Fariba Mostajeran

MR Pharmacy

Development and Evaluation of Therapeutic Mixed Reality Applications

MR Pharmacy

Development and Evaluation of Therapeutic Mixed Reality Applications

Dissertation

submitted by

Fariba Mostajeran

Human-Computer Interaction
Department of Informatics
Faculty of Mathematics, Informatics, and Natural Sciences
Universität Hamburg

2021

First Reviewer: **Prof. Dr. Frank Steinicke**
Second Reviewer: **Prof. Dr. Albert "Skip" Rizzo**
Chair of the Examination Committee: **Prof. Dr. Simone Frintrop**
Deputy Chair of the Examination Committee: **Prof. Dr. Eva Bittner**

Date of the Defense: **15.06.2021**

Abstract

Mixed Reality (MR) Pharmacy aims at examining psycho-physiological effects of exposure to and interaction with a collection of MR-based therapeutic applications. The scope of the research was narrowed down to personal prevention and management of specific disorders, which mainly fall in the category of mental disorders, i.e., Anxiety Disorders, Psychotic Disorders, Mood Disorders, Alcohol Use Disorders, and Cognitive Disorders. Moreover, in order to take into account the ageing population, a particular age-related category of disorders, namely Balance Disorders was included in this work as well.

For each of these disorders, MR-based therapeutic interventions were designed, developed, and evaluated. Along the way, some of the clinical-MR challenges, such as validation, usability, and acceptance were tackled. To measure different aspects of these interventions, including their effects on human psychology, physiology, and cognition, several user studies were conducted. Additionally, the developed MR-based interventions were validated against in-vivo, control, or non-immersive alternative conditions. The usability of the applications was also measured along with the user experience and acceptance.

The results show that MR provides enormous advantages over tested alternatives. Depending on the disorder and the needs of the therapy receivers, MR-based interventions can be used to decrease or increase psycho-physiological reactions such as anxiety. For instance, for treating Social Anxiety Disorders, one can increase Social Anxiety in training situations, whereas for treating Depressive and Mood Disorders, one can elevate mood and decrease anxiety levels. Moreover, we observed positive effects of exposure to virtual natural environments on human cognition. The MR-based applications were well accepted as well, as users found them novel, encouraging, and stimulating. Furthermore, obtained functional brain connectivity between brain regions as a result of training in MR, imply that MR may be most suited for long-term brain training interventions. Finally, the outcomes of this research provide evidence for practice and facilitate forming the foundations of planning future clinical trials.

Zusammenfassung

Mixed Reality (MR) Pharmacy zielt darauf ab, die psychophysiologischen Auswirkungen der Exposition gegenüber und der Interaktion mit einer Sammlung von MR-basierten therapeutischen Anwendungen zu untersuchen. Die Forschungsarbeit wurde auf die persönliche Prävention und Behandlung bestimmter Störungen fokussiert, die hauptsächlich in die Kategorie der psychischen Störungen fallen, d.h. Angststörungen, psychotische Störungen, Stimmungsstörungen, Alkoholkonsumstörungen und kognitive Störungen. Um den demografischen Wandel zu berücksichtigen, wurde in dieser Arbeit darüber hinaus eine bestimmte altersbedingte Kategorie von Störungen einbezogen, nämlich Gleichgewichtsstörungen.

Für jede dieser Störungen wurden MR-basierte therapeutische Interventionen entworfen, entwickelt und bewertet. Insbesondere wurden klinische MR-Herausforderungen wie zum Beispiel Validierung, Benutzerfreundlichkeit und Akzeptanz addressiert. Es wurden mehrere Benutzerstudien durchgeführt, um verschiedene Aspekte dieser Interventionen zu messen, einschließlich ihrer Auswirkungen auf die menschliche Psychologie, Physiologie und Kognition sowie Benutzerfreundlichkeit und Akzeptanz. Zusätzlich wurde die Wirksamkeit der entwickelten MR-basierten Interventionen gegen Real-, Kontroll- oder nicht-immersive Alternativen validiert.

Die Ergebnisse zeigen, dass MR enorme Vorteile gegenüber den Alternativen bietet. Abhängig von der Störung und den Bedürfnissen der Therapieempfänger*innen können MR-basierte Interventionen verwendet werden, um psychophysiologische Reaktionen wie z.B. Angstzustände zu verringern oder zu verstärken. Zum Beispiel lässt sich zur Behandlung von sozialen Angststörungen die soziale Angst in Trainingssituationen erhöhen, während sich zur Behandlung von depressiven und Stimmungsstörungen die Stimmung verbessern und das Angstniveau senken lässt. Wir beobachteten positive Auswirkungen der Exposition gegenüber virtuellen natürlichen Umgebungen auf die menschliche Kognition. Die MR-basierten Anwendungen wurden durch die Benutzer*innen als neuartig, ermutigend und anregend wahrgenommen. Darüber hinaus impliziert die sich aus dem MR-Training ergebene funktionelle Gehirnkonnektivität zwischen Gehirnregionen, dass MR am besten für langfristige Gehirntrainingsinterventionen geeignet sein kann. Somit liefert diese Arbeit die Grundlage und wichtige Erkenntnisse für die Planung künftiger klinischer Studien.

Acknowledgements

PhD was a dream, which came true only with the generous support of ***Prof. Dr. Frank Steinicke***, who trusted and welcomed me in his group, when I was at the lowest level of confidence in my life. I cannot express enough gratitude for his trust, support, and scientific advice. On top of that, he has created a research group with the friendliest atmosphere I have ever experienced, in which doing research is a real joy.

He also engaged me in our first collaboration with the Department of Psychiatry and Psychotherapy at the UKE, through which I started my close collaboration with ***Prof. Dr. Simone Kühn***. Without a doubt, without her generous time and kind scientific advice, major contributions of this thesis could not be achieved.

Also, my deepest gratitude goes to all my colleagues at the HCI group, who acted more than mere colleagues and have become my close friends; to all students who chose me as their thesis and project adviser and tolerated my learning process of becoming an adviser; to all my Holobalance project partners who generously provided me with a platform to perform studies beyond the borders of my located research group; to all my co-authors for fruitful collaborations; and to all participants of our studies who provided the value of human research to this work.

Finally, I would like to thank my family for loving support throughout my entire education, specially my mother who envisioned me always as a great scientist.

Contents

List of Figures

List of Tables

List of Acronyms and Abbreviations

2D Two-dimensional.

3D Three-dimensional.

AAL Anatomical Automatic Labelling.

AAT Approach Avoidance Training.

ABGS American and British Geriatric Society.

APA Activity Planning App.

AR Augmented Reality.

ARET Augmented Reality Exposure Therapy.

ART Attention Restoration Theory.

ATT Auditory Training Tool.

AUD Alcohol Use Disorder.

AV Augmented Virtuality.

BDI Becks Depression Inventory.

BOLD Blood-Oxygen-Level Dependent Imaging.

BPH Balance Physiotherapy Hologram.

CBT Cognitive Behavioural Therapy.

CET Cue-Exposure Therapy.

CF Compassion Focused.

CFI Compassion-Focused Imagery.

CTEG Cognitive Training and Exercise Games.

DHI Digital Health Interventions.

DLPFC Dorsolateral Prefrontal Cortex.

