intended to aid in making more informed design decisions within the context.

6.1.1 Summarization the literature study

The literature study will be summarized here to provide an overview of potential problems and advantages that should be considered when applying VR to assess cognitive function. These problems and advantages concerns the use of VR from a holistic perspective, taking multiple factors into account. Eight potential problems were identified, and nine possible advantages. Figure 6.1 provides an overview of all the identified problems and advantages, all of which are discussed in further detail in section 6.3 and 6.4, respectively. One could note that the factor 'ergonomics' appears as both a problem, and an advantage. A lack of consideration for ergonomics can cause discomfort when using VR. However, if taking proper precautions to mitigate these issues through design, VR could also facilitate user comfort. For example when conducting a navigation task, elderly or disabled users could be allowed to sit down and navigate within a virtual environment, instead of walking through a physical environment, which may be tiring.

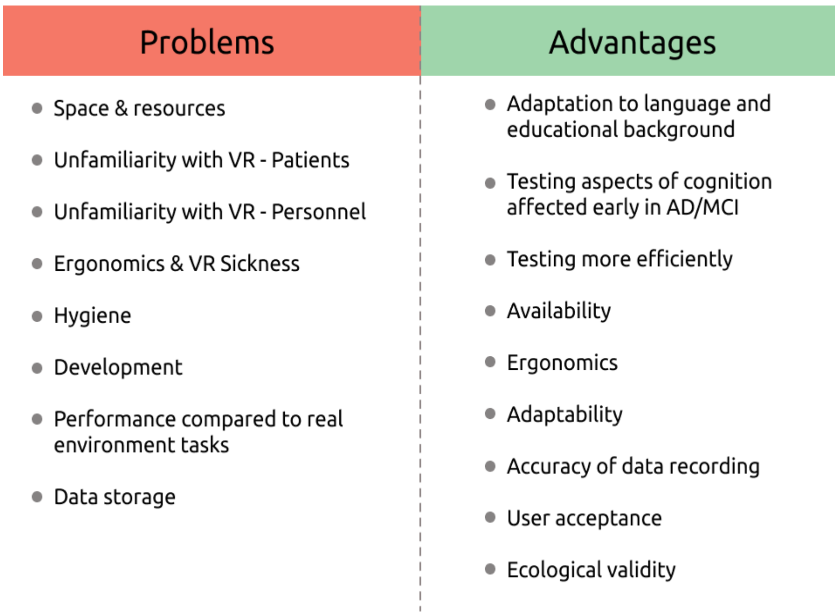

Figure 6.1: Identified problems and advantages of utilizing VR for cognitive assessment

6.2 Problems

6.2.1 Space and resources

VR systems come in several shapes and sizes and levels of complexity, ranging from full-sized immersive simulators, HMDs and controllers that can be used with a personal computer or gaming systems, or smartphone HMDs such as Google Cardboard. Mobile VR are VR systems that are easily portable and allow the user to use the system almost anywhere, at any time, by placing their smartphone in a HMD. Location-based VR, on the other hand, usually requires more set-up, and are not as portable, they can be systems that can be used at home, or larger set-ups as seen in arcades or VR entertainment centers, for example. (Jerald, 2016) Selecting between mobile and location-based VR comes with certain trade-offs: location-based systems are usually more advanced and can provide better immersion and tracking, but can be large and cumbersome to set up, and are not easily moved. Mobile VR systems, on the other hand, are very easy to move around, with the minimal requirement of a HMD and a smartphone they can be used in almost any location. On the other hand, Mobile VR systems are usually not as high-end as location-based systems (ibid). With regards to monetary resources, VR systems vary greatly in price depending on which type of system and quality - Mobile VR systems are generally fairly cheap, while location-based VR may be more expensive.

6.2.2 Unfamiliarity with VR - Patients

VR systems have existed in various forms for quite a long time, but in the 1990's the interest for VR rose with numerous companies investing and developing VR systems, mainly for location-based entertainment systems or research purposes. However in the early 2000's, the so-called VR winter began, partially due to technological shortcomings, where VR gained little attention outside of academia, military and corporate research settings. It is only in fairly recent years, starting from 2012, interest in VR has increased, and VR systems have become widely available to consumers (Jerald, 2016). In other words, it is not unreasonable to assume that many have little to no experience of VR systems. According to Jerald (2016), not being used to VR is a risk factor for VR Sickness, which lessens the more familiar one becomes with VR.

Being unaware of how the HMD is supposed to fit, and how the image on the screen is meant to look can cause VR Sickness, which means users need to be instructed to properly secure and adjust the HMD. A poorly fitting headset can result in discomfort and headaches, and be especially uncomfortable if the user is wearing eyeglasses underneath. If the headset is too loose, it could also wiggle and cause some distortion to the image, increasing risk of VR Sickness (ibid).

Knowledge of how to interact with the system is also important, however fairly system-dependent since there are several possible methods of interaction. Facilitating interactions require careful design considerations in order to aid the users in achieving their goals within the system. Ideally, there should be an onboarding task where users can first learn how the interactions work, and help them feel secure and capable before taking on the real assessment task. A poorly designed task could potentially lead to misleading results, if not enough care is taken to lessen the risk of VR Sickness or other aspects related to not being accustomed to VR.

6.2.3 Unfamiliarity with VR - Medical personnel

Not being accustomed to VR can also affect medical personnel, who will have to learn to use a new system. The systems can range in complexity, and setting up the system can be troublesome. Some systems may be easier to use than others, such as mobile VR, which only requires a headset and smartphone.

Furthermore, possibly having to calibrate the system to the patients interpupillary distance (IPD) could be bothersome - some systems require manual hardware calibration, while some have software-based calibration (Jerald, 2016). Assuring proper headset fit, and ensuring that the image looks correct can be difficult from an outside perspective, and would require instructing the patient to ensure they feel comfortable. Furthermore, some VR systems may require the user to walk or move around in the real environment while immersed in VR. This can be a safety concern, since the person wearing the HMD will be unaware of the real world surroundings, and may require an outside person keeping track and ensuring that they do not trip or otherwise hurt themselves (Jerald, 2016). In other words, there may need to be a third person as a safety precaution, aside from the patient and the person performing the assessment.

6.2.4 Ergonomics and VR Sickness

The risk of being susceptible to VR Sickness increases with age (Jerald, 2016). This becomes especially relevant to consider when one notes that AD typically onsets after the age of 65, although some cases may have an earlier onset (Funke & Willbold, 2011).

Furthermore, people in poor health may also be more susceptible to feeling sick when using VR - for example, respiratory symptoms, flu, hangover, sleep deprivation, emotional stress, fatigue, dehydration, blockage of ears or ear infection, as well as certain medications - may all exacerbate VR Sickness (Jerald, 2016). Keeping these aspects in mind are important in order not to introduce VR Sickness, which may not only be uncomfortable, but could also potentially affect performance on the VR task.

Other factors to consider is physical fatigue, such as so-called Gorilla Arms, fatigue resulting from keeping one's arms in a raised position during an extended period, which can result from using VR, as well as physical discomfort related to positioning of the body and head. Another factor is interpupillary distance, the distance between the eyes typically ranging between 45-80 millimeter, which needs to be calibrated for each user in order to avoid discomfort and eye strain (Jerald, 2016).

6.2.5 Hygiene

Hygiene is an important factor to keep in mind, especially within a medical setting. HMDs can harbour viruses, fungi and bacteria and risk transmitting diseases between users. The material closest to the face is typically porous and difficult to clean, even with alcohol. However, there are options. The lenses can easily be cleaned between users, and there are removable liners for the parts closest to the face available that can either be washed or disinfected with alcohol (Jerald, 2016) Another option could be to use the mobile headsets made out of cardboard, as a disposable headset thrown away after use - this is, however, not a very environmentally conscious choice, and may be considered wasteful.

6.2.6 Development

The process from beginning to investigate the possibility of utilizing VR to screen for cognitive impairment to being able to actually implement and being using it in a clinical setting may be quite long. Developing and designing a prototype, testing and evaluating it, and evaluating based on a normative sample to make the results of the prototype possible to generalize, and establishing a baseline of normal performance are all necessary steps. After following these steps, one could discover if it is possible to use for its' intended purpose. These types of steps are typically required when developing other types of assessment and screening tests, however, VR has the additional layer of ergonomic issues and having to design to avoid VR Sickness. These issues may not be as prevalent in pen and paper tasks, or tasks performed on a regular desktop computer or touchscreen. This means that developing a VR assessment task may require a number of special considerations which could make the development of a prototype more time consuming.

6.2.7 Performance compared to real environment tasks

Cushman et al. (2008) created a navigational task in a hospital lobby, to be performed in either VR or in the real environment. The authors found that all participants performed worse in the VR task, in comparison to its real world counterpart task. However, this dip in performance was found across all participant groups - healthy elderly controls, healthy young controls, as well as the MCI and AD groups, respectively. Similarly, Manera et al. (2016) constructed a VR training task for selective and sustained attention, and compared it to a pen and paper task, and found that the VR task yielded a lower performance score. Allain et al. (2014) created a virtual kitchen to assess abilities related to instrumental activities of daily living. The task consisted of making a cup of coffee in this virtual kitchen, completing 14 steps in a specific order. Allain et al. (2014) also found that scores between the real world task and the virtual task consistently differed between healthy controls and AD participants - the VR task resulted in lower scores for both groups,
In other words, VR risks resulting in a lower performance, which could be important to keep in mind, however by establishing a normalized baseline this does not have to cause any issues, as long as one is aware of the possibility. This lowered performance could potentially result from not being familiar with VR, or possibly that VR tasks are more difficult. Allain et al. (2014) points out the risk of VR increasing the cognitive load, or that VR requires more attentional resources.

6.2.8 Data storage

Digital assessment may entail storing data, which could become problematic, especially in a healthcare setting. Ensuring that patient data is stored securely and that it cannot be accessed wrongfully is imperative. Furthermore, it is important that both patients and medical personnel feel that the system is transparent and trustworthy.

6.3 Advantages

6.3.1 Ability to test independent of language or educational background

Few of the studies discussed here actually discuss the possibility of utilizing VR to bridge language barriers. However, it is not unfathomable that VR could be used to create assessment tasks or screening tests that are usable regardless of a patients native language or level of education. Franzen et al. (2019) suggest that VR may be beneficial in assessing people with a lower educational background, since it can be created with a higher ecological validity.

The native language of a patient may affect the result of some cognitive tests and neuropsychological assessment tasks. In VR, one could either make a purely non-verbal task, or one could create multiple versions of the same task in different languages. Creating the same task in different languages, or adapted for different cultural backgrounds would need for the task to be evaluated in each version. The results could still be presented to the clinician in a language they understand, while keeping the task in the patient's language, lessening a possible language barrier. Creating a purely non-verbal task could have the benefit of being able to assess independent of native language, and assessing based on performance of the task. For example, Allain et al. (2014) created a coffee-making task that evaluated performance with regards to time, number of steps accomplished, commission and omission errors, and total number of errors. One can, however, note that many of the studies include some form of verbal instructions, or verbal assessment after the virtual task. For example, Bellassen et al. (2012) created a non-verbal spatial and temporal navigation task, where participants navigated a maze. Participants were asked to perform and then reproduce a sequence of turns within the maze, and then trace the path on a map. However, participants were also asked to name cues they had seen in the environment, and place these on the map - the naming of cues may prove difficult to those with a different language background, however although an interpreter could aid.

6.3.2 Ability to test aspects of cognition affected early in AD/MCI

According to Cushman et al (2008), loss of navigational ability and disorientation are common early signs of AD, however few behavioural measures exist to assess these aspects, which risks hindering the early detection of AD. The authors claim that VR could be a useful tool to detect these navigational issues. Allain et al. (2014) points out that VR could be used to detect more subtle deficits otherwise difficult to detect. Howett et al. (2019) conducted an entorhinal cortex-based VR path navigation task study to differentiate MCI. The study found that the task had superior accuracy in classifying prodromal AD, compared to commonly used cognitive tests that assess episodic memory, attention and processing speed.

6.3.3 Ability to test more efficiently

According to Lesk et al. (2014) navigation tasks investigating spatial navigation can be particularly useful to detect Alzheimer's disease. Navigation tasks in real environments have the downside of being changeable - they may look different on different days, with different stimuli present. VR environments, on the other hand, are more consistent, and there is a greater degree of control over the environment. Cushman et al. (2008) also point out that VR tasks could be performed in a less time-consuming, and less difficult, manner compared to real world navigational tasks which require moving around a physical space.

Computerized tasks have the risk of being too demanding of fine motor skills. These typically decrease with age and can be affected by disabilities brought on by aging, which can make interacting with the typical input devices (mouse and keyboard) more difficult, particularly for those not accustomed to it (Allain et al. 2014). VR can be made to demand less of a patients' fine motor skills, by using larger gestures with hand controllers, and making the items interacted with on screen larger than they would be on a standard computer monitor, for example. This could potentially make the assessment process faster to some degree.

6.3.4 Availability

Laske et al. (2015) points out that neuroimaging often can be unavailable outside of specialist clinics, as well as expensive. In terms of expenses, VR has a large price range depending on the type of system - mobile VR are quite inexpensive, while location-based VR may be more expensive. VR can also be comparatively portable. In other words, VR could be made more easily available outside of specialized settings, due to its' portability and varying price-range.

6.3.5 Ergonomics

While ergonomics and VR Sickness may cause an issue, one can take measure to lessen the risk. Flynn et al. (2003) found that people with AD are not especially susceptible to VR Sickness, and that participants generally felt in control of the environment and safe in their ability to interact with it. However, Flynn et al. (2003) also took measure to minimize the risk of VR Sickness beforehand such as interviews before the experiment to establish psychological well-being, screening for susceptibility to motion sickness, vertigo and epilepsy, as well as limiting the time spent in VR and restricting traveling speed in the virtual environment to further reduce risk of adverse effects. The participants also had a carer or relative with them at all times that could help them feel more comfortable, and offer support if the participant became distressed.