DM Dashboard Module.

DMN Default Mode Network.

DSM Diagnostic and Statistical Manual of Mental Disorders.

ECG Electrocardiogram.

EEG Electroencephalography.

EPI Echo-Planar Imaging.

ET Exposure Therapy.

fALFF fractional Amplitude of Low-Frequency Fluctuation.

FC Functional Connectivity.

FD Framewise Displacement.

fMRI functional Magnetic Resonance Imaging.

fNIRS functional Near-infrared Spectroscopy.

FOV Field of View.

FSS Flow Short Scale.

FWE Family Wise Error.

G-AAT Gamified AAT.

GOe Global Observatory for eHealth.

GSR Galvanic Skin Response.

HCI Human-Computer Interaction.

HMD Head-Mounted Display.

HQ-I Hedonic Quality-Identity.

HQ-S Hedonic Quality-Stimulation.

HR Heart Rate.

HSC Health System Challenges.

IAT Implicit Association Test.

ICA Independent Component Analysis.

ICD International Classification of Diseases.

ICT Information and Communication Technologies.

IPQ Igroup Presence Questionnaire.

LSAS Liebowitz Social Anxiety Scale.

MCWS Motion Capture and Wearable Sensors.

MINI Mini International Neuropsychiatric Interview.

MNI Montreal Neurological Institute.

MR Mixed Reality.

MRI Magnetic Resonance Imaging.

MRM MR Medication.

N-AAT Non-Gamified AAT.

NASA-TLX NASA Task Load Index.

NICE National Institute of Clinical Excellence.

NIMH National Institute of Mental Health.

NSDUH National Survey on Drug Use and Health.

PACT People, Activities, Contexts, and Technologies.

PANAS Positive Affect Negative Affect Schedule.

POMS Profiles of Mood States.

PQ Pragmatic Quality.

PSS Perceived Stress Scale.

RDoC Research Domain Criteria.

ROI Region of Interest.

RSN Resting State Networks.

RV Reality-Virtuality.

SA Social Anxiety.

SAD Social Anxiety Disorder.

SAM Self-Assessment Manikin.

SCID Structured Clinical Interview for DSM.

SCR Skin Conductance Response.

SCS Self-Compassion Scale.

SeedFC Seed-based Functional Connectivity.

SIT Social Impact theory.

SPDQ Social Phobia Diagnostic Questionnaire.

SSA State Social Anxiety.

SSQ Simulator Sickness Questionnaire.

SSSQ Short Stress State Questionnaire.

STADI State Trait Anxiety Depression Inventory.

STADI-S State Trait Anxiety Depression Inventory-State.

STADI-T State Trait Anxiety Depression Inventory-Trait.

SUS System Usability Scale.

TR Tele-Rehabilitation.

TSST Trier Social Stress Test.

UEQ User Experience Questionnaire.

UKE The University and Clinical Center Hamburg-Eppendorf.

UN United Nations.

UX User Experience.

VA Virtual Audience.

VE Virtual Environment.

VH Virtual Human.

VR Virtual Reality.

VRET Virtual Reality Exposure Therapy.

WG Whole Game.

WHO World Health Organization.

Chapter 1

1 Introduction

1.1 Digital Health Interventions

According to the United Nations (UN), health is a global issue [371], as at least half of the world's population does not receive essential health services [403]. Within the UN system, the World Health Organization (WHO) leads the efforts to address this global issue [372]. According to WHO [401], health systems are associated with eight categories of challenges, which affect both consumers as well as providers of health goods and services. Table 1.1 provides a list of these global Health System Challenges (HSC).

Fortunately, Information and Communication Technologies (ICT) have enormous potential to bear positively on HSC [398]. In fact, ICT is increasingly being integrated into health systems and services worldwide and improvements in health as a direct benefit of ICT are being made every day across the world [399].

WHO has also acknowledged the growing importance of digital health and thus, has carried out a number of key actions to facilitate its integration in health systems and services [398]. For instance, the Global Observatory for eHealth (GOe) has been initiated to monitor digital health progression and guide its developments around the world [400]. WHO has also provided a classification of Digital Health Interventions (DHI), which are being used to support health system needs [401]. In order to effectively show how ICT can address identified health needs, it has been recommended to use DHIs in tandem with the identified HSCs.

In particular, DHIs have great potentials to address some of the HSCs by overcoming the limitations of traditional health interventions and bringing additional advantages to this domain [95, 13, 346, 233]. For instance, they could provide anonymous, cost-effective, and flexible (in self-direction and self-pacing) solutions, which can make interventions more accessible and acceptable to the general public [95, 65, 345, 13]. In addition, multi-sensory (e.g., visual and auditory) interactions can make these interventions more appealing and attractive [95], addressing, for instance, the "insufficient patient engagement" HSC (see Table 1.1).

Table 1.1: Health System Challenges (adapted from [401]).

Challenge	Example
Information	• Lack of population denominator • Delayed reporting of events • Lack of quality/reliable data • Communication roadblocks • Lack of access to information or data • Insufficient utilization of data and information • Lack of unique identifier
Availability	Insufficient supply of: • Commodities • Services • Equipment • Qualified health workers
Quality	• Poor patient experience • Insufficient health worker competence • Low quality health commodities • Low health worker motivation • Insufficient continuity of care • Inadequate supportive supervision • Poor adherence to guidelines
Acceptability	• Lack of alignment with local norms • Ignorance of individual beliefs and practices
Utilization	• Low demand for services • Geographic inaccessibility • Low adherence to treatments • Loss to follow up
Efficiency	• Inadequate workflow management • Lack of or inappropriate referrals • Poor planning and coordination • Delayed provision of care • Inadequate access to transportation
Cost	• High cost of manual processes • Lack of effective resource allocation • Client-side expenses • Lack of coordinated payer mechanism
Accountability	• Insufficient patient engagement • Unaware of service entitlement • Absence of community feedback mechanisms • Lack of transparency in commodity transactions • Poor accountability between the levels of the health sector • Inadequate understanding of beneficiary populations

1.2 Mixed Reality Technologies

The technological concept of Virtual Reality (VR) was first formulated at the mid of 1960s by Ivan Sutherland, who described VR as an “ultimate display”, through which a user perceives a Virtual Environment (VE) as if it is real [356]. Ever since, several definitions have been formulated for VR, which all highlight three common features: (i) immersion, (ii) sense of presence, and (iii) interaction with the VE [130, 120, 82, 42, 226, 227, 151, 43, 24, 333, 334, 354, 338].

Immersive VR systems provide users with an omnidirectional view of the VE, enabling them to perceive, feel, and interact similar as in reality [338]. This is usually achieved by combining multiple sensory channels, such as sight, sound, and touch. Typically, immersive VR systems include a Head-Mounted Display (HMD), a.k.a., VR glasses or goggles, which displays VEs while blocking the users' Field of View (FOV) to the physical world. The head tracking sensors of a typical VR system compute head position to corresponding views in the VE. Additional input devices (a.k.a., VR controllers) enable interaction with VE through touching, grabbing, or manipulating the virtual objects [348].

An essential feature of VR is its ability to evoke a sense of presence, i.e., a sense of being physically present in the VE [337]. This characteristic leads to a human behavior that is similar to the behavior shown in real world settings. Importantly this sense of presence may facilitate transfer of knowledge and skills acquired in VR to similar real world situations, which would make VR an ideal choice for therapeutic interventions [51]. In addition, experiments, training, and therapy sessions that use immersive VR would be able to reach a high level of ecological validity for situations, which may be too dangerous, expensive, or impossible to create otherwise [292]. Moreover, experimenters and therapists may benefit from a systematic control over the experiment and/or therapy. The low-cost VEs, which can be duplicated and distributed easily, would also make VR-based therapeutic applications accessible at a larger scale [293]. Thus, VR can complement the human research by maximizing the benefits of lab-based (e.g., control over independent variable) and field-based (e.g., realistic stimuli) experiments [340].

In contrast to VR, in which user's entire FOV to the real world is blocked, in Augmented Reality (AR), users can see the real world (either through cameras or see-through displays) augmented with additional virtual objects [314]. Therefore, AR systems should register real and virtual objects with each other while running interactively and in real-time [22]. Typically, virtual objects in AR environments should provide an indistinguishable simulation of the reality. For instance, an augmented virtual spider should appear as real as in the real world and an augmented Virtual Human (VH) should look and behave as similar as to a real human being. This way, AR may evoke similar to reality human behaviors towards augmented virtual objects [49, 175, 59, 404].

The concept of VR and AR systems might be best explained by the Reality-Virtuality (RV) Continuum (see Figure 1.1), which was introduced by Milgram and Kishino [241] in 1994. Fully real and virtual environments span a continuum and any display, which merges any degree of reality and virtuality, lies in between and can be explained by the general term of Mixed Reality (MR). Therefore, AR is created when some degree of virtuality is added to reality, whereas Augmented Virtuality (AV) is created by means of adding some degree of reality to purely virtual worlds. Since the pure virtuality as conceptualized by Ivan

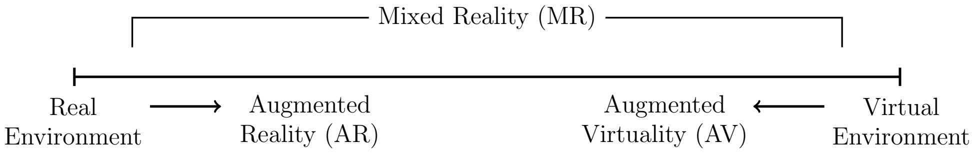


Figure 1.1: Reality-Virtuality Continuum (adapted from [242]).

Sutherland, in which, for example, a "virtual bullet would be fatal", has not been achieved so far, the general term of MR could be used to include AR and all current VR systems.

1.3 Clinical-MR

MR technologies, in particular, VR and AR have enormous potentials to contribute to health services by, for instance, being used as DHIs. In fact, these potentials have been realized as early as mid 1990's and until today, numerous scientific findings have been reported on the applications of VR for therapy and rehabilitation [76]. Ever since, quite a number of scientific journals, conferences, and handbooks have been dedicated to the subject of clinical-VR as well [51].

Successful application of VR in clinical contexts have been demonstrated, for instance, for treating psychological disorders [292, 75], pain remediation [158], neurocognitive therapy [51], and rehabilitation (e.g., recovery after stroke [68]). It has been also suggested that users show higher preference and acceptance of VR versus in vivo therapy. For instance, 81-89% of the subjects of a study on Arachnophobia (i.e., fear of spiders) [126] and 76% of the participants of another study on specific phobias [125] prefer Virtual Reality Exposure Therapy (VRET) over in vivo exposure. In the latter study by Garcia et al. [125], the refusal rates for the in vivo Exposure Therapy (ET) (27%) were also significantly higher than for the VRET (3%). Therefore, VR provides a preferred alternative and could potentially increase the number of individuals, who seek and undergo treatment.

Compared to VR, AR has a more recent research and application history [76]. Nonetheless, its efficacy in different research fields including clinical treatments such as the treatment of phobia to small animals through Augmented Reality Exposure Therapy (ARET), has been shown by researchers of the field [49, 175, 59, 404].

Reported advancements in clinical-MR research has been also reflected in practice and practical guidelines, such as the American Psychological Association's Practice Guideline for the Treatment of Patients with Acute Stress Disorder and Post-traumatic Stress Disorder [36] as well as its Presidential Task Force on Evidence-based Practice in Psychology [258]. In addition, a survey with 70 psychotherapy experts has revealed that VR will be among the top five intervention methods, which will be increasingly used in practice, with other computer-supported methods, such as smartphone applications, occupying four out of the top five methods [254].

As a matter of fact, significant investments in ICT has shifted the VR research towards the so-called clinical-VR era in which clinical applications of VR have become the leading

Table 1.2: Clinical-MR Challenges (adapted from [380]).

Challenge	Example
Validation of Clinical Efficacy	• Poor study design • Small sample sizes • Lack of control groups • Variable technical standards • Lack of data regarding adverse reactions • Lack of focus on economic feasibility • Varying MR hardware and software • Lack of data regarding long-term effects • Need for quantification of efficacy associated with MR vs. conventional treatments • Need for "virtual pharmacy"
Cost/Accessibility	• Cost of MR equipment • Expense of developing MR software • Rapid obsolescence of MR equipment • Potential inefficacy of lower-end equipment • Unpredictable reimbursement by health insurance
Usability	• Technical glitches and malfunctions • Need for training to achieve technical proficiency • Cybersickness • Limited exposure to reduce cybersickness • Ergonomics (e.g., weight and bulkiness) of HMDs • Difficulty interacting with MR interface • Tripping hazard created by MR equipment • Lack of computer/video-game literacy • Need for dedicated space/interferes with usability of environment for other purposes • Potential for interference by pets or children • Interference of MR with other electronic equipment • Calibration of MR equipment
Technical Capabilities	• Limitation of technical capabilities • Limitations of input data via user's voice • Limitations to incorporate data from patient portals • Uncanny valley effects of unrealistic avatars
Acceptance	• Unclear application in everyday clinical practice • Unknown extent of patient acceptance • Reluctance of older patients to use MR • Patient discouragement from confusion, discomfort, and difficulty operating MR equipment • Perception of unclear benefits by patients • Preference for traditional educational approaches

research field, making it one of the most investigated ever topics in the field [76].

Yet, there are several challenges, which hinder comprehensive use of MR potentials for therapy and rehabilitation. Ventola [380] has summarized these challenges for clinical-VR (see Table 1.2), which could be relevant for clinical-MR technologies in general.

1.4 MR Pharmacy

The main motivation of this thesis is to advance the research on DHIs, in particular, MR application for physical and mental health. As addressing all HSC is out of the scope of this work, the focus was narrowed down to some main challenges. To shed a light on these challenges, a number of interviews, discussions, and focus groups with health professionals were performed. The data was collected from interviews with the local (i.e., The University and Clinical Center Hamburg-Eppendorf (UKE)), national (e.g., The Rehab Hospital Hansenbarg, The Clinic Bremen-Ost, The University and Clinical Center Freiburg), and EU-wide (e.g., The National and Kapodistrian University of Athens, University College London) clinics and poly-clinics as well as a charity in London (i.e., Age UK London). As a result, the following challenges were identified as relevant within the scope of our collaborative projects:

- limited availability of the trained health professionals,
- low adherence and motivation of the therapy receivers and patients to the therapeutic plans, and
- limited availability and high costs of individualized interventions.