The design of the system can also affect the risk of motion sickness, and by taking precautions when designing the VR experience, one can minimize these risks. For example, allowing the user to be in control of navigation, avoiding visual acceleration, designing to avoid frequent head motions, and keeping the VR experience short. Furthermore, aspects such as standing up or walking instead of sitting down can also cause VR Sickness (Jerald, 2016).

VR navigation tasks also have the benefit of not requiring moving through a physical space. Real environment navigation tasks often require a participant to walk around in an environment, which can become tiring and difficult for elderly or disabled patients. Cushman et al. (2008) solved this by placing participants of a real environment navigation task in a wheelchair, while the VR task was performed sitting down. Laske et al. (2015) points out that currently used diagnostic tools can be time-consuming (neuropsychological assessment), invasive (lumbar puncture), or expensive (neuroimaging). VR has the benefit of being non-invasive, and as long as special care is taken to minimize the risk of VR Sickness and other ergonomic issues, it could be a good option with regards to patient comfort.

6.3.6 Adaptability

VR could provide more control of the environment and the stimuli present, in comparison to pen and paper tasks or real world navigation tests, as well as the ability to control the complexity (Lesk et al. 2014). Allain et al. (2014) also points out the ability to alter the complexity of the task in virtual environments, while maintaining control over measurements.
VR allows for more precise control over the stimuli presented, in comparison to the real world which may be full of distractors. This control can be used to adapt the difficulty level - for example, if a patient starts out at the middle level of difficulty, and performs poorly, they can be allowed to try a lower difficulty level, or conversely, if they perform very well, the difficulty level can be increased. One could think of it like levels of a game, where the clinician can observe the performance and decide if the patient should level up, or go down a level.

6.3.7 Accuracy of data recording

Digital testing in general can also provide more accurate measurements, for example time, number of errors, and other data, in comparison to relying on manual note-taking and timing (Lesk et al. 2014). By having a computer keeping track of the number of errors, timing, and other measurements, the medical personnel could also be able to focus more on assessing what the patient actually does and how they solve the task. In addition, one could implement additional measurements that are otherwise difficult to observe, using for example eye tracking

6.3.8 User acceptance

Manera et al. (2016) conducted a study to investigate the feasibility to use highly realistic image-based VR with people with MCI and dementia, in terms of usability, user acceptance and interest, and compared it to a pen and paper task. The results were generally positive, participants reported low levels of discomfort, fatigue and anxiety, and high levels of satisfaction, security and interest. While the participants performed worse than in the corresponding pen and paper task, 70% of participants still reported that they preferred the VR task.
Mrakic-Sposta et al. (2018) conducted a study to evaluate the effects of a combination of physical activity and VR-based cognitive training in people with MCI. They found a high user acceptance for the VR tasks, and that participants reported high levels of engagement and motivation, rating the experience as both acceptable and enjoyable. The participants reported that interacting with the technology was intuitive, and few reported sickness related to VR.
In other words, VR used as an assessment tool can be designed to be enjoyable, which could potentially increase the comfort of patients.

6.3.9 Ecological validity

Several studies point out the possibility of utilizing VR to provide a higher ecological validity in neuropsychological assessment and cognitive screening. For example, Lesk et al. (2014) points out that spatial navigation tasks performed with pen and paper lacks ecological validity, while VR can provide a more ecologically valid platform for assessing spatial navigation without the problems that arise when assessing navigation in a real environment, Furthermore, Plancher et al. (2012) points out that it may be beneficial to make neuropsychological assessment more closely resemble real life.

Allain et al. (2014) point out that tasks that more closely resemble real environments may increase ecological validity when assessing function related to instrumental activities of daily living, and found significant relations between the scores of a real environment and a virtual environment task. In other words, VR can create environments more closely resembling real situations, in comparison to, for example, certain types of pen and paper tasks, while at the same time being more convenient than some types of tasks, particularly assessing navigation and orientation.

6.4 Interviews

The stakeholder interviews had two main purposes: gaining a further understanding of how cognitive ability typically is assessed in practice, and to inform the design of the prototype. The interviews were semi-structured, with some questions prepared but still allowing for free discussion. The early stage interviews focused on understanding the process of assessing cognitive function, the various tools and tests used in the process, as well as their benefits and drawbacks. This also included showing different types of tools and screening tests used to detect cognitive impairment, such as a computerized Useful Field of View Test, Koh's Block Test, the Boston Naming Test, Corsi blocks and The Clock Drawing Test. Koh's Block Test consists of colored blocks that the patient is asked to arrange to create designs and patterns (Loebach Wetherell et al. (2002). The Boston Naming Test consists of 60 line drawings of items the patient is asked to identify, ranging from more common to uncommon items (Hall, O'Caroll & Frith, 2010) and Corsi block-tapping test consists of tapping on colored blocks in a sequence, assessing visual memory (Kessels et al. 2000). Some of the key benefits of these types of tests were also discussed, and whether these benefits possibly could be carried over to VR.

The notes from the interviews were analyzed to find common themes. Some of the identified key benefits of currently used tests include: *simplicity*, as tasks of quite a simple nature, like Kohs Block Test, can still provide the person performing the assessment with a lot of information by observing how the patient performs and solves the task. *Familiarity* can also be a benefit, since some tests include aspects that are already familiar to the patient, such as blocks played with during childhood, which can make them easier to facilitate. Furthermore, *adaptability* was considered a strong point. Some tasks allow for starting at a medium level of difficulty, and after observing what the patient does and how they perform, and can be adapted quite easily to be either more difficult, or less difficult, depending on the patients' performance.

Furthermore, expectations of how VR could be used were discussed, as well as potential drawbacks and benefits of using VR. One possible benefit of VR was suggested to be that it can be created to be more 'neutral' in comparison to other tests, which can be affected by native language or cultural background. Furthermore, that VR could provide easier interactions, after a learning period, since touchscreens, such as those on smartphones, often can be too small to interact with, with precision, for some patients. Some possible issues of utilizing VR were also discussed, such as the risk of hindering cognitive activities. Humans have a tendency to some degree to off-load cognitive activities by using additional resources, such as counting on your fingers, and VR could hinder this by, for example, placing controllers in the patients hand. Furthermore, concerns regarding familiarity with VR were raised. People are often familiar with blocks as used in Kohs Block Test, however VR may be more difficult to learn to interact with if one is not familiar with it, which could cause issues.

6.5 Design process

The stakeholders from Minnesmottagningen and Centrum för Digital Hälsa have been involved in the entire design process, providing input and feedback throughout the process. Furthermore, design decisions were informed by literature and meetings with the stakeholders. The process has by no means been linear, but rather dependent on continuous stakeholder feedback and re-design throughout.

6.5.1 Ideation

A large part of the ideation phase took place during the literature study, gathering insights from previous studies. The stakeholder meetings moved towards becoming more informal ideation sessions aimed at gathering requirements and reasoning around which types of tasks could be beneficial, which types of pitfalls to avoid, and a suitable platform for developing the prototype. Within these meetings, the idea emerged to keep the virtual environment in a familiar situation, in order to provide a specific context which users would be familiar with and can make associations, utilize contextual cues and previously learned behaviours. Several different contexts were suggested, such as a wardrobe, baking a cake or making a cup of coffee. However, it was finally decided to create a grocery store setting. Not only is it a setting most people are familiar with, but there are examples of other studies, for example Mrakic-Sposta et al. (2018), and Yeh et al. (2012), which have utilized a grocery or convenience store as a virtual environment.

The requirements for the design were gathered from the stakeholder meetings as well as literature, and condensed to encompass the most important factors. The goals were to create a relatively fast and simple test, reducing discomfort, attempting to reduce reliance on language, as well as create a simple overview of the results of the test. From the stakeholder meetings, a number of desired qualities emerged: Keeping the task simple, being able to scale it up or down, and adapt the task, depending on patient factors, providing adequate feedback to the user, keeping the interaction simple. Furthermore, that creating a combination of a memory task and a navigation task, with some distracting elements, could be a good route to take.

Based on the literature and the stakeholder meetings, a first design was created, in sketch form, and reviewed by stakeholders in order to uncover further needs and to gain further input on issues such as how one should move through the environment and how much guidance of navigation should be provided. Also, the question of whether the user should be able to see all the items gathered, in a shopping basket.

After discussing the sketch, it was decided to create a digital prototype of a grocery store task, in which the user is asked to memorize a shopping list, and the order of the items on the list, before navigating the store in order to gather all of the items. Creating a navigation tasked appeared to be the most appropriate choice, based on the literature, since VR has the advantage of allowing the user to navigate, without actually navigating a physical environment. Keeping the cartoon-ish feeling of the sketch in the digital version of the prototype was deemed a good choice, in order to obtain some sense of playfulness, rather than attempting to create a more photorealistic environment.

6.5.2 Prototyping

Based on the sketch, the low fidelity prototype began to take shape in A-Frame. A-Frame is an open-source web framework based on top of HTML for creating WebVR, and is available on multiple platforms - web browsers, Cardboard, HTC Vive,

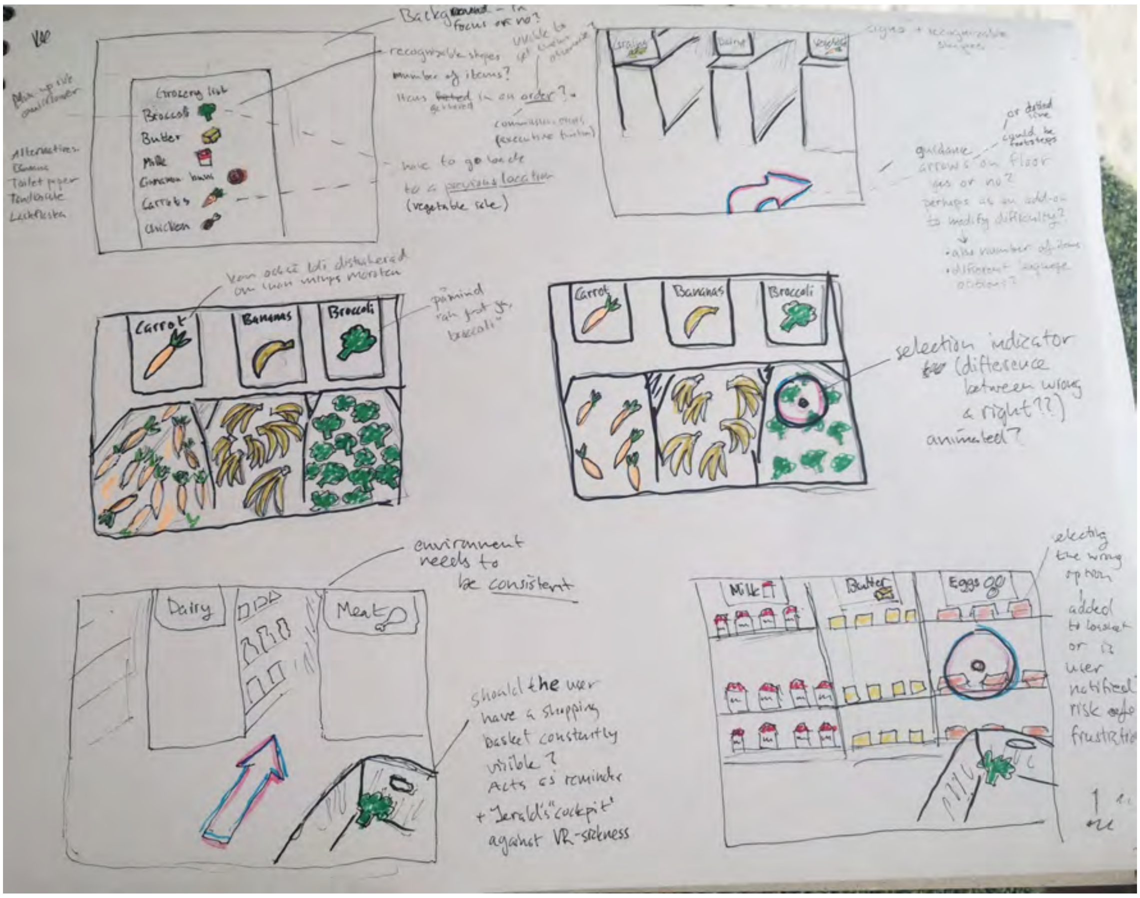

Figure 6.2: The first iteration of the prototype, in sketch form

Oculus Rift, Gear VR, amongst others, and can be interacted with using controllers and headsets, or simply on desktop or mobile (aframe.io, retrieved 2020-05-08).

A-Frame was selected because it provides a relatively simple platform to create WebVR. In order to facilitate evaluation, the prototype was developed to be used with a mobile VR HMD, such as Google Cardboard, or on a smartphone without a HMD only utilizing the sensors. This made the prototype more portable, and more easily accessible for evaluation. Furthermore, keeping the prototype web-based allowed for continuous evaluation by stakeholders, who could view the prototype online.

Building the prototype consisted of creating the virtual environment, based on simple geometrical shapes. The floor of the grocery store consists of a 30 x 30 meter plane, planes as outer walls, and with rectangles acting as shelves and inner walls. Interactivity was added via a cursor, which is consistently placed in the middle of the screen, and 'clicks' on items after hovering over them for a while. Animations, a built-in component of A-Frame, were added to certain items, in order to provide some feedback to the user, and to show which items can be interacted with by, for example, changing colors when hovered on. Simple images were created to act as textures on the interiors of the grocery store, for example the various shelves and the grocery items the users are meant to locate.

Development of the prototype lasted approximately four to five weeks. The prototype has been continuously evaluated throughout based on feedback from stakeholders, and adapted accordingly. One early decision was to change the language of the prototype from English to Swedish, in order to facilitate evaluation.