These challenges confirm some of the previously mentioned HSC regarding Availability, Quality, and Costs (see Table 1.1). Thus, they were chosen as the main challenges to focus upon in this work.

Furthermore, in consultation with the clinical experts, the scope of the health conditions was narrowed down to personal prevention and management of a few disorders, which mainly fall in the category of mental disorders based on the International Classification of Diseases (ICD) [402]. Thereby, one disorder from some selected categories of mental disorders, i.e., Mood Disorders (Chapter 3), Cognitive Disorders (Chapter 4), Anxiety Disorders (Chapter 5), Psychotic Disorders (Chapter 6), and Alcohol Use Disorders (Chapter 7) were studied.

Moreover, the continuous improvement of the health systems has successfully extended the human longevity globally [403]. This phenomena along with reduction in levels of fertility has resulted in an inevitable increase of the ageing population and a shift in the distribution of the population age from younger to older ages [373]. Therefore, to take this unignorable increasing group of individuals into account, a particular age-related category of disorders, namely Balance Disorders (Chapter 8), was included in this work as well.

For each of these disorders, MR-based DHIs were designed, developed, and evaluated. Along the way, some of the clinical-MR challenges (see Table 1.2), such as Validation, Usability, Technical Capabilities, and Acceptance were tackled as well. Several user studies were conducted to measure different aspects of MR-based therapeutic applications, including their effects on human psychology, physiology, and cognition. Additionally, the developed MR-based interventions were validated against in-vivo, control, or non-immersive alternative conditions (Chapters 2 - 7). The usability of the applications was also measured along with the user experience and acceptance (Chapter 7 and Chapter 8).

To summarize, the contributions of this thesis are:

- design and development of a collection of therapeutic MR interventions for different disorders,
- inclusion of clinical as well as Human-Computer Interaction (HCI) measures in the evaluation,
- validation of MR techniques against in-vivo, control, or non-immersive alternatives for clinical applications, and
- inclusion of end users in the design, development, and evaluation process.

1.4.1 Publications

Major parts of this thesis have been published in peer-reviewed journals and conference proceedings.

Main Authorship

To the following list of publications, I contributed as main author, responsible for experiment design, development, data collection, analysis, and authoring the paper:

- **F. Mostajeran**, J. Krzikawski, F. Steinicke, and S. Kühn, "Effects of exposure to immersive videos and photo slideshows of forest and urban environments," *Scientific Reports*, vol. 11, no. 3994, pp. 1–14, Nature Publishing Group, 2021.
- **F. Mostajeran**, F. Steinicke, O. J. Ariza Nunez, D. Gatsios, and D. Fotiadis, "Augmented reality for older adults: Exploring acceptability of virtual coaches for home-based balance training in an aging population," in *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, pp. 1–12, Association for Computing Machinery (ACM), 2020.
- **F. Mostajeran**, M. B. Balci, F. Steinicke, S. Kühn, and J. Gallinat, "The effects of virtual audience size on social anxiety during public speaking," in *Proceedings of the 2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 303–312, IEEE, 2020.
- **F. Mostajeran**, N. Katzakis, O. Ariza, J. P. Freiwald, and F. Steinicke, "Welcoming a holographic virtual coach for balance training at home: two focus groups with older adults," in *Proceedings of the 2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR), Presented at IEEE VR Workshop on Applied VR for Enhanced Healthcare (AVEH)*, pp. 1465–1470, IEEE, 2019.
- **F. Mostajeran**, A. Kirsten, F. Steinicke, J. Gallinat, and S. Kühn, "Towards gamified alcohol use disorder therapy in virtual reality: A preliminary usability study," in *Proceedings of the 2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR), Presented at IEEE VR Workshop on Applied VR for Enhanced Healthcare (AVEH)*, pp. 1471–1476, IEEE, 2019.

- **F. Mostajeran**, A. Kirsten, F. Steinicke, J. Gallinat, and S. Kühn, "Virtual alcohol use disorder therapy: a pact analysis and two focus groups," in *Mensch und Computer 2019-Workshopband*, pp. 595–599, Gesellschaft für Informatik e.V., 2019.

Co-Authorship

In addition, I was not the lead author of the following publications. Yet, this dissertation makes use of parts of them as I contributed to the critical parts of the works such as experiment design and development, data collection and analysis, as well as authoring the paper:

- L. Ascone, K. Ney, **F. Mostajeran**, F. Steinicke, S. Moritz, J. Gallinat, and S. Kühn, "Virtual reality for individuals with occasional paranoid thoughts," in *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems*, pp. 1–8, Association for Computing Machinery (ACM), 2020.
- T. Androutsou, I. Kouris, A. Anastasiou, S. Pavlopoulos, **F. Mostajeran**, D.-E. Bamiou, G. Genna, S. G. Costafreda, and D. Koutsouris, "A smartphone application designed to engage the elderly in home-based rehabilitation," *Frontiers in Digital Health*, vol. 2, p. 15, Frontiers, 2020.
- K. M. Tsiouris, D. Gatsios, V. Tsakanikas, A. A. Pardalis, I. Kouris, T. Androutsou, M. Tarousi, N. Vujnovic Sedlar, I. Somarakis, **F. Mostajeran**, N. Filipovic, H. op den Akkeri, D. D. Koutsourisa, and D. I. Fotiadis, "Designing interoperable telehealth platforms: bridging iot devices with cloud infrastructures," *Enterprise Information Systems*, pp. 1–25, Taylor & Francis, 2020.
- C. G. Forlim, L. Bittner, **F. Mostajeran**, F. Steinicke, J. Gallinat, and S. Kühn, "Stereoscopic rendering via goggles elicits higher functional connectivity during virtual reality gaming," *Frontiers in Human Neuroscience*, vol. 13, p. 365, Frontiers, 2019.

Other

Finally, the following list presents my further publications as main or co-author, which are not used as part of this dissertation:

- J. P. Freiwald, Y. Göbel, **F. Mostajeran**, and F. Steinicke, "The cybersickness susceptibility questionnaire: predicting virtual reality tolerance," in *Proceedings of the Conference on Mensch und Computer*, pp. 115–118, Association for Computing Machinery (ACM), 2020.
- T. Machulla, A. Treskunov, F. Lang, S. Rings, C. Prasuhn, **F. Mostajeran**, H. Klapperich, M. Hassenzahl, and C. Geiger, "Virtuelle und augmentierte realität für gesundheit und wohlbefinden," in *Mensch und Computer 2019-Workshopband*, pp. 576–578, Gesellschaft für Informatik e.V., 2019.