6.5.3 Evaluation

The goal of the usability study was to uncover potential issues with the prototype with regards to user experience and usability. Since the prototype is at an early stage, involving participants with cognitive impairment was deemed inappropriate. At this early stage, using healthy participants to identify usability issues was considered the most viable option in order to shed light on the most critical issues. The original planning included a Think-aloud protocol to evaluate the prototype in person, with experts, in order to immediately be able to detect issues regarding usability. Due to unforeseen circumstances, the evaluation had to be performed completely online. However, there was no shortage of expert input, due to the continuous feedback during the development of the prototype.

Using WebVR on mobile devices allowed for the prototype to be evaluated online, with or without a mobile VR HMD. A small group of 11 participants, consisting of healthy adults ranging between 20 to 59 years old were recruited for the non-VR condition. Two of the 11 participants later dropped out when attempting to evaluate the prototype, one due to technical issues, the other out of frustration from not being able to locate the milk. In addition, four participants, in which two stakeholders are included, evaluated the prototype with a HMD, in an attempt to detect issues specifically related to VR. These participants evaluated the prototype on their own, before providing written feedback.

The non-VR participants were instructed to use their smartphones, in order to take advantage of the sensors in the phone, to make the experience slightly more authentic. This allowed them to 'look around' in the virtual environment, by turning their phones or bodies, the same way as in VR. Furthermore, the interactions for the non-VR condition were the same as if they had had a HMD on, they moved around and interacted with the items in the same way as in the VR condition. This means that the main difference between the non-VR and the VR conditions were the level of immersion.

The non-VR participants first tried the prototype, by following the instructions provided within the prototype and completing the task of memorizing and navigating around the grocery store. This was followed by a questionnaire, consisting of an adapted version of the User Experience Questionnaire (UEQ), followed by an additional short questionnaire asking questions more specific to the prototype itself.

6.5.3.1 User experience questionnaire

The User Experience Questionnaire (UEQ) is a tool developed to measure user experience. In its original form, it is based on six scales with 26 items. The scales consist of Attractiveness, Perspicuity, Efficiency, Dependability, Stimulation and Novelty, and thus measures both pragmatic qualities (Perspicuity, Efficiency and Dependability), as well as hedonic qualities (Stimulation and Novelty). Attractiveness is considered a pure valence, the immediate overall impression (Schrepp, 2019)

The questions are based on a seven item scale where the product is rated between two opposing attributes - one positive and one negative. For example, one item can consist of a scale between attractive and unattractive, where the most positive rating is +3 and thus closest to attractive, and the most negative rating would be -3, unattractive. The middle is a neutral zero, and would make the rating equally attractive and unattractive. The UEQ is intended to be performed rather quickly, with the participant basing their judgment on their first impression. (ibid).

The UEQ was selected because it provides a foundation to quickly evaluate both hedonic and pragmatic qualities of the experience of the prototype and get a good grasp of how the prototype was perceived. Furthermore, it is quite a simple evaluation task, which was deemed appropriate to be used for online evaluation.

The UEQ used in this study was adapted to be even shorter and faster than the original 26 item questionnaire. This was done partially in order to not deter participants with too many questions, since the UEQ was combined with additional questions more specific to the prototype. According to the UEQ Handbook (Schrepp, 2019) entire scales can be removed from the UEQ if they are deemed irrelevant, but not single items. As such, two scales were removed - Dependability, and Novelty. The Dependability scale was removed because several of the items were deemed to be confusing or irrelevant (for example, rating the prototype as ‘secure’ or ‘not secure’ can easily be misinterpreted as being in regards to data security, rather than that the user feels secure and in control of the interaction). The Novelty scale was removed because many of the items were considered irrelevant in this particular scenario.

As such, the remaining four scales were: Attractiveness, Efficiency, Perspicuity and Stimulation, making up a total of 18 items.

The Stimulation scale contained the following four items: inferior/valuable, boring/exciting, not interesting/interesting, and motivating/demotivating.

The Efficiency scale contained: slow/fast, inefficient/efficient, impractical/practical, cluttered/organized.

The Perspicuity scaled contained: not understandable/understandable, difficult to learn/easy to learn, complicated/easy, and confusing/clear.

The Attractiveness scale contained six items: annoying/enjoyable, bad/good, unlikeable/pleasing, unpleasant/pleasant, unattractive/attractive, unfriendly/friendly.

These items were deemed to make up a good combination of pragmatic and hedonic attributes, as well as the pure valence of the Attractiveness scale which can provide insight into overall impression. These sections of the UEQ appear to cover all the core issues that were of interest in the evaluation - however, the Perspicuity section may be the most important, since it concerns ease of learning, whether the prototype is understandable, how complicated the prototype is, and whether it is more clear or confusing, all of which are key issues at this stage. The hedonic attributes and the Attractiveness scale are also important, as one wants to avoid creating a negative experience that makes people feel annoyance or unpleasantness.

The Swedish translation of the UEQ was used, since the prototype was made in Swedish, evaluation had to be performed by Swedish-speakers.

6.5.3.2 Additional questionnaire

In order to investigate more specific aspects of the prototype, an additional questionnaire followed the UEQ. This additional questionnaire consisted of eight questions. Seven of these questions were based on statements, where participants were to answer on a 1 to 7 Likert scale. 1 represents "do not agree" and 7 represents "agrees completely". The eight question consisted of checkboxes where participants were asked to cross off the items on the shopping list they found particularly difficult to find. Finally, the participants were given the option to add any additional comments.

The seven Likert scale questions intended to investigate how easy the prototype was to learn, attitude toward the prototype, how the interaction worked, whether the instructions within the prototype were adequate, and if the task was at an adequate difficulty level. The statements were as follows: *1. The instructions within the prototype were clear*, *2. I knew exactly what to do after reading the instructions*, *3. Learning how the prototype worked was fast*, *4. It was difficult to learn how to move through the virtual environment*, *5. Learning how to navigate in the virtual environment was fast*, *6. It was easy to find all items on the shopping list*, and finally *7. It was easy to remember the order of the items on the shopping list.*

The multiple choice question asks participants to recall which items were difficult to find - this was done in order to discover how many items were reasonable to find, and to uncover if any particular item was more difficult to find within the grocery store. The answer options lists all items that were on the shopping list, an option that states "all items were easy to find", and and option to add their own answer.

7 Results

7.1 Results from the prototype

7.1.1 The prototype

The grocery store was mapped out on in order to determine the layout of the store. The layout of the store was intended to place several of the items on the shopping list slightly hidden and far apart, in order to make the prototype slightly more challenging. The store contains several sections and shelves, only some which needs to be interacted with or visited in order to obtain the items on the list. Viewed from above, the grocery store resembles a very simple labyrinth, but from a first-person perspective, where the camera is placed at approximate average eye-level (1.6 meters from the ground), the items are more obscured by the shelves in the store.

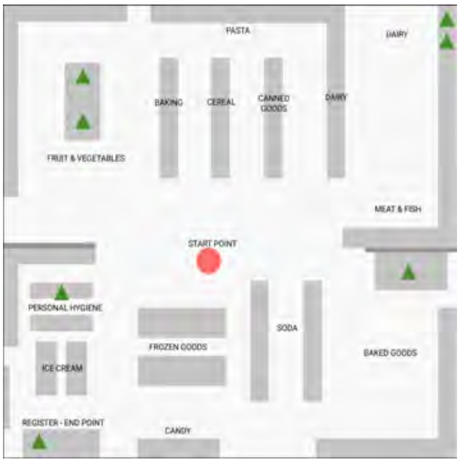

Figure 7.1: The layout of the grocery store, with all the locations the user is intended to visit marked with triangles - The fruit and vegetable section to find carrots and broccoli, the dairy section with milk and butter, the toothbrush in the personal hygiene section, the cinnamon bun in the baked goods section, and finally, the register, where the user ends the test.

The task begins with the user being presented with a list of six grocery store items to memorize, both in terms of the items themselves and in the order in which they appear. The user can take as much time as they need to attempt to memorize the list. The items on the prototype shopping list were carrot, milk, butter, broccoli, cinnamon bun, and a toothbrush. The list is illustrated with cartoon-like images to further assist the user in memorizing the items, and to aid recognition both when memorizing the list, and when moving through the grocery store and finding the items. On this first page there are also instructions on how to move within and interact with the prototype, as well as instructions on how to end the task, along with illustrations. The task begins once the user feels ready, by hovering over the start button. The user interacts with the prototype with a cursor placed in the middle of the screen. When hovering over an item, the cursor animates to indicate that the item can be interacted with.

Figure 7.2: The starting page of the prototype, with the shopping list and instructions.

When entering the store, the user can 'teleport' short distances by hovering over footprints placed out on the floor. Keeping the teleportation distances relatively short intends to give the user some sense of walking through a store, while attempting to minimize the risk of VR Sickness resulting from linear locomotion or acceleration through the virtual environment (Jerald, 2016). Electing to use footprints as teleportation portals is intended to show the user where they can walk, without leading the way in any particular direction (as arrows would, for example) and making the test too easy. Keeping the teleportation distances short was also intended to lessen the risk of getting lost by teleporting too large distances at a time. The footprints have a slight animation, changing color when hovered over, in order to indicate that they can be interacted with.

When the user moves through the store to retrieve the items on the shopping list, a shopping basket follows along, visible in the right corner. This is intended to create a rest frame, an item that remains stable relative to the user as the user moves around the environment, which can mitigate the risk of VR Sickness (Jerald, 2016). Furthermore, the items the user locates and picks up 'flies' into the basket from the shelf and remains visible within the shopping basket, which can act as a reminder as to which items the user has located, and which they have left to find. Having the chosen item 'fly' into the shopping basket provides feedback that the user has successfully gathered an item. This feature was added after feedback from stakeholders, who noted that it was difficult to see whether one had been successful

Figure 7.3: The start position. When entering the grocery store, the user ends up close to the middle of the store.

in picking up an item or not.

The user moves through the store to find all items they can recall, in the correct order. Some of the items are placed rather close to each other - the carrots are on the same shelf as the broccoli, and the butter is placed right next to the milk. The reasoning is that this may help remind users of the items they need to find - if they remembered the milk, they may see the butter right next to it. The order, on the other hand, may become more difficult when seeing the items right next to each other.

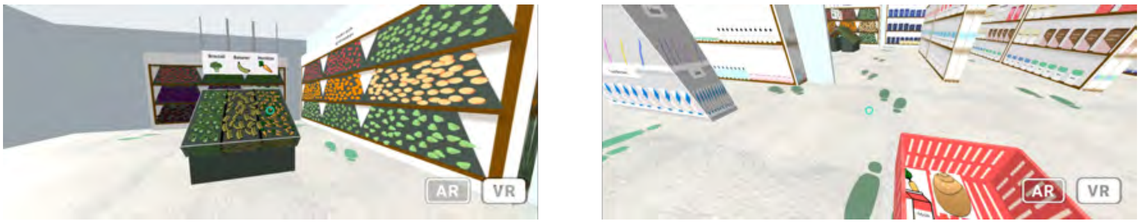

Figure 7.4: Screenshots depicting the fruit and vegetable section, and the shopping basket.

When finishing the VR task, the user travels towards the register, and hovers the cursor over a button that says 'pay'. This was done to emulate the experience of going to the grocery store - usually, a customer walks around and finds all their items, before paying and leaving. Once the pay-button has been clicked, the user gets a list of all items gathered, in the order they were gathered, with time-stamps for each item, and a total time of the whole test.
Having to gather all the items in a certain order adds another component of memory assessment - not only memorizing a list, but also memorizing the order of the items on the list. Gathering the items requires the user to navigate through the store. The carrot and broccoli on the shopping list requires the user to first find the carrot, then navigate to other sections of the store to find other items, before

Figure 7.5: The register and the pay-button that ends the test.

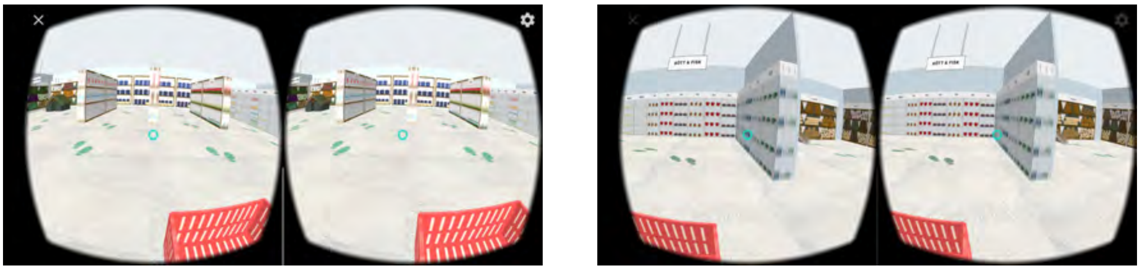

Figure 7.6: The grocery store viewed in VR mode.

finding their way back to the fruit and vegetable-section of the store to pick up the broccoli. This requires the users to remember the location of the broccoli, and be able to find their way back to the previously visited location. In addition, upon first visiting the fruit and vegetable section to find a carrot, the user may see the broccoli and be reminded that broccoli is also on the shopping list. This could either aid the user, or make them pick up the broccoli in the wrong order. Including a sequential memorizing task was inspired by Allain et al. (2014) and Bellassen et al. (2012), who utilized a similar approach where participants were asked to memorize a sequence of tasks, in the correct order. The sequence of which to gather items was also created in order to make the participants to first navigate to one location (the fruit and vegetable section), and then re-visit it the same location at a later stage. By having to navigate to the same location twice, they have to recall that location and the items located there.