- J. Zhang, N. Katzakis, **F. Mostajeran**, P. Lubos, and F. Steinicke, "Think fast: Rapid localization of teleoperator gaze in 360 hosted telepresence," *International Journal of Humanoid Robotics*, vol. 17, no. 03, p. 1950038, World Scientific Publishing Company, 2020.
- J. Zhang, N. Katzakis, **F. Mostajeran**, and F. Steinicke, "Localizing teleoperator gaze in 360 hosted telepresence," in *Proceedings of the 2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 1265–1266, IEEE, 2019.
- N. Katzakis, L. Chen, **F. Mostajeran**, and F. Steinicke, "Peripersonal visual-haptic size estimation in virtual reality," in *Proceedings of the 2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 1010–1011, IEEE, 2019.
- **F. Mostajeran**, P. Joschko, and J. Göbel, "Heuristic optimisation and simulation as decision support for operation and maintenance of offshore wind farms," in *Proceedings of the International Conference on Harbour, Maritime and Multimodal Logistic Modelling and Simulation, organized within the International Mediterranean and Latin American Modelling Multi-Conference*, 2017.
- **F. Mostajeran**, P. Joschko, J. Göbel, B. Page, S. Eckardt, and T. Renz, "Offshore wind farm maintenance decision support," in *Proceedings of EnviroInfo*, pp. 241–247, 2016.
- F. Ritter, R. Porzel, **F. Mostajeran**, and R. Malaka, "Brain-computer-interface for software control in medical interventions," in *Proceedings of Die Deutsche Gesellschaft für Computer- und Roboterassistierte Chirurgie (CURAC)*, 2015.

Chapter 2

2 Stereoscopic Rendering via Goggles Elicits Higher Functional Connectivity During Virtual Reality Gaming

This chapter presents a study to address the question of whether brain connectivity differs between VR stimulation via goggles and a presentation from a screen via mirror projection. In addition, the question of whether stereoscopic goggle stimulation, where both eyes receive different visual input, elicit stronger brain connectivity than a stimulation in which both eyes receive the same visual input (monoscopic) was investigated as well. The motivating assumption is that the condition that elicits higher brain connectivity values should be most suited for long-term brain training interventions given that, extended training under these conditions could permanently improve brain connectivity on a functional as well as on a structural level.

To do so, multiple analyses approaches were taken so that different aspects of brain connectivity could be covered: fractional low frequency fluctuation, independent component analysis, seed-based functional connectivity and graph analysis. As a result, in goggle presentation (mono and stereoscopic) as contrasted to screen, differences in brain activation in left cerebellum and postcentral gyrus as well as differences in connectivity in the visual cortex and frontal inferior cortex (when focusing on the visual and default mode network) was found. When considering connectivity in specific areas of interest, higher connectivity between bilateral superior frontal cortex and the temporal lobe, as well as bilateral inferior parietal cortex with right calcarine and right lingual cortex was observed. Furthermore, superior frontal cortex and insula/putamen to be more strongly connected in goggle stereoscopic vs. goggle monoscopic was found.

2.1 Introduction

The study presented in this chapter was conducted to test to what extent VR technology for games presented via goggles can be used in a Magnetic Resonance Imaging (MRI) scanner. It was set out to investigate whether VR visual stimulation using MRI compatible goggles with 3D stereoscopic stimulation (in which the image is rendered separately for each eye creating the illusion of depth and 3D effect) differs in terms of brain connectivity from more commonly applied presentation forms using goggles with Two-dimensional (2D) monoscopic presentation (in which both eyes receive the same visual input) and a conventional screen back-projection via a mirror. It was hypothesized that the condition that elicits higher brain connectivity values should be most suited for long-term brain training interventions given that, extended training under these conditions could permanently improve brain connectivity. To study potential brain connectivity differences elicited by Three-dimensional (3D) stereoscopic, 2D monoscopic and screen stimulations, multiple methods were chosen, each of them being able to reveal different aspects of brain connectivity: Independent Component Analysis (ICA) a data-driven technique to extract whole brain networks, Seed-based Functional Connectivity (SeedFC) that calculates the brain network related to specific regions of interest and graph analysis that characterizes the topology of the brain networks. Additionally, brain activation in the domain of low frequency fluctuation of the BOLD signal using fractional Amplitude of Low-Frequency Fluctuation (fALFF) was tested. Before the publication of this study, there was no previous research using games and functional connectivity in MRI to address this question. Most of the previous studies either used different stimulus material to investigate different degrees of spatial presence [30, 31, 91, 148, 214] and/or used Electroencephalography (EEG) [31, 86, 148, 194, 339]. Gaebler and colleagues [123], employed the same movie as stimulus delivered in 3D stereoscopic and monoscopic 2D and evaluated subject experience and inter-subject correlation of brain activity finding higher immersion and more realistic the stimulus was associated with higher inter-subject correlations. Nevertheless, the connectivity between brain areas was not investigated.

Most interesting for the present endeavor and the question whether VR elicits higher brain connectivity, is an EEG study that compared brain signals during navigation either on a desktop PC (2D) or on a large wall projection in 3D [194]. The 3D condition was accompanied by higher cortical parietal activation in the alpha band, whereas the 2D condition was accompanied by stronger functional connectivity between frontal and parietal brain regions, indicating enhanced communication. In two additional EEG studies, in which different modes of presentation have been compared, but in which brain activity instead of connectivity was the focus of investigation, contradictory evidence has been gathered. Likewise in a navigation task comparing a condition in which participants wore 3D glasses and watched a screen vs. a 2D screen condition, higher theta power in frontal midline structures was observed in the 3D VR condition [339]. In contrast a study investigating paper folding (origami) learning with 3D glasses vs. with a 2D film showed that the 2D condition displays a higher so-called cognitive load index computed as the ratio of the average power of frontal theta and parietal alpha. A last study focused on intra-hippocampal EEG recordings comparing real world navigation vs. VR navigation and demonstrated that oscillations typically occurs at a lower frequency in virtual as compared to real world navigation [46]. Brain regions that have repeatedly received attention in the endeavor to explain differences between VR-related presence experiences in comparison to 2D or less immersive environments are the prefrontal cortex, the parietal cortex as well as the hippocampus [30, 31, 35, 46, 86, 194]. For this reason, we used whole brain connectivity analysis approaches as well as Region of Interest (ROI)-based approaches focusing on the effects of the type of the display (conventional 2D screen, MRI goggles with monoscopic view effect, MRI goggles with stereoscopic view effect) which capture two different degrees of immersion in a technical [338] and subjective [123] point of view.

2.2 Methods

2.2.1 Participants

Twenty-six healthy participants were recruited from a local participant pool. After complete description of the study, the participants' informed written consent was obtained. Exclusion criteria for all participants were abnormalities in MRI, relevant general medical disorders and neurological diseases. Additional exclusion criteria were movement above the threshold of .5 [276] during the scanning section and completion of all 3 conditions. After the exclusion criteria were applied, the number of subjects dropped to twenty three (mean age = 26.5, SD = 4.8 years, female:male = 11:12). The protocol was approved by the local psychological ethics committee of the University Medical Center Hamburg-Eppendorf, Germany. All subjects gave written informed consent in accordance with the Declaration of Helsinki.

2.2.2 Game description and procedure

While situated in the scanner participants played our FloVR game (see Chapter 3, Section 3.3 for more details). In brief, FloVR was designed to induce positive mood and reduce depressivity. In the game, players should fly through a nature scene with the goal of making the landscape blossom (Figure2.1). This can be achieved by the player flying close to flowers, which are surrounded by a coloured halo, to virtually pollinate them so that more flowers grow in the surrounding area. Visual and auditory elements associated with positive affect were integrated in the game.