The measured data was also inspired by Allain et al. (2014), by timing each item collected, the total time, the total number of items and the order in which the items were gathered. Timing each item is intended to illustrate which items were difficult to find, while the total time could indicate difficulty of the total experience. The total number of items gathered shows which and how many items were remembered, and the order in which they were gathered is intended to keep track of the ability to remember the sequential order of the list.

Designing the interior of the grocery store and the items that were to be located demanded some consideration. Mrakic-Sposta et al. (2018) designed a grocery shopping-based task, and found that the design of the packaging of the items caused some issues for the participants - they did not look like the items the participants usually bought. In order to mitigate similar issues, the products in the prototype were attempted to look fairly neutral, without brand-names. The interior of the grocery store attempted to have a balance between a number of 'decoy' shelves acting as distractions, and maintaining a fairy realistic grocery store experience without it being too cluttered and having an overwhelming amount of distracting elements. The VR environment is intended to look friendly and slightly playful in order to feel safe and calm.

The items on the grocery list were selected because they have a clear visual profile, and are designed to be recognizable and distinct. Attempting to keep items distinct, and supplementing with text and additional cues, such as animations of the buttons and footprints, could also aid those with colorblindness. The text and images in the prototype are made to be large and easily visible to account for older adults, which was an issue detected during the continuous feedback sessions.

7.2 Evaluation of the prototype

7.2.1 VR condition

The four participants evaluating the VR condition provided written feedback on the prototype, indicating a mixed result. Generally, they reported that the experience was enjoyable. However, most of the feedback focused on the issues that were detected. One comment was that the shopping list was quite long and difficult to remember. Another note was that the milk was difficult to find, because of the categorization. However, some VR-specific issues were detected. One noted that the shopping basket was positioned too far down which resulted in discomfort. The positioning of the basket caused them to bend the neck and make awkward movements with their head, which was found to be uncomfortable, and it was feared that it may exacerbate the risk of VR Sickness. Another recurring issue was the footprints used to move through the store. Some were too far apart, which caused issues when attempting to navigate an area with too few footprints. Another comment was that the footprints were too small to hit with the cursor from a certain distance. The biggest issue regarding the footprints, however, was that it was difficult to keep the cursor stable while using a HMD. This created issues with hovering over the footprints for long enough to produce a 'click', thus making teleportation difficult. One comment was also that the grocery store was too big, and another that there were a lot of impressions within the store, it may feel cluttered.

Another comment was that one participant could clearly see pixels when using a smartphone with a 1080p screen. The participant felt that it affected the aesthetic aspects of the prototype negatively, and made the experience feel less immersive.

7.2.2 Non-VR condition

7.2.2.1 Questionnaire

The results from the questionnaire were separated into results from the UEQ, and results from the additional questionnaire, in order to provide an overview of the user experiences based on the UEQ, and over the additional questions more specific to the prototype. 9 participants, aged between 20 to 59, completed the questionnaire.

7.2.2.2 User experience questionnaire - UEQ

The UEQ Data Analysis tool was used to extract the user experience data for the different items. Looking at the specific items can provide insight as to which aspects could be improved.

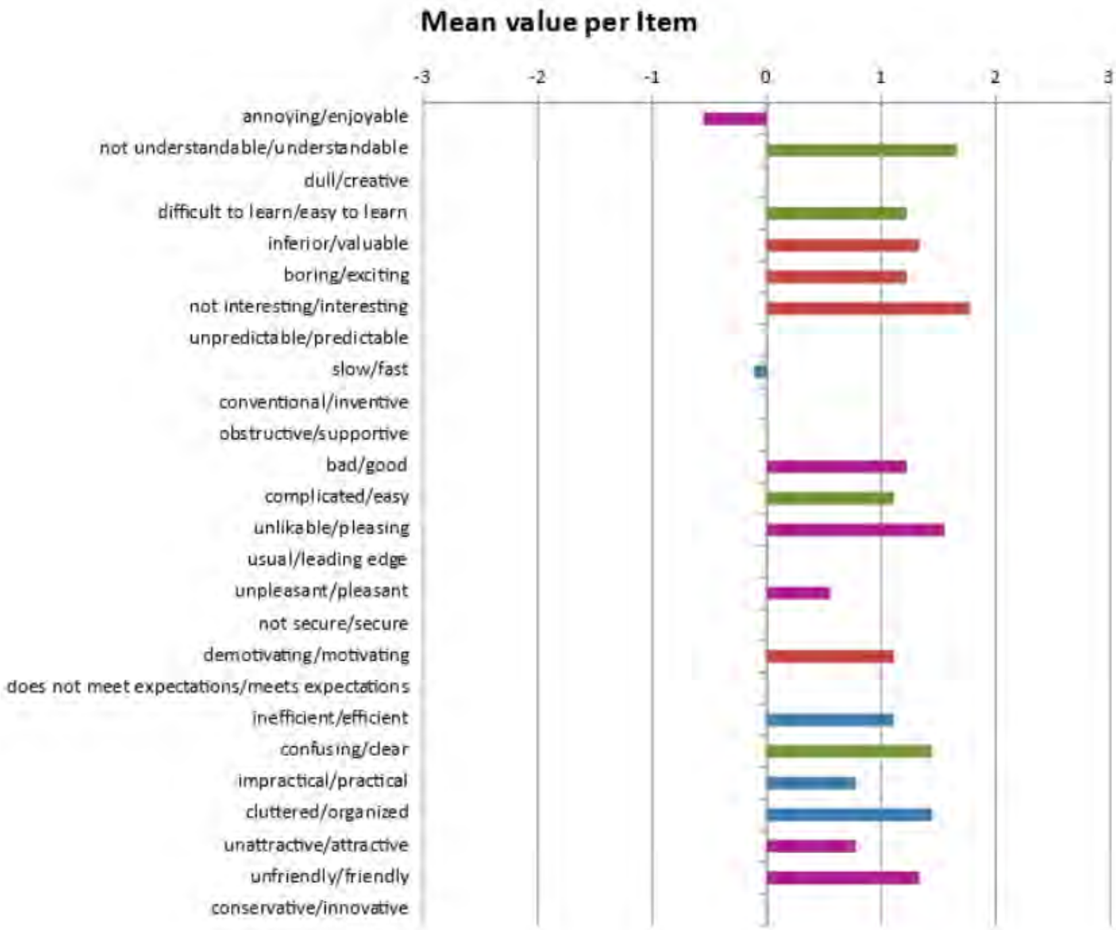

Figure 7.7: Bar graph depicting the mean value per item of the UEQ. The colors of the bars represent which scale each item belongs to. The purple color is the Attractiveness scale measuring pure valence, the green color is the Perspicuity scale, red is the Stimulation scale, and blue is Efficiency. Blank items are items belonging to the two removed sections of the UEQ.

The graph (Figure 7.7) shows a generally positive user experience based on the mean value per item. The rating ranges from -3 to 3, where -3 would be an exceptionally bad rating, and +3 would be exceptionally good. A UEQ score between -0,8 and 0,8 is generally considered a neutral score, while a score exceeding 0,8 is interpreted as a positive score, and below -0,8 is interpreted as negative. The lowest scoring item, annoying/enjoyable, has a mean score of -0,6, which based on UEQ guidelines would be considered a neutral score. The second lowest mean score is the item slow/fast, with a mean of -0,1. The highest mean score is the not interesting/interesting item with a mean score of 1,8, followed by the not understandable/understandable item with a mean score of 1,7. The blank items are the items from the scales that were excluded from the UEQ - Dependability and Novelty.

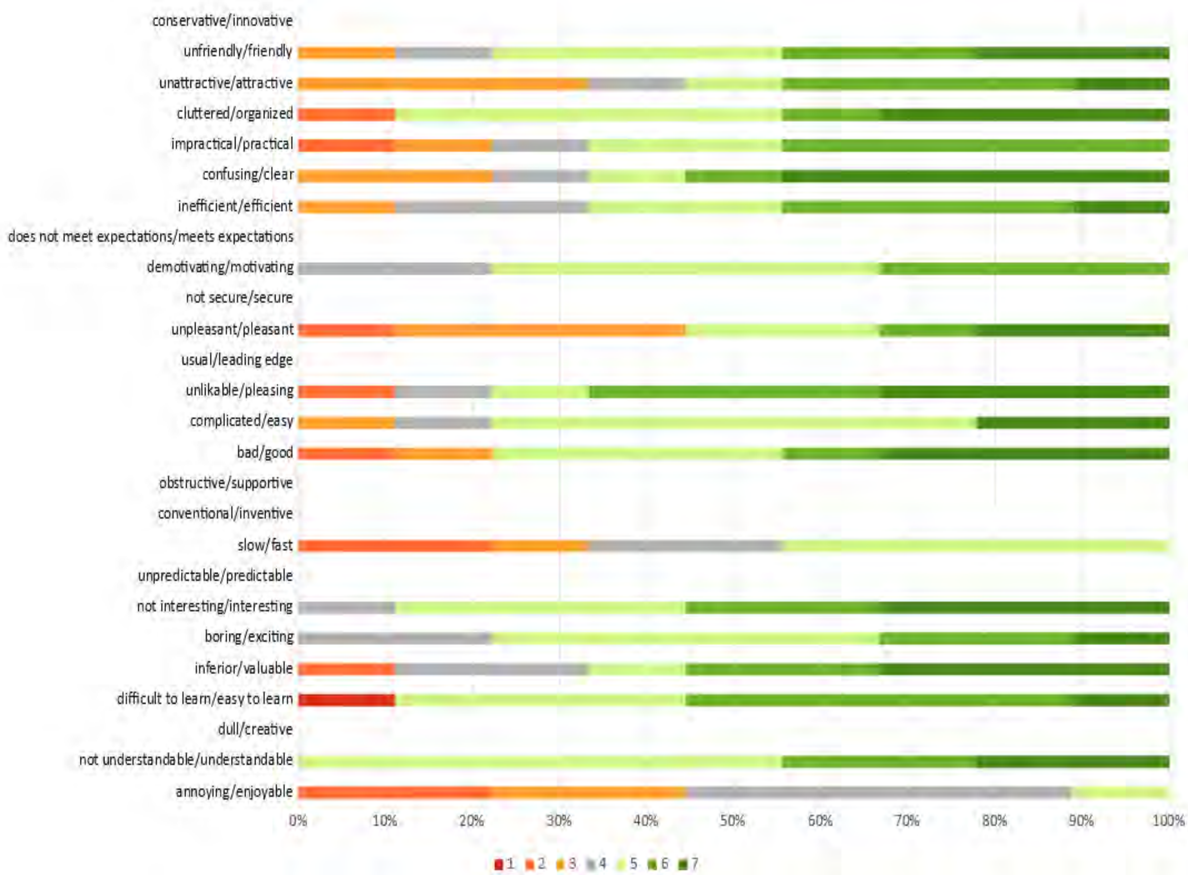

Figure 7.8: Answer distribution for all items. Blank items are items belonging to the two removed sections of the UEQ. Red and orange shades indicate lower values, green shades indicate higher values. Gray represents neutral

The distribution of answers (Figure 7.8) illustrate more specifically how the prototype was rated for each item. The dark red color represent the lowest score, the dark orange represents the second lowest score,and the lighter orange represents the third lowest. Gray is a four on the 7-point scale, meaning neutral, or equally good and bad.

As Figure 7.9 shows, the error bars indicate quite a large confidence interval per scale, indicating a low precision of the UEQ. The Perspicuity and the Stimulation

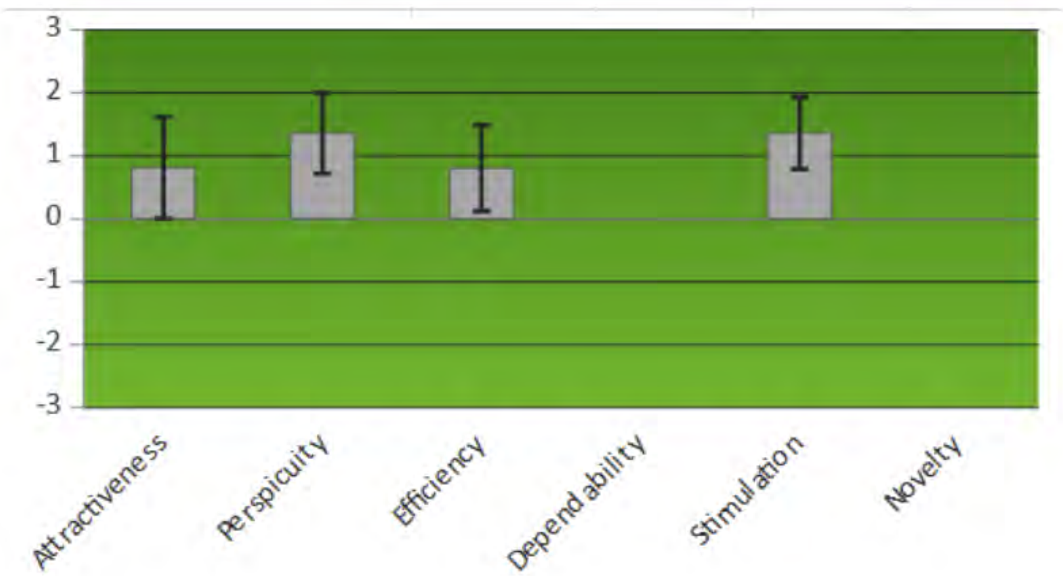

Figure 7.9: Error bars illustrating the confidence interval, per scale of the UEQ. The Dependability and Novelty scales were excluded from the questionnaire.

scales both received a mean of 1,361, while the Attractiveness scale received a mean of 0,815, and the Efficiency scale had a mean of 0,806.

7.2.2.3 Additional questionnaire

The supplementary questionnaire focused on more specific aspects of the prototype, in order to investigate aspects of learnability and interaction. The questions were based on statements which were to be rated on a 1 to 7 Likert scale, where 1 was 'do not agree' and 7 was 'agrees completely'. The first statement was *1. The instructions within the prototype were clear.* The majority (66,7%, or six participants) of participants answered 7, 'agrees completely'. Two participants (22,2%) answered a 6, and one participant (11,1%) answered with a 5. In other words, a majority of the participants responded positively to the clarity of the instructions. The second statement *2. I knew exactly what to do after reading the instructions* acted as a control question to further investigate the instructions, and detect answer discrepancies between the first and second statement. Seven participants filled in 'agrees completely', one participant answered with a 6, and, interestingly enough, one participant answered with a 3, weighing more towards 'do not agree'.

The third statement, *3. Learning how the prototype worked was fast*, fourth statement *4. It was difficult to learn how to move through the virtual environment*, and fifth statement *5. Learning how to navigate in the virtual environment was fast*, aimed to take a closer look at the learnability of the prototype as well as ease of interaction and navigation. The responses to the third statement appear to lean strongly towards 'agrees completely'. The fourth statement was formulated to specifically detect difficulties regarding moving through the virtual environment,

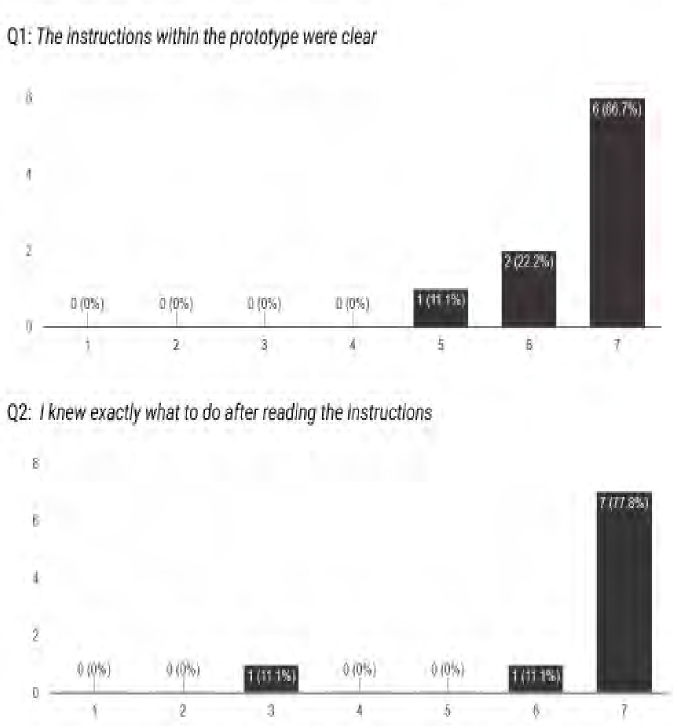

Figure 7.10: Answer distribution for question 1 and question 2.

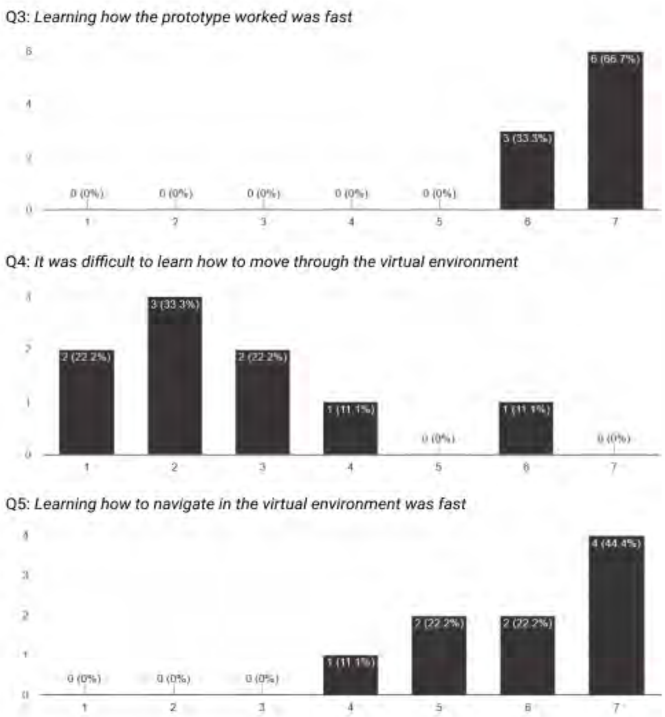

Figure 7.11: Answer distribution for question 3, question 4 and question 5.

and shows a larger answer distribution, however leaning towards 'do not agree'. The fifth statement also shows a wider distribution of answers, with four responses indicating 'agrees completely', two responses answering with a 6, and two responses answering with a 5. One response answered with a 4, indicating neutral. The sixth statement, *6. It was easy to find all items on the shopping list*, and the seventh statement *7. It was easy to remember the order of the items on the shopping list* aimed to investigate the task itself, whether locating all the items was experienced as easy or difficult, and if the ordering of the list was easy or difficult to remember. The responses to question 7 are quite distributed between alternatives.

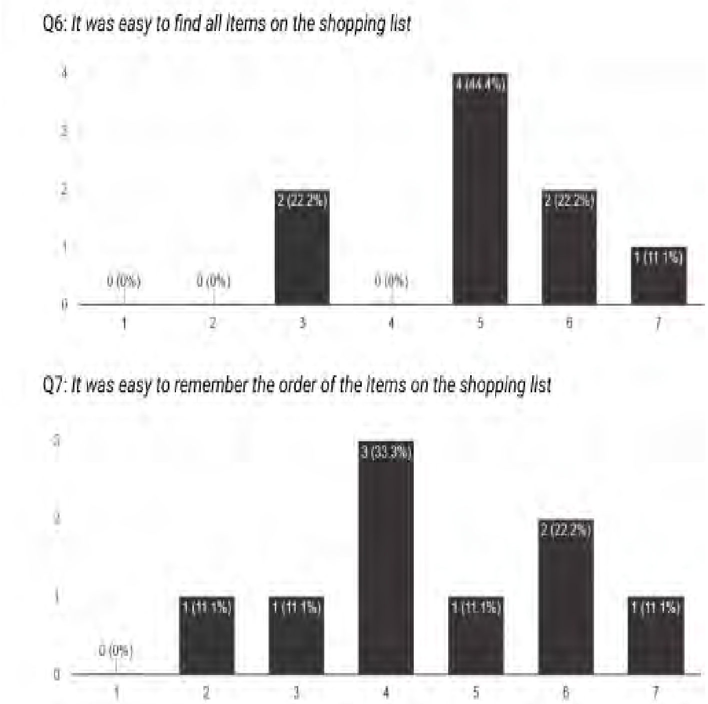

Figure 7.12: Answer distribution for question 6 and question 7.

Figure 7.13: Answer distribution for question 8.

The final question, question 8, *Were any of the items on the shopping list particularly difficult to find?* consisted of checkboxes where participants could indicate which, if any, item(s) from the shopping list were difficult to locate. The respondents were provided with six options for each of the items on the list, one option to indicate that all items were easy to locate, as well as the option to write their own response. The option 'all items were easy to find' received 4 responses, followed by milk, which received 3 responses, and butter, which received one. One custom response was added, saying 'fairly easy, but I forgot the milk'.

7.2.2.4 Additional comments

Some additional comment from participants also need to be included, in order to gain further insight into the experiences of the participants. For example, one participant wrote that they accidentally picked up the wrong item by looking around the grocery store, and felt that the risk of accidentally placing an item in the shopping basket stressful. Several mentioned issues with the teleportation and experienced that the footprints were difficult to hit with the cursor, and one participant mentioned having trouble with grabbing some of the items off the shelf. Two participants also found the dairy section confusing, and ended up on the wrong side of the shelf before realizing they had to walk around the shelf to find the milk and butter. Some comments on the questionnaire also concerned the number and positioning of the footprints which affected the ability to move through the grocery store. One comment mentioned wanting to be able to take bigger steps at the time to go through the prototype faster.

7.3 Factors to consider

The literature study, interviews and prototype lead to the uncovering of a number of factors that need to be considered when designing and using a VR task for cognitive assessment, in a clinical setting. This section will present these eight factors.

7.3.1 VR Sickness

The risk of VR Sickness can be reduced through design, by for example designing the task not to be performed in a standing position, allowing the user to be in control of navigation, and designing to avoid frequent head motions (section 6.3.5).

Reducing the risk of VR Sickness is imperative, as it may not only cause discomfort, but may also affect the performance of the assessment task. Taking careful measures to mitigate the risk of VR Sickness may be especially important when utilizing VR in a clinical setting to assess cognitive function. The potential users of such a system may often be older adults, or be in a poor state of health, who could be particularly susceptible to VR Sickness (section 6.2.4). This indicates that taking measures to reduce the risk of adverse effects, throughout the design process, may be especially important in a clinical setting.

The grocery store prototype utilized two main methods to counteract VR Sickness: teleportation rather than linear motion, and having the shopping basket act as a rest frame as the user moves through the environment. These two tactics were gathered from Jerald (2016) (section 7.1.1). The evaluation however, indicated that the shopping basket was unsuccessful and rather counter-productive, as it was placed in such a way that the user had to make head movements which could exacerbate the risk of VR Sickness (section 7.2.1) There were no mentions of VR Sickness or other adverse effects stemming from teleportation through the environment, in the VR condition evaluation.

7.3.2 Physical discomfort

The use of VR could also lead to physical discomfort, if not designed with great care. For example, physical fatigue or discomfort resulting from a prolonged positioning of the body while interacting in VR. This is also an important factor to consider when designing a system that may be used by elderly or disabled users (section 6.2.4), such as a system that is intended for use in a clinical setting. Much like the case of VR Sickness, physical discomfort stemming from use of VR could be reduced through careful design (section 6.3.5).

The grocery store prototype was designed not to require a lot of physical movement, besides the head and neck, as interaction was done via a cursor rather than hand-held controllers. However, in the VR condition evaluation, the shopping basket was found to cause some discomfort due to its' placement (section 7.2.1). Furthermore, the grocery store task could be performed while sitting down, but would still require some turning of the body to look around the environment. This could be mended further by, for example, placing the user in a swivel chair.

7.3.3 The method of interaction

Selecting which type of input method to be used with the VR system needs careful consideration. From the interviews, it was gathered that VR hopefully could provide easier interactions, compared to touchscreens which many patients have trouble interacting with (section 6.4. In order to for this to be the case, one needs to carefully consider the different options for user interaction, since there may be advantages and disadvantages of these options. For example, hand controllers could provide better control over the interaction, however may also cause physical discomfort, if care is not taken when designing the system (section 6.2.4). Furthermore, hand-held controllers could hinder cognitive abilities such as counting on your finger (section 6.4). Taking away the option to utilize these types of cognitive activities, that are normally available to the user, could possibly affect the results of the assessment. On the other hand, a cursor-based interaction may lessen the risk of physical discomfort, but, as the evaluation of the prototype showed, provide less control of the interaction (section 7.2.2.4). The evaluation of the grocery store prototype indicate that there were a number of issues with the cursor-based interaction. For example, the risk of accidentally interacting with the environment, and difficulties picking up items (section 7.2.2.4), as well as issues with keeping the cursor stable (section 7.2.1).

7.3.4 Movement through the environment

The method of movement through the virtual environment needs to be considered carefully when designing this type of system, for example not to cause issues with VR Sickness (section 6.2.4), but also to allow the user to efficiently move through the environment. One option are tasks which have the user walk around in the physical environment while immersed in the VR (section 3.4) which could provide a higher degree of immersion and lessen the risk of VR Sickness, but it is also associated with safety risks (section 6.2.3). Furthermore, it may be less efficient and less comfortable than keeping the movement purely in VR (section 3.4).

The grocery store prototype utilized teleportation, via portals placed on the floor in the virtual environment, to lessen the risk of VR Sickness, as well as to attempt to keep a sense of 'walking' through the grocery store (section 7.1.1). However, this resulted in a number of issues regarding movement through through the grocery store, detected in the evaluation, both in the VR condition and the non-VR condition (section 7.2.1 and section 7.2.2.4). This indicates that the design of the movement through a virtual environment needs to be well designed in order not to cause similar issues. There are numerous methods of movement through virtual environments that can be explored.

7.3.5 Adaptability

The interviews found the importance of being able to create a system that can be scaled up or down in terms of difficulty level, depending on how the user performs. This type of adaptability often exists in traditional methods of assessment, so when designing this type of task for VR, there may be a need to include some form of adaptability, to be able to provide a similar flexibility, to allow the evaluator to find what they are looking for when assessing a patients' cognitive ability (section 6.4).

The grocery store prototype was not adaptable in terms of difficulty, however it could relatively easily be re-designed to be adaptable. For example, one could create different versions of the task with more, or less, items on the shopping list, by creating a smaller or larger store, or having a store with more, or less, distracting stimuli (for example, the number of shelves or items that are not intended to be interacted with).

7.3.6 Familiarity with VR

The familiarity with VR is an important consideration both for the end-users (patients) and other users (personnel). For the patients, the level of familiarity may affect their susceptibility to VR Sickness as well as their ability to interact with the system, as well as how comfortable they feel while interacting (section 6.2.2). This indicates the need for a task that is performed before the real task to allow the user to familiarize themselves with VR, in order not the affect the results of the real task.

From the point of view of the personnel, they would need to learn a new system which may be entirely new to them, to be able to set up and modify the system, while ensuring the comfort of the patient (section 6.2.3). In other words, they would need the time to familiarize themselves with the system in order to feel able to use it efficiently without too much additional workload. Ideally this aspect should be considered early in the design process, by designing the system to be easy to set up and modify. The grocery store prototype was created to be used with a mobile VR headset and a smartphone, and accessed via a web page, which may have made the prototype more accessible (section 6.5.2).