Before entering the scanner, the participants were given information about the gameplay. The participants did not practice outside the scanner. They were asked to imagine to be a bee, whose goal is to make the landscape flourish by flying above it and touching flowers or trees which are surrounded by a bright halo. That would lead to pollen popping out of the flowers and new flowers, bushes or trees growing all around. They were told to follow the path of the flowers and collect as many flowers as possible, however, they were likewise instructed to not mind when missing out on a flower. The number of pollinated flowers was not counted and no score was kept. For this reason no behavioural measures of the game were reported.

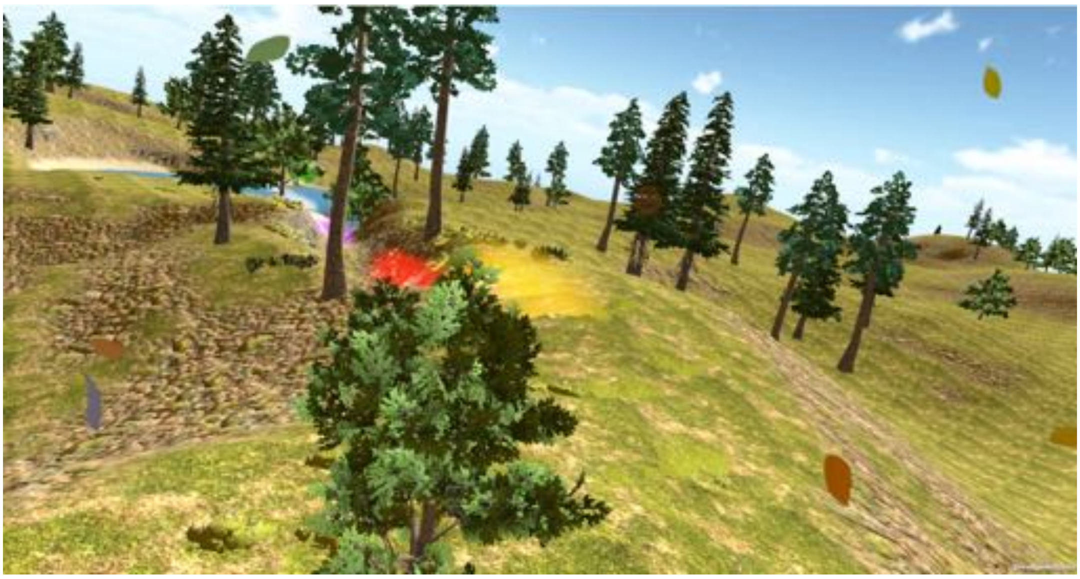


Figure 2.1: A screenshot of the player's view when flying in the virtual landscape. The colorful halos indicate the positions of the next flowers to be pollinated.

Participants used an MRI compatible button box with four buttons in a row from Nordic Neuro Lab to navigate in the game. The user had to hold the controller with both hands and use the two left buttons for flying upwards and downwards and the other two for flying to the left and right. Regardless of the number of pollinated flowers, the speed of the flight was kept constant and each run lasted about 5 min.

Participants underwent three conditions, one in which the game was projected from a screen via mirror projection, one in which MRI compatible goggles were used either with a 3D stereoscopic stimulation (in which the image is rendered separately for each eye creating the illusion of depth and 3D effect) and a 2D monoscopic presentation (in which both eyes received the same visual input). MRI compatible goggles used were the VisuaStim digital [166], with a resolution of 800×600 (SVGA) and FOV of 30 degrees horizontal and 24 degrees vertical. The order of the conditions was randomly assigned to the participants. Due to subject exclusion the order of condition, monoscopic-stereoscopic – screen were performed for 3 subjects instead of 4 (23 subjects distributed among 6 orders of condition).

2.2.3 Scanning Procedure

Structural images were collected on a Siemens Prisma 3T scanner (Erlangen, Germany) and a standard 32-channel head coil was used. The structural images were obtained using a three-dimensional T1-weighted magnetization prepared gradient-echo sequence (MPRAGE) (repetition time = 2500 ms; echo time = 2.12ms; TI = 1100 ms, acquisition matrix = $240 \times 241 \times 194$, flip angle = $9°; 0.8 \times 0.8 \times 0.94$ mm voxel size). Resting state data was acquired after the T1 image. We acquired whole brain functional images while participants were asked to keep their eyes closed and relax for 5 min. We used a T2*-weighted Echo-Planar Imaging

(EPI) sequence (repetition time = 2,000 ms, echo time = 30 ms, image matrix = 64 × 64, field of view = 216 mm, flip angle = 80°, slice thickness = 3.0 mm, distance factor = 20 %, voxel size = $3 \times 3 \times 3mm^3$, 36 axial slices, using GRAPPA). Images were aligned to the anterior-posterior commissure line.

2.2.4 Functional MRI Data Analysis

Preprocessing

To ensure for steady-state longitudinal magnetization, the first 10 images were discarded. Slice timing and realignment were performed in the remaining images. The individual anatomical images T1 were coregistered to functional images and segmented into white matter, gray matter, and cerebrospinal fluid. Data was then spatially normalized to Montreal Neurological Institute (MNI) space and spatially smoothed with an 8-mm FWHM to improve signal-to-noise ratio. Signal from white matter, cerebrospinal fluid and movement were regressed. To reduce physiological high-frequency respiratory and cardiac noise and low-frequency drift data was filtered in the bandwidth of (0.01 – 0.08 Hz) and, finally, detrended. All steps of data preprocessing were done using SPM12 (Wellcome Department of Cognitive Neurology) except filtering that was applied using REST toolbox [343]. In addition, to control for motion, the voxel-specific mean Framewise Displacement (FD) was calculated according to Power and colleagues [276]. We excluded from the analyses participants who had an FD above the recommended threshold of 0.5.

Fractional Amplitude of Low-Frequency Fluctuation (fALFF)

fALFF is not a measure of connectivity between areas, but rather it accounts, voxel-wisely, for changes in the amplitude of low frequency spontaneous fluctuations in the Blood-Oxygen-Level Dependent Imaging (BOLD) signal in the whole brain. fALFF represents the relative contribution of low frequency oscillations to the total detectable frequency range and is calculated taking the power within a frequency range and dividing it by the total power in the whole detectable frequency range [413]. For that, first the voxel time series are transformed into the power domain, then the amplitudes within a specific low frequency bandwidth are summed. Finally, this value is divided by the sum of the amplitudes in the entire detectable frequency range. To create standardized maps, each subject maps is transformed into Z-scores. We calculated the fALFF using REST toolbox [343]. Subject-specific fALFF maps were taken to the second level analysis in SPM12.

Independent component analysis (ICA)

ICA a data-driven technique to extract whole brain networks. We examined the resting-state networks given by the spatial grouping of voxels with temporally coherent activity calculated in a data-driven fashion using ICA. ICA decomposes blindly, in multiple independent components, the brain activity. Each component is a spatial grouping of voxels with temporally coherent activity (connectivity) and according to this spatial grouping of voxels, the components are associated with sources that can be either related to brain activity or to noise such as movement, blinking, breathing, and heartbeat. The main brain activity-related sources resemble discrete cortical functional networks and are named Resting State Networks (RSN). The RSN comprise the Default Mode Network (DMN), basal ganglia, auditory, visuospatial, sensory-motor, salience, executive control, language and visual networks. Here ICA was performed in GIFT software [1] using Infomax algorithm. The number of spatially independent resting-state networks (N) was estimated by the GIFT software (N=26). The identification of the networks was done automatically using predefined GIFT templates and later the resting-state networks of interested, the DMN and visual networks, were chosen by two specialists. For every resting-network of interest, subject-specific spatial connectivity maps were taken to the second level analysis in SPM12.