7.3.7 Language, culture and educational background

Part of the aim of the prototype was to be able to create a more 'neutral' task in terms of language and culture, in comparison to traditional assessment tasks which may be affected by the patients' native language or cultural background (section 6.4, such as the MMSE and the Clock Drawing Test (section 3.2, and many traditional tasks are developed for a Western population (section 3.4.1). Furthermore, the proposed higher ecological validity of VR could make it a beneficial tool for assessment of people with a lower educational background (section 6.3.1). Moreover, it could be possible to create purely non-verbal tasks in VR, or multiple versions of a task, to cater to more groups of people. In other words, it is important to attempt to design an assessment task that can be accessible to more people, in order to facilitate assessment of various groups of people.

The grocery store prototype attempts to lessen the reliance on language, however it does include some text, such as the shopping list and instructions to the task. However, images were also utilized, in an attempt to make the items easier recognizable (section 7.1.1).

7.3.8 Resemblance to real environment

VR tasks could resemble real situations more closely than pen and paper tasks or desktop-based tasks, and as such increase the ecological validity of the assessment tasks (section 6.3.9). In order to increase the ecological validity, it may be sensible to attempt to mimic familiar settings and contexts in VR. A higher degree of ecological validity could be beneficial (section 6.3.9) and could aid when assessing patients with a lower educational background (section 6.3.1). Furthermore, a familiar setting could aid the user in making associations, utilizing previously learned behaviours and contextual cues, to facilitate the interaction (section 6.5.1).

7.4 Knowledge contribution

The eight identified factors (section 7.3) aim to shed light on considerations related to the design and use of a VR system for cognitive assessment in a clinical setting. This can be considered the main knowledge contribution of the project, as these factors could provide future projects with considerations that should be kept in mind during the design process, and could help avoid certain pitfalls.

8

Discussion

8.1 Results

8.1.1 Literature study

The results from the literature study attempts to summarize the identified potential issues and advantages of using VR to screen for cognitive impairment, for example in cases of suspected dementia. The potential issues can help draw conclusions regarding feasibility, based on a more holistic perspective and not purely the accuracy of the screening methods themselves.

Potential issues that can arise when utilizing VR. For example, some studies (Cushman et al. (2008); Allain et al. (2014); Manera et al. (2016)) indicate that participants' performance was lower, for all participants, when comparing a virtual environment task with the same task performed in a real environment. There are numerous possible explanations as to why, for example, the participants may have been unfamiliar or uncomfortable with VR, that the VR tasks were more demanding of cognitive resources, or that the VR tasks were more difficult. However, this does not necessarily have to become an issue, as long as there is awareness of the possibility. For example, Cushman et al. (2008) notes that this lowered performance was consistent across all participant groups, both young and older healthy control groups, as well as the MCI and AD groups. Keeping the possibility in mind that VR could result in a lowered performance is nevertheless important.

Issues not related to the VR assessment tasks in themselves, but rather related issues that need to be considered when creating a VR assessment task. For example, the risk of adverse effects of VR, such as VR Sickness or physical discomfort is related to the creation and application of VR assessment tasks, but not necessarily related to their ability of assessing cognitive impairment. The risk of making people experience discomfort is an important issue to keep in mind, not only to avoid affecting the results of the assessment, but to minimize patient discomfort.

Unfamiliarity with VR. This aspect can affect both the patients being assessed, and the medical personnel performing the assessment. From the patients point of view, it may be an issue of experiencing discomfort, and not being able to know what looks correct due to inexperience (which in turn can exacerbate discomfort). Not being familiar with the methods of interaction could also create insecurities. This means that care needs to be taken to produce a practice task, allowing the patient to familiarize themselves with VR before the real assessment task. Having an onboarding task could aid in making the patient feel confident and capable before the assessment begins. The medical personnel would also need to be familiar with VR in order to be able to use the tool for assessment. They would also need to know what looks correct in order to aid the patient, and be able to guide the patient with placement of the HMD and how the screen is supposed to look. Furthermore, depending on the type of VR system used (mobile VR or location-based VR), and the design of the system, setting up the system may need some practice. Furthermore, they would need to be aware of the various risks of discomfort as well as the hygiene risks of utilizing VR. The storing of patient data is also a possible issue which needs a robust solution in order not the violate confidentiality and the GDPR.

All of these identified issues also create a separate issue - the development process of VR to assess cognitive impairment. The process of developing a neuropsychological assessment task or cognitive screening test is, in itself, a rigorous process. Utilizing VR adds on to the requirements of issues that needs to be considered when designing such a task, such as the risk of VR Sickness or physical discomfort and the risk of a lowered performance compared to real environment tasks. These aspects are typically not as prevalent in pen and paper or desktop tasks, while utilizing VR adds an additional layer of considerations that need to be carefully investigated when designing the task. This could make the development process quite time consuming, and whether this investment of time and resources is sensible, needs to be weighed against the possible advantages of using VR.

For instance, one of the main identified advantages of VR is the ability to assess aspects of cognition that typically arise early in AD and MCI, such as loss of navigational ability and disorientation, as well as more subtle deficits. For example, Howett et al. (2019) found that their entorhinal-cortex based VR navigation task was able to differentiate prodromal AD with superior accuracy compared to commonly used cognitive tests of episodic memory, attention, and processing speed. Navigation and orientation tasks appear to be a particularly promising application of VR. VR could provide a higher ecological validity of navigation tasks, compared to pen and paper based tasks, while being performed more conveniently without having to navigate a physical environment. Furthermore, VR could also allow for a higher level of control over the environment and the stimuli presented. Recording data may also be an advantage in VR, by being able to measure more accurately without having to rely on manual note-taking and timing, and obtaining other measurements that are difficult to observe manually. The ability to adapt the environment does not only concern the stimuli presented, but there are also possibilities to make adaptions based on language, however few studies have actually investigated this.

Patient comfort is another possible advantage of utilizing VR. Certain procedures are uncomfortable, such as the lumbar puncture, and can be stressful or time consuming. Especially real environment navigation tasks may cause discomfort for elderly or disabled patients, if they have to walk or move for a longer period of time. In VR, navigation tasks can be performed without actually having to move through a physical environment. Furthermore, it was noted that computerized tasks utilizing mouse and keyboard, or touchscreens, may be too demanding of fine motor skills, which can be affected by aging. VR can utilize other input methods such as hand-controllers or cursors placed in the middle of the users' field of view, and make the interface larger than typically seen on touchscreens or computer monitors. This can facilitate interaction as well as visibility. However, this absolutely requires taking proper precautions and careful design measures to mitigate the risk for ergonomic issues and VR Sickness, otherwise it may be counter-productive and quite uncomfortable. Another possible benefit of VR is the possibility to make the experience enjoyable and interesting, which also could aid in increasing patient comfort. For example, Manera et al. (2016) found that participants with MCI and dementia preferred a VR task over a pen and paper task, and Mrakic-Sposta et al. (2018) evaluated VR tasks for cognitive training for participant with MCI, where the VR tasks were reported to have high levels of engagement and motivation, as well as enjoyment. However, there is a risk of the high reported levels of user acceptance being related to the novelty of using VR. The relative newness of VR may make it more enjoyable and exciting, compared to, for example, pen and paper tasks. On the other hand, Mrakic-Sposta et al. (2018) performed their study during a longer period of time, six weeks, where participants trained in VR three times a week, while still reporting a high user acceptance and enjoyment. In other words, it is possible that the effects of novelty may have worn off in this case. Another aspect to consider is how often a user would be exposed to VR in a clinical setting, since this may affect the perceived novelty of the experience, and as such, the user experience.

In other words, there are numerous considerations one needs to make when utilizing VR to assess cognitive function, however VR also shows some promise, especially in cognitive tasks related to navigation and orientation.

8.1.2 The prototype

The aim of the prototype was explorative in nature, by attempting to design a VR assessment, based on existing studies as well as interviews with stakeholders. Furthermore, to uncover possible design considerations that need to be kept in mind, especially regarding usability and user experience.

The prototype received mixed feedback in the evaluation, and numerous potential issues were discovered. The prototype, being an early-stage prototype rather than a close-to-finished product, means that this was expected. Finding these issues at an early stage may be beneficial to be able to adapt and improve.

Having such a small sample size in the UEQ means that the precision of the UEQ results may be low and may not produce accurate measurements. However, at this prototype stage the results may still be enough to identify possible areas of improvement. For example, the fairly low mean score,-0,6, of the annoying/enjoyable scale could indicate that some aspects of the prototype have room for improvement. Similarly, the slow/fast item, which had a rating of -0,1 could indicate that there is a need for optimization of virtual environment. The Perspicuity scale and Stimulation scales had the highest mean values, 1,361. The Perspicuity scale intends to measure how easy the prototype was to get familiar with and learn, and a mean of 1,361 is a fairly positive result, as it is, according to the UEQ Handbook (Schrepp, 2019) rare to receive a mean exceeding +2. Making the prototype simple and easy to learn was quite an important goal, and the UEQ indicates that this was fairly successful. This is also supported in the additional questionnaire, where the questions regarding learnability received fairly high scores, as did the questions regarding the instruction page of the prototype. The Stimulation scale intended to measure whether the prototype was motivating and exciting, and judging from the UEQ, it appears to have succeeded fairly well. The Attractiveness scale only ended at a mean of 0,815, indicating that there is room for improvement regarding the overall impression. The Efficiency scale had the lowest mean, 0,806, attempting to measure if the prototype required a lot of effort to complete. When performing a cognitive test, some amount of effort may be required, however the effort should be directed to the task at hand rather than the tool used to perform the test with, and in this case the effort may be related to the tool itself rather than the task, based on the items in the scale, indicating that there may be aspects that need to be examined further. According to the UEQ Handbook (ibid), a score exceeding 0,8 can be interpreted as a positive score, however, both the Attractiveness and the Efficiency scale very barely made the cut.

The teleportation method was one of the biggest issues both in VR and in non-VR. Many experienced that the footprints were placed too far apart, or were too small to hit with the cursor, to efficiently move through the grocery store. The teleportation issue could also be related to the comment in the VR condition, that the grocery store was perceived to be too big, making it difficult to traverse the store efficiently. By having quite a large space to move through, but inadequate possibilities of motion likely contributes to a poor user experience. One possible reason for the difficulty of hitting the footprint is that the prints are placed as planes, almost parallel to the grocery store floor, giving them a relatively small clickable area, and difficult to reach from certain angles.The idea behind this was to hinder users from travelling too-large distances across the store, and to avoid the issue of hovering over a footprint by accident for too long and teleporting by accident. One possible solution could be to place invisible boxes on top of the footprints, which could provide a larger area to hit as well as making them easier to reach from a distance. This, however, increases the risk of accidental teleportation.

The participant who commented that they accidentally picked up an item simply by looking around within the store discovered an important problem with the cursor-based interaction, as well as the user experience since the participant mentioned feeling stressed by this. One possible option would be to instead utilize some form of hand controller, or making items more difficult to pick up by making the action of picking up an item a two-step process - first allowing the user to inspect the item, and then verify by putting it in the basket. Another feedback comment was wishing for the ability to put items back from the basket. This version of the prototype does not have this possibility, however it would likely be beneficial to include. Being able to put items back could also make the risk of accidentally putting an item in the basket less stressful.

Another recurring issue was the placement of the milk. Out of all the items on the list, the milk was the most difficult for participants to find. This was expected, since the milk was intentionally placed quite far away and hidden among other shelves. The milk was located in one of the corners, hidden behind other dairy shelves. Furthermore, on the dairy shelves there were other packages that looked too similar to the milk packaging. One participant even dropped out of the study out of frustration from not being able to find the milk, which may be a good indication that it was too well-hidden.

Another interesting result came from the VR condition, where issues regarding the placement of the shopping basket emerged. The idea behind the shopping basket was to act as a stable component to minimize the risk of VR Sickness, however the effect became the opposite when the basket was placed in such a way that it required the user to bend their neck in order to see the items within, instead causing head motions which have been tied to VR Sickness.

One of the hopes for VR was to be able to make the assessment less language-dependent. This iteration of the prototype does not fully reach that goal, since it contains a fair amount of text. For example, the instructions and shopping list consists of written text in Swedish, as well as the text on the button used to end the experience, and the signs within the store. The illustrations on the instruction page, having illustrations of the items on the shopping list, and some of the signs are however intended to lessen the reliance on text.

The issues detected during the evaluation of the prototype serve to illustrate the importance of carefully selecting and planning the method of movement through the virtual environment. Moving through the environment was found to be one of the most prevalent issues within the prototype, indicating that another method is required, or that the currently used teleportation-method needs to be improved. Another important lesson gathered from the prototype is to add constraints to the interaction, or otherwise lessen the risk of accidental interaction (such as picking up items just by looking around the environment). Another option is to include hand controllers to provide a better sense of control, and to provide a better user experience, since the risk of accidentally picking up items could make the experience not only stressful, but frustrating. Adding the ability to remove items could have a similar effect.

8.1.3 Factors to consider

The eight identified factors are the main knowledge contribution of this project, and could be considered when designing a VR task for assessment of cognitive function in a clinical setting. These factors were based on the experiences from this project, the design and evaluation of the prototype, the stakeholder interviews, and the literature study. This means that they may have some limitations, and in some cases may be quite specific for this particular context.

For example, the need to include adaptability in terms of difficulty level, could perhaps be quite specific for the context, and may be less relevant in other contexts. Adaptability was found to be important in the stakeholder interviews, from a neuropsychological perspective, in order to offer a similar flexibility as traditional assessment tasks. However, this factor is only based on the interviews, which could indicate that this factor may not be as general as some of the other identified factors. The resemblance to the real environment is also quite specific for the cognitive assessment context, as it could increase the ecological validity of the assessment. This factor may not be as relevant outside of the context of assessing cognitive function, however it could possibly also be applied to, for example, cognitive training tasks.

Creating a VR task that is less reliant on language, culture and educational background could appear to be quite specific for this particular context, as these factors could affect the results of cognitive assessment tasks. However, making a system accessible to as many people as possible may be beneficial, or even necessary, in many types of contexts, and as such this factor may be more general and applicable to many different systems. Keeping in mind the risk of VR Sickness, and taking steps to mitigate it, throughout the design process is also a more general factor that can be beneficial to consider when designing a VR system of any kind. The risk of VR Sickness does become especially important to consider when designing a system that may be used by people who may be more susceptible to VR Sickness, such as older adult or people in a poor state of health. The same reasoning applies to the risk of physical discomfort stemming from VR use. While these factors may be particularly important to consider when working with these user groups, it may also be sensible to keep them in mind when designing a VR system for other user groups, as not to cause unnecessary discomfort.

Carefully considering the method of movement through the environment, and the input method to interact with the environment, are also factors that can be beneficial to keep in mind when designing a VR system. There are many different interaction methods, and methods of movement, and selecting which to use requires careful consideration. The different methods may be more or less suitable depending on which type of system or VR experience is being designed. Comparing the benefits or drawbacks of each, based on the context as well as needs or abilities of different user groups, is a necessary consideration to make. This also ties in with the familiarity with VR factor. Considering the level of familiarity with VR among the users may help determine which type of movement and interaction could be appropriate, as some may be easier to grasp than others. The level of familiarity with VR also has more specific considerations for a clinical setting, as the personnel who would facilitate the use of a VR assessment task would need to be considered. The personnel needs to be able to set up and modify the task, and ensure the comfort of the patient, which would need to be kept in mind throughout the design process.

8.2 Process

The process has deviated slightly from the details of the initial plan, however the plan was followed but with some adaptations due to time constraints and outside factors. The literature study made up a large part of the project, as intended, and acted as a basis for the design. The design process has been heavily focused on the research and understanding, with heavy involvement of stakeholders from a different perspectives throughout. The stakeholders have continuously evaluated and provided feedback both regarding the design of the prototype and the literature study, and been key to informing the design decisions from the first sketch to final iteration of the prototype. The ideation consisted of literature as well as meetings where the prototype and various possibilities were discussed before the first sketch was created. This has also made it possible to detect possible issues in the prototype early on, and continuously throughout the development. For example, adding more animations within the prototype in order to provide stronger feedback, enlarging text and images, ideas on how to end the experience, and electing to use the teleportation method to transport the user in the environment to avoid ending up upside down. Electing to use A-Frame to create the prototype seems to have been a sensible choice, since it allowed to create an interactive prototype, provided a lot of flexibility for the design, and ensured that the prototype could be evaluated online.

The evaluation of the prototype had to be conducted online, which was not the original plan, and relied on the users testing and providing written feedback rather than a Concurrent Think-Aloud. It is possible that additional aspects could have been discovered if utilizing the Think-Aloud protocol and performing the evaluation exclusively in VR rather than a mix of both VR and non-VR. Having a non-VR condition was due to the fact that evaluation had to be performed online, and recruiting participants would become more difficult if they needed to have access to a mobile VR headset. The non-VR condition did not, however, differ greatly from the VR condition, aside from the level of immersion. The VR condition did, however, find some issues specific to VR, and having more participants evaluating in VR could potentially have found additional issues with the prototype.

Utilizing the UEQ and the additional questionnaire could still extract some interesting results, however a more qualitative approach could likely have yielded more, since the many of the more interesting issues were raised in the 'additional comments'-section. Furthermore, having so few respondents to the questionnaire mean that it is difficult to obtain a statistically reliable result on the UEQ, however the results can be viewed as indicators of possible issues. Another possible issue is that the participants' level of familiarity with VR was not investigated. It could have been beneficial to know whether they were familiar with VR, or first time users, as it could have affected their answers. Having users from both groups, familiar and unfamiliar with VR, could provide different types of perspectives into issues with the prototype. First time users could provide a better insight into how easy the prototype was to learn to interact with, while more experienced VR users could have previous experiences to compare with.

8.3 Future work

If applying this type of assessment in a clinical setting for first-time users, there would need to be an onboarding task to acquaint the user with the system before launching into the real assessment task, in order not to let any possible unfamiliarity with VR affect the results. This onboarding task could, for example, consist of a miniature version of the real task, or it could be a start-up task related to the main task. One example based on the grocery store task could be starting in the entryway of a virtual home, picking up some items needed to go to the grocery store - keys, wallet, and the shopping list - before walking out the door to practice moving through the virtual environment. This could also provide a first opportunity to attempt to memorize the shopping list, adding a learning-component to the assessment, which may be needed to make the assessment task more accurate. Another option could be to have a completely unrelated task where the user gets to learn how to interact with a virtual environment, by moving around and picking up items. The grocery store task could also be expanded, for example by adding a second task, where the participant is asked to remember where all the items were located, and place them out on a map, much like Bellassen et al. (2012). Another alternative could be to remove the items in a second task and ask participants to re-visit the correct locations i VR.

There are many possibilities to create VR tasks for cognitive assessment. A-Frame may be particularly accessible, since it is relatively easy to learn and can be used to create both complex and simple virtual environments, for different platforms. The graphical aspects of the environment can be created in the editing tool of choice, or utilize actual photographs, and the objects in the environment can be created in 3D. Audio could also be incorporated, which may enhance the immersion. The grocery store prototype was quite simple, with only one environment, but there are possibilities of linking between environments in VR, which could be used to, for example, create a multi-step task with increasing difficulty without having to leave VR. Furthermore, the task does not necessarily have to be based around a grocery store, one could create other familiar environments, like a kitchen, a wardrobe, a park, or, for example, a labyrinth, depending on which type of cognitive abilities one is attempting to assess. Navigation and orientation appears to be particularly useful to assess, based on the literature study, however it is possible that other cognitive abilities could be assessed utilizing VR, which may require a completely different type of environment.

To make the prototype less excluding based on language, one could create multiple versions of the grocery store in different languages, while keeping the result page in a language the medical personnel understands. Ideally, the patient should perhaps not see the result of the assessment, if this were to be used in a clinical setting. Different cultural backgrounds could also affect the result as it is now, since the way people shop for groceries may differ, and for example affect the expected location of certain items. One possible solution would be to create an even more neutral task. Another issue that was noted was that the shopping list may be too long. This issue could also be solved by creating multiple versions of the grocery store, with various difficulty levels (for example, number of distracting elements, the size of the grocery store, the length of the shopping list). This would also allow for a better adaptation based on the patients' performance, since the task could be scaled up or down based on the performance.

The prototype clearly has many areas that can be improved, and if attempting to create a similar VR task to assess cognitive function one would need to re-design and evaluate the prototype before attempting to test the prototype over multiple groups of participants, establishing a baseline for performance, regarding time, number of items, and order of items. Furthermore, it could be interesting to measure which route the participants take through the store, to further investigate navigation and wayfinding. Another measurement that could possibly be implemented is the time it takes to memorize the shopping list, as users are allowed to attempt to memorize the list for as long as they want to. Then, perhaps, could it be evaluated with participants with cognitive impairment.

Regarding the literature study, a further in-depth investigation into the identified issues and advantages could be beneficial. The list of identified issues and advantages could for example be utilized in a SWOT analysis in order to further come to terms with what each issue and advantage entails for the possible use of VR to assess cognitive function. Furthermore, the literature study was heavily focused on literature pertaining to VR specifically used for detecting cognitive impairment pertaining to AD or MCI, since the study had to be limited to some degree. There are numerous studies focused on other types of impairment, or VR used with other purposes, such as therapy or cognitive training, and these studies may also be beneficial to investigate further, in order to gain a more extensive picture of the field as a whole.

There are several possible advantages for the stakeholders. Being able to assess cognitive abilities using VR in a clinical setting could, in the future, make assessment more efficient, provide a higher degree of ecological validity, and be able to assess with less reliance on native language. Furthermore, there is a possibility of creating a new type of assessment task, from scratch, which could provide the freedom to create tasks based on their own needs and requirements.

8.4 Ethical aspects

One ethical aspect pertaining to the design and evaluation of the grocery store prototype was that one participant experienced slight discomfort due to the placement of the shopping basket. It is ethically problematic to subject participants to discomfort, even if it was unintended. One of the purposes of the shopping basket was to lessen the risk of VR Sickness, but it had the opposite effect for this participant, which points to that it needs further work. However, detecting this type of issue early in the prototyping stage also means the possibility to apply careful measures to mitigate the risk of similar issues in future projects. Another aspect of the evaluation of the prototype is that it may be stressful to be told that the test is intended to assess cognitive function. The participants were informed, in writing, that the grocery store task was only evaluated based on their experiences of the interaction, and that the results of the grocery store task were not stored, and did not indicate anything about their cognitive function. However, this could perhaps have been clarified further, in order to reduce the risk of feeling stressed due to their performance on the task.

Working in healthcare, there are issues of privacy and confidentiality, which are important ethical issues. Digitized assessment may entail the storing of data, which could become problematic, thus it is imperative that patient data is kept secure. Within the prototype, no participant data was stored, and all participants were anonymous. It could theoretically be possible to implement a similar system, where no information pertaining to the identity of the user is stored. However, this may not be practical if the person performing the assessment wants to be able to save and analyze the results. The protection of personal data is regulated by GDPR (General Data Protection Regulation), and the results of an assessment may be considered especially sensitive personal data, putting additional emphasis on the need for data security.

One important consideration is the well-being of the patient, which mean we need to thoroughly investigate the various ergonomic issues with VR. If a VR system for cognitive screening would be implemented in practice, it is important to be mindful of aspects that could affect the results of the screening, as to not be mislead when assessing cognitive function. The potential downsides of utilizing VR needs to be carefully considered and accounted for within this project, and how these downsides compare to currently used methods of cognitive assessment, in terms of patient comfort. Another potential issue to be considered is hygiene, by introducing new equipment into clinical environments, routines for hygiene becomes critical. The HMD needs to be properly sanitized between uses. VR is being used in various public settings, which have systems to sanitize the HMDs, however, one still needs to take hygiene especially carefully into consideration when working in healthcare, to avoid the spreading of diseases.

Accessibility is also an ethical aspect, as it affects the well-being of the patient, and may affect the results of the tests. For example, using certain colors as the only cue, or not using enough contrast, could make things difficult to people with color blindness. Furthermore, there is the question of how VR works when you are visually impaired, even perhaps nearsighted, which needs to be considered as it may affect results. VR set-ups requiring walking, making certain types of movements, or too quick movements could also prove problematic, especially if assessing elderly or disabled patients. This requires careful design considerations. Even if no patients were involved in this project, these types of considerations still need to be kept in mind throughout the design process of this type of system, as these issues primarily can result from the design of the task itself.

Regarding sustainability, VR has both possible advantages and drawbacks. Many neuropsychological assessment tasks are based on pen and paper, or physical objects made out of plastic, and VR could lessen the use of these types of tools, and perhaps even be used to digitize some of the tasks that require these tools. Naturally, a VR task would need to be thoroughly evaluated and tested before even attempting to phase out use of these traditional assessment tools, but perhaps it could be possible in the future. However, VR is not necessarily less demanding of resources, since they require electrical power, and the parts may be made out of plastic, electronics and other materials that require resources that may be difficult to recycle. One notable exception would be Google Cardboard HMDs, made out of cardboard. Cardboard headsets have quite a simple construction and can easily be recycled. The downside is that they cannot be cleaned easily, putting disinfectant on cardboard may disintegrate the material, so cardboard headsets would perhaps need to be single-use items, which may not be ideal. Another drawback of cardboard headsets is that the focus normally cannot be adapted to suit the user, which is an option often available on more robust HMDs. One option could be to use washable HMD covers to lessen the risks associated with hygiene, while producing less waste. Social sustainability is another factor. VR could provide the option to cater to people of different educational backgrounds or native language, which could make the experience of cognitive assessment easier for both the person being assessed, and the person performing the assessment. Furthermore, if the task could be performed without depending on language, the assessment could become more accurate which could aid in, in combination with additional examinations, finding the correct diagnosis and treatment. This, in turn, could make it possible to plan out care, to reach a better quality of life for the individual.

9
Conclusion

The main research question for this project was: *How feasible, both technologically and from a user perspective, is it to assess cognitive function utilizing Virtual Reality, and how could such a system be designed?*. There was also a subquestion, to further explore the possible advantages or disadvantages of using VR within this context: *How does Virtual Reality compare to currently used methods of cognitive assessment?*

The literature study attempts to answer the first part of the research question, *How feasible, both technologically and from a user perspective, is it to assess cognitive function utilizing Virtual Reality.* The results from the literature study, as compiled into a list of identified problems and advantages, indicates that VR shows promise to be used as a tool to assess cognitive function in a clinical setting. The literature study also attempts to answer the subquestion *How does Virtual Reality compare to currently used methods of cognitive assessment?*, by comparing various advantages and disadvantages of VR to advantages and disadvantages of currently used methods of assessment. VR appears to be particularly useful in tasks related to navigation and orientation, when compared to real-environment navigation tasks. VR could also have advantages regarding language-independence, adaptability, comfort, and efficiency. Moreover, VR could provide a higher degree of ecological validity to cognitive assessment tasks, in comparison to certain types of currently used assessment tasks. However, there are also potential problems that are necessary to keep in mind when deciding to pursue VR to assess cognitive function. For example, ergonomic aspects, unfamiliarity with VR both among patients and personnel, as well as the risk of lowered performance compared to real environment tasks and the space and resources required. Some of these issues can, however, be mended through well-informed design decisions.