Seed-based Functional Connectivity (SeedFC)

Functional Connectivity (FC) is one of the most popular methods to infer connectivity in neuroimaging. When FC is calculated by means of the temporal correlations (Pearson's correlation) between a region of interest to the other voxels in the whole brain, it is known as Seed-based Functional Connectivity (SeedFC). In this way, SeedFC calculates the brain network related to specific regions of interest. We investigated the seed-based connectivity maps, using as seed the brain regions of interest in VR, namely, bilateral superior and middle frontal cortex, bilateral hippocampus, bilateral superior and inferior parietal cortex, areas shown in previous works that may be explain differences between VR-related presence experiences and less immersive environments [30, 31, 35, 46, 86, 194]. The seed areas were defined using the Anatomical Automatic Labelling (AAL) atlas. Seed-based connectivity maps were obtained by correlating the seed timeserie with all voxels in the brain. For that, first we extracted the timeserie of all voxels within the corresponding region of interest delimitated by the AAL atlas and then we took the average. Next, we calculated the Pearson's correlation coefficient between the seed timeserie and all other voxels in the brain. Finally, Fischer transformation were applied to the individual FC maps obtaining Z scores to improve normality. The individual SeedFC maps were calculated in MATLAB R2012b [4]. The Z score maps were taken to the second level in SPM 12 [388].

Graph analysis

To examine differences in the topology of the brain networks we performed graph analysis [63]. The first step was to construct the functional connectivity matrices, where nodes and links should be defined. Nodes were brain regions created based on the AAL116 atlas [369] and the links were the connectivity strength between nodes calculated using Pearson's correlation coefficient. The node-averaged time series used to infer the connectivity strength were extracted for each subject using the REST toolbox [343]. To avoid false positive links, connectivity values that were not statistically significant (p-value $\geq .05$) were excluded. Once the functional matrices are built, graph analysis can be applied in order to characterize their topology. At this stage, thresholding was applied, namely density threshold ranging from 0.1 to 0.8 in steps of 0.1. Thresholding means that only links with the highest connectivity strengths are kept until the desired density is reached, e.g. a threshold of 0.1 means 10% of the links with the highest connectivity were kept and the remaining ones were set to 0. Graph analysis were then applied to these thresholded matrices using the Brain Connectivity toolbox (brain-connectivity-toolbox.net). The main graph measures were chosen: betweenness centrality that measures the fraction of all shortest paths that pass through an individual node; characteristic path length which accounts for the average shortest path between all pairs of nodes; efficiency which is the average inverse shortest paths and transitivity that measures the relative number of triangles in the graph, compared to total number of connected triples of nodes. For a complete description of the graph measures please refer to [63, 300].

Statistics

We were interested in two particular contrasts: (1) VR visual stimulation using MRI compatible goggles with 3D stereoscopic stimulation and 2D monoscopic presentation vs. screen via mirror projection, to investigate for brain differences during goggles and screen, (2) 3D stereoscopic stimulation, in which the image is rendered separately for each eye creating the illusion of depth and 3D effect vs. 2D monoscopic presentation, most commonly applied presentation in which both eyes receive the same visual input, so that we could investigate the effect of stereoscopic view in the brain. The resulting individual maps of each analysis were taken to the second level analysis in SPM12 [388] with mean FD as covariate. Using a Family Wise Error (FWE) threshold on the cluster level of $p < .05$ we ran the two contrasts in both the positive and the negative direction. For the graph analysis, repeated measures one-way ANOVA was used. FDR were used for multiple comparison correction.

2.3 Results

Since we were mostly interested in the direction of potential brain connectivity differences between conditions, we employed multiple different analysis pipelines. We employed four methods that focus in different aspects of the brain signal, the intrinsic frequency fluctuation and the connectivity given by the spatial grouping of voxels with temporally coherent activity and by the temporal correlations between areas, respectively, fALFF that measures the amplitude of the low frequency fluctuation of the BOLD signal, ICA that uncovers the resting-state brain networks, the seed-based FC that calculates the brain network related to specific regions of interest and graph analysis that characterizes the topology of the brain networks.

2.3.1 fALFF

In fALFF we found significantly higher fALFF (Figure 2.2) in left cerebellum (VI, -20, -62, -26), left postcentral gyrus (-42, -40 60) in the goggles (monoscopic+stereoscopic) condition compared with the screen condition. In the reverse contrast we observed higher fALFF in right superior orbital frontal cortex (-18, 56, -4) (all results family wise error corrected $p<.05$).

2.3.2 ICA

In the visual networks (Figure 2.3), we found increased in connectivity in the primary and higher networks in the left calcarine (-8, -94, 10) and left lingual (-4, -64, 4), respectively, in the goggles (monoscopic+stereoscopic) as compared to screen condition, that is, when using goggles, the grouping of voxels in left calcarine and left lingual were more strongly correlated to the signal of the source identified, respectively, as primary and higher visual network A decreased in connectivity was observed in the left middle occipital (6, -90, 30) in the primary visual network and in the bilateral cuneus (0, -78, 24) in the higher visual one, which means, when using goggles, the grouping of voxels in left middle occipital and in the bilateral cuneus were less correlated to the signal of the sources identified, respectively, as primary visual and higher visual network. Investigating the DMN (Figure 2.3), we found an increase in the connectivity in the inferior frontal (-34, 0, 28) and bilateral lingual (-2, -66, 0) for goggles (monoscopic+stereoscopic) as compared to screen condition, meaning stronger correlation of these areas and the signal of the source identified as DMN. No significant difference between the contrast monoscopic vs. stereoscopic was found in the visual networks nor in the DMN. FDR was used to account for multiple comparison correction due to multiple networks.

2.3.3 SeedFC

In a ROI-based seed analysis we used the following ROIs: bilateral superior frontal cortex, middle frontal cortex, hippocampus, superior parietal cortex, inferior parietal cortex. We found significant seed-based connectivity (Figure 2.4) between bilateral superior frontal cortex and the left superior temporal lobe (-48, 16, -12) for the MRI goggle contrast goggles (monoscopic+stereoscopic) > screen and to left insula and putamen (-34, 10, 10) in the stereoscopic view contrast (stereoscopic > monoscopic; Figure 2.4). The bilateral inferior parietal cortex was more strongly connected to right calcarine cortex (18, -98, 4; 20, -88, 2) and right lingual cortex (26, -88, -6) in the goggles vs. screen condition (Figure 2.5). All Seed-based ROI analysis results were family wise error corrected at $p < .05$. To account for multiple comparison correction due to multiple seeds, FDR was used.