To answer the second part of the research question: *how could such a system be designed?*, a low-fidelity prototype was developed, and eight factors that need to be considered were identified. The prototype aimed to explore the possibility of creating a VR task to assess cognitive function, from a design perspective, and to uncover issues that may arise. Several issues were discovered during the evaluation, which could illustrate the importance of taking care when designing the movement through the environment, as well as the interaction with the environment, and taking care to mitigate the risk of causing discomfort for the user. Furthermore, the prototype indicates the importance of having the ability to adapt the difficulty level of the task, since some participants found that the shopping list was too long or that the grocery store was too big. It may also be beneficial to investigate currently used assessment methods and discern positive and negative aspects of these methods, and involving experts with knowledge of neuropsychological assessment. The interviews provided insight into the process of currently used methods for assessing cognitive function, and concerned the different issues regarding use of VR, as well as expected benefits. The interviews also helped identify positive aspects of currently used methods of assessment that potentially could be carried over into VR, such as potentially creating tasks with adaptable difficulty levels. Furthermore, to find issues of currently used assessment methods that one could attempt to mitigate using, such as the influence of native language or level of education, which may affect the result of traditional methods. The prototype is at an early stage, and was only evaluated based on user experience and interaction, rather than ability to detect cognitive impairment. In other words, there is no definite answer to how a system used to assess cognitive function should be designed, however the grocery store prototype could serve as a foundation for one such possible solution in the future.

The eight identified factors (section 7.3) could also help guide the design and use of VR tasks for cognitive assessment, in a clinical setting, and can be considered the main knowledge contribution of the project. The eight factors aim to encompass both design considerations, as well as considerations that need to be made when using this type of system in practice: *1: VR Sickness*, *2: Physical discomfort*, *3: The method of interaction*, *4: Movement through the environment*, *5: Adaptability*, *6: Familiarity with VR*, *7: Language, culture and educational background*, and *8: Resemblance to real environment*.

Bibliography

A-Frame (2020) https://aframe.io/, retrieved 2020-01-24

Agrell, B. & Dehlin, O., The Clock Drawing Test (1998) Age and Ageing 27, pp. 399-403. doi: 10.1093/ageing/afs149.

Allain, P., Foloppe, D., Besnard, J., Yamaguchi, T., Etcharry-Bouyx, F., Le Gall, D., . . . Richard, P. (2014) Detecting Everyday Action Deficits in Alzheimer's Disease Using a Nonimmersive Virtual Reality Kitchen, Journal of the International Neuropsychological Society, 20(5), pp. 468-477. doi:10.1017/S1355617714000344

Astell, A. J., Bouranis, N., Hoey, J., Lindauer, A., Mihailidis, A., Nugent, C., Robillard, J. M., & Technology and Dementia Professional Interest Area ... (2019). Technology and Dementia: The Future is Now. Dementia and geriatric cognitive disorders, 47(3), 131–139. https://doi.org/10.1159/000497800

Bellassen, V., Iglói, K., Cruz de Souza, L., Dubois, B., Rondi-Reig, L. (2012) Temporal Order Memory Assessed during Spatiotemporal Navigation As a Behavioral Cognitive Marker for Differential Alzheimer's Disease Diagnosis, Journal of Neuroscience 8 February 2012, 32(6) pp. 1942-1952; DOI: 10.1523/JNEUROSCI.4556-11.2012

Costello, P.J. (1997), Health and Safety Issues Associated With Virtual Reality - A Review of Current Literature, Advanced VR Research Centre, Loughborough University, Leicestershire

Cushman, L. A., Stein, K., & Duffy, C. J. (2008). Detecting navigational deficits in cognitive aging and Alzheimer disease using virtual reality. Neurology, 71(12), pp. 888–895. doi:10.1212/01.wnl.0000326262.67613.fe

De Greve, M. (2018) VR Prototyping with Marvel, retrieved 2020-01-24 from blog.marvelapp.com/vr-prototyping-marvel/

Dyer, E., Swartzlander, B.J., & Guggliucci, M.R. (2018) Using Virtual Reality in Medical Education to Teach Empathy, Journal of the Medical Library Association 106(4), pp. 498-500. DOI: 10.5195/jmla.2018.518

Edwards, J.D., Vance, D.E., Wadley, V.G., Cissell, G.M., Roekner, D.L., & Ball, K.K., (2005) Reliability and Validity of Useful Field of View Test Scores as Administered by Personal Computer, Journal of Clinical and Experimental Neuropsychology 27(5), pp. 529-543, https://doi.org/10.1080/13803390490515432

Edwards, J.D., Ross, L.A., Wadley, V.G., Clay, O.J., Crowe, M., Roekner, D.L., & Ball, K.K., (2006) The Useful Field of View Test: Normative Data For Older Adults, Clinical Neuropsychology 21, pp. 275-286. DOI: doi:10.1016/j.acn.2006.03.001

Evans, R. W., (1998) Complications of Lumbar Puncture, Neurologic Clinics 16(1), pp. 83-105, DOI: 10.1016/S0733-8619(05)70368-6

Flynn, D., van Schaik, P., Blackman, T., Femcott, C., Hobbs, B., & Calderon, C. (2003) Developing a Virtual Reality–Based Methodology for People with Dementia: A Feasibility Study, CyberPsychology & Behavior 2003 6(6), pp. 591-611

Franzen, S., Van den Berg, E., Goudsmit, M., Jurgens, C., Van de Wiel, L., Kalkisim, Y., . . . Papma, J. (2019). A Systematic Review of Neuropsychological Tests for the Assessment of Dementia in Non-Western, Low-Educated or Illiterate Populations. Journal of the International Neuropsychological Society, 1-21. doi:10.1017/S1355617719000894

Funke, S.A. & Willbold, D (2011), Quantitation of Amyloid-B Oligomers in Human Body Fluids for Alzheimer's Disease: Early Diagnosis or Therapy Monitoring? Ronson, C.E., eds, Alzheimer's Diagnosis, Nova Science Publishers Incorporated, ISBN: 978-1-61209-846-3

García-Betances, R.I, Arredondo Waldmeyer, M.T., Fico, G., & Cabrera-Umpiérrez, M.F., (2015) A Succinct Overview of Virtual Reality Technology Use in Alzheimer's Disease, Frontiers in Aging Neuroscience,DOI: https://doi.org/10.3389/fnagi.2015.00080

Gaver, W. W (2012) What Should We Expect From Research Through Design, CHI'12, May 5–10, 2012, Austin, Texas, USA, DOI: 10.1145/2207676.2208538

Github, auxdesigner (2016) github.com/auxdesigner/Sketch-to-VR, retrieved 2020-01-24

Google Blocks (2020) arvr.google.com/blocks/, Google, retrieved 2020-02-2

Hall, J., O'Carroll, R., & Frith, C. (2010). Neuropsychology. Companion to Psychiatric Studies, Eigth edition, pp. 121-140, Elsevier Ltd., ISBN:

978-0-7020-3137-3

Howett, D., Castegnaro, A., Krzywicka, K., Hagman, J., Marchment, D., Henson, R., Rio, M., King, J.A., Burgess, N., & Chan, D. (2019), Differentiation of Mild Cognitive Impairment Using an Entorhinal Cortex-based Test of Virtual Reality Navigation, Brain 2019: 142, pp. 1751-1766, DOI: 10.1093/brain/awz116

IDEO (2015), The Field Guide to Human-Centered Design, IDEO.org, Canada, ISBN: 978-0-9914063-1-9

Jack, C. R., Jr, Knopman, D. S., Jagust, W. J., Petersen, R. C., Weiner, M. W., Aisen, P. S., Shaw, L. M., Vemuri, P., Wiste, H. J., Weigand, S. D., Lesnick, T. G., Pankratz, V. S., Donohue, M. C., & Trojanowski, J. Q. (2013). Tracking pathophysiological processes in Alzheimer's disease: an updated hypothetical model of dynamic biomarkers. The Lancet. Neurology, 12(2), 207–216. DOI: https://doi.org/10.1016/S1474-4422(12)70291-0

Jerald, J., (2016), The VR Book: Human-Centered Design for Virtual Reality, ACM Books series 8, the Association for Computing Machinery and Morgan and Claypool Publishers

Kandalaft, M. R., Didehbani, N., Krawczyk, D. C., Allen, T. T., & Chapman, S. B. (2013). Virtual reality social cognition training for young adults with high-functioning autism. Journal of autism and developmental disorders, 43(1), 34–44. DOI: https://doi.org/10.1007/s10803-012-1544-6

Kessels, R.P.C., van Zandvoort, M.J.E., Postma, A.L., Kappelle, J., & de Haan, E.H.F., (2000) The Corsi Block-Tapping Task: Standardization and Normative Data, Applied Neuropsychology, 7(4), pp 252-258, DOI: 10.1207/S15324826AN0704_8

Laske, C., Sohrabi, H. R., Frost, S. M., López-de-Ipiña, K., Garrard, P., Buscema, M., Dauwels, J., Soekadar, S. R., Mueller, S., Linnemann, C., Bridenbaugh, S., Kanagasingam, Y., Martins, R. M., & O'Bryant, S. E., (2015). Innovative Diagnostic Tools For Early Detection of Alzheimer's Disease, Alzheimer's & Dementia 11, pp. 561-578. http://dx.doi.org/10.1016/j.jalz.2014.06.004.

Lesk, V.E., Wan Shamsuddin, S.N., Walters, E.R. et al.(2014), Using a virtual environment to assess cognition in the elderly, Virtual Reality 18, pp. 271–279. https://doi.org/10.1007/s10055-014-0252-2

Loebach Wetherell, J., Reynolds, C.A., Gatz, M., & Pedersen, N.L. (2002), Anxiety, Cognitive Performance, and Cognitive Decline in Normal Aging, The Journals of Gerontology: Series B, 57(3), pp 246–255, DOI:

https://doi.org/10.1093/geronb/57.3.P246

Manera, V., Chapoulie, E., Bourgeois, J., Guerchouche, R., David, R., Ondrej, J., Drettakis, G., & Robert, P. (2016) A Feasibility Study with Image-Based Rendered Virtual Reality in Patients with Mild Cognitive Impairment and Dementia, PloS one, 11(3), DOI: https://doi.org/10.1371/journal.pone.0151487

Martin, B. & Hanington, B. (2012) Universal Methods of Design: 101 Ways to Research Complex Problems, Develop Innovative Ideas, and Design Effective Solutions, Rockport Publishers

Mrakic-Sposta, S., Di Santo, S.G., Franchini, F., Arlati, S., Zangiacomi, A., Greci, L., Moretti, S., Jesuthasan, N., Marzorati, M., Rizzo, G., Sacco, M. & Vezzoli, A. (2018) Effects of Combined Physical and Cognitive Virtual Reality-Based Training on Cognitive Impairment and Oxidative Stress in MCI Patients: A Pilot Study, Frontiers in Aging Neuroscience, 01 October 2018, DOI: https://doi.org/10.3389/fnagi.2018.00282

National Collaborating Centre for Mental Health (UK) (2007) Dementia: A NICE-SCIE Guideline on Supporting People With Dementia and Their Carers in Health and Social Care. Leicester (UK): British Psychological Society, NICE Clinical Guidelines, No. 42(4), DEMENTIA.

Petersen, R.C, Lopez, O., Armstrong, M.J., Getchius, T.S.D., Ganguli, M., Gloss, D., Gronseth,G.S, Marson, D., Pringsheim, T., Day, G.S, Sager, M., Stevens, J., & Rae-Grant, A. (2018),Practice guideline update summary: Mild cognitive impairment,Neurology Jan 2018, 90(3), pp. 126-135.
DOI: 10.1212/WNL.0000000000004826

Picchi,L., Vista, M., & Mazzoni, M., Milestones and Difficulties in Alzheimer's Diagnosis, Ronson, C.E., eds, Alzheimer's Diagnosis, Nova Science Publishers Incorporated, ISBN: 978-1-61209-846-3

Plancher, G., Tirard, A., Gyselinck, V., Nicolas, S., & Piolino, P., (2012) Using Virtual Reality to Characterize Episodic Memory Profiles in Amnesic Mild Cognitive Impairment and Alzheimer's Disease: Influence of Active And Passive Encoding, Neuropsychologia 50(5) 592–602.
https://doi.org/10.1016/j.neuropsychologia.2011.12.013

de Ribaupierre, S., Kapralos, B., Faizal, H., Stroulia, E., Dubrowski, A., & Eagleson, R. (2014) Healthcare Training Enhancement Through Virtual Reality and Serious Games, Ma, M. et al, (eds), Virtual, Augmented Reality and Serious Games for Healthcare 1, Intelligent Reference Library 68, Springer-Verlag, Berlin, Heidelberg, DOI:10.1007/978-3-642-54816-1_2, Schrepp, M. (2019), The User Experience Questionnaire Handbook: All You Need to Know to Apply the UEQ Successfully in Your Projects, Version 8 (31.12.2019). Online resource, retrieved 2020-05-07 from https://www.ueq-online.org/

Tarnanas, I., Schlee, W., Tsolaki, M., Müri, R., Mosimann, U., Nef, T. (2013), Ecological Validity of Virtual Reality Daily Living Activities Screening for Early Dementia: Longitudinal Study, JMIR Serious Games 2013;1(1):e1, DOI: 10.2196/games.2778

Tombaugh, T. N.,& McIntyre, N.J., (1992). The Mini Mental Status Examination: A comprehensive review. The American Geriatric Society, JAGS. 40(9), pp. 922–935. doi:10.1111/j.1532-5415.1992.tb01992.x

Unity 3D (2020) https://docs.unity3d.com/Manual/VROverview.html, Unity, retrieved 2020-05-21

Unreal Engine (2020) https://www.unrealengine.com/en-US/vr, Unreal Engine, retrieved 2020-05-21

Yeh, S., Chen, Y., Tsai, C., & Rizzo, A. (2012), An innovative virtual reality system for mild cognitive impairment: Diagnosis and evaluation, 2012 IEEE-EMBS Conference on Biomedical Engineering and Sciences, Langkawi, pp. 23-27.