2.3.4 Graph Analysis

Despite no topological differences consistent across threshold were found (Figure 2.6), the threshold 0.1 representing 10% of the highest connections, goggles condition presented higher mean betweenness than screen, as well as transitivity and characteristic path length. Betweenness is measure of centrality, transitivity of segregation and characteristic path length a measure of integration. This means that, for networks formed by the highest connectivity links during VR local information processing (transitivity) was higher, whereas, at the same time, the global exchange of information (characteristic path length) in the brain was also increased. In addition, a higher mean betweenness revealed more central nodes in the network. The local efficiency of the network was also higher during VR as compared to screen, when considering the 30% highest connectivity values as nodes. These results reported above were not corrected for multiple comparison. No significant differences were found after correcting for multiple comparison.

2.4 Discussion

Within the scope of the present study we set out to unravel the effects of VR stimulation presented via goggles on functional brain connectivity in MRI. In order to do so, we used our FloVR game, in which the player had the task to fly through a natural scene with the goal to make the landscape blossom. This was designed with the goal to decrease negative affect and induce positive affect. To disentangle the effects of presenting the visual stimuli via MRI-compatible goggles, we compared the goggle stereoscopic and monoscopic condition with the screen condition. We considered that the goggles, which are mounted on the user's head and have the ability to display stereoscopic images, are objectively more immersive than a back-projection of a Screen via a mirror system [338] and therefore can elicit the sense of presence [123]. As in the back-projection of a screen, the participant receives only 2D images and can still see the scanner bore and oftentimes even the staff operating the scanner next to the presented stimuli. In order to investigate differences in functional brain connectivity during gaming induced by the stereoscopic view, namely the fact that the image is rendered separately for each eye creating the illusion of depth and 3D effect, we contrasted the stereoscopic and monoscopic condition directly. With the goal to obtain a comprehensive picture of brain connectivity we chose four common approaches to analyze resting state functional Magnetic Resonance Imaging (fMRI) data, namely the assessment of the amplitude of low frequency fluctuation (fALFF [413]), independent component analysis [66], seed-based functional connectivity analysis and graph analysis [63].

In line with our hypothesis, the goggles and the stereoscopic contrast revealed stronger brain connectivity for the respective condition with more immersion in the technical [338] and subjective [123] point of view, that is goggles (stereoscopic and monoscopic) compared to screen and stereoscopic compared to monoscopic generally elicited higher brain connectivity. We found higher fALFF in left cerebellum and postcentral gyrus for goggles compared to the screen. In the visual networks, we found an increase in connectivity in the left calcarine and left lingual for the same contrast, meaning higher correlation of the grouping of voxels in these areas to the signal associated with visual networks and in the DMN there was increased connectivity in the inferior frontal cortex and bilateral lingual gyrus, which means lower correlation of these areas and the signal associated with the DMN activity. Additionally, in the seed-based analysis we found higher connectivity between bilateral superior frontal cortex and the temporal lobe, as well as bilateral inferior parietal cortex with right calcarine and right lingual cortex for the two goggle vs. screen conditions. Furthermore, we found superior frontal cortex and insula/putamen to be more strongly connected in stereoscopic vs. monoscopic view. When looking at the screen condition we found higher brain activity, that is higher fALFF values, in right superior orbital frontal cortex.

These results can be viewed as in line with the hypothesis proposed by Jäncke and colleagues [170] stating that prefrontal cortex is involved in the experience of presence. In particular, bilateral Dorsolateral Prefrontal Cortex (DLPFC) activity was shown to be negatively correlated with the subjective report of the experience of presence in adults [30]. Moreover, when focusing on connectivity between brain areas, the authors report the results of an effective connectivity analysis from which they conclude that the right DLPFC is involved in down-regulating the activation in the dorsal visual processing stream. Furthermore the authors interpret the observed increase of activity in the dorsal visual stream during presence experience as a sign of higher action preparation in the virtual reality because the brain responds to it similarly as in real-life situations [170]. However, on the same dataset the left DLPFC was shown to be positively connected to brain regions that are part of the default-mode network (such as medial prefrontal cortex, anterior cingulate cortex, thalamus, brain stem, nucleus caudatus and parahippocampus). Due to the involvement of the latter brain regions in self-referential processing the interpretation is that higher left DLPFC activation when participants experience less presence leads to an up-regulation of self-referential processing which reflects the detachment from the VR experience.

In contrast to previous studies, in which the focus was on the subjective feeling of presence and brain activity, we set out to investigate differences in brain connectivity between objectively different conditions of stimulation during gaming. A major disadvantage of the previous design, with exception of one [123], was that the stimulation used to elicit different degrees of presence was not the same. In addition, in all of them, the participants were only passively watching the displays. For this reason, we confronted participants with the same interactive VR game in all three conditions with the only difference being the hardware presentation procedures used to display the respective environment.

Our results can be viewed as being in line with the findings and interpretations shown in association to perceived experience of presence [30, 170], since we also find more fALFF in right superior orbital frontal cortex in the 2D screen condition compared to the two goggle conditions. Next to these previous fMRI results our results can also be perceived as in line with results from an EEG study in which the same spatial navigation task in a virtual maze was compared between a projections onto a large wall which was supposedly more immersive than a display on a small Desktop PC screen [194]. The authors report stronger functional connectivity between frontal and parietal brain regions in the Desktop display condition. When comes to comparing only stereoscopic and monoscopic condition, our results differ to the one from Gaebler and colleagues [123] in terms of brain areas. This fact cannot being seen as a surprise considering that they investigated areas that showed common brain activity across subjects by means of the intersubject correlations, which were the temporal lobe, right inferior occipital cortex and right precuneus, whereas we focused on the connectivity between brain areas. Interestingly, the direction of the contrast where differences were found were the same in Gaebler and colleagues and in the present study: stereoscopic > monoscopic, meaning that stereoscopic elicited higher intersubject correlations as well as higher connectivity as compared to monoscopic.

However, the rationale for the present study was slightly different from the studies presented before. We set out to test whether overall the stereoscopic VR presentation elicits higher degrees of functional brain connectivity than monoscopic and a screen display.

Our hypothesis was that the condition that elicits the most brain connectivity should be most suited for long-term brain training interventions, assuming that extended training under these conditions could permanently improve brain connectivity on a functional as well as a structural level. Our results show that the majority of contrasts and functional connectivity indicators resulting from different analysis pipelines reveal higher brain connectivity between brain regions in the goggle condition and the stereoscopic condition in particular, which we interpret as a hint that training in VR environments in contrast to environments displayed on a screen may be superior in eliciting and therewith facilitating brain connectivity in intervention studies.

At present, a major drawback of the implementation of VR in an MRI environment is the fact that the VR experience is limited to the stereoscopic input to the eyes without the experience of movement in space. Usually the stereoscopic view in VR is accompanied by the fact that individuals can freely turn their head and move in space while the visual input is adapted to the individual's movements. However, since the head cannot be freely moved in the MRI scanner, due to its resulting movement artefacts in the images, the actual differences in brain connectivity between a VR and a screen presentation of an environment might be underestimated. Future research may attempt to use motion tracking systems to enable this movement related visual feedback, while at the same time correct for the occurring motion artefacts in the acquired MRI images [352]. The limited field of view and resolution of the MRI compatible goggles introduces yet another limitation to such studies. As it can affect the level of immersion for both monoscopic and stereoscopic conditions. Specifically, to the design of the present study, the question still remains of whether and how frequently the participants noticed the 3D effect when, for example, the bee was not flying close enough to the virtual objects.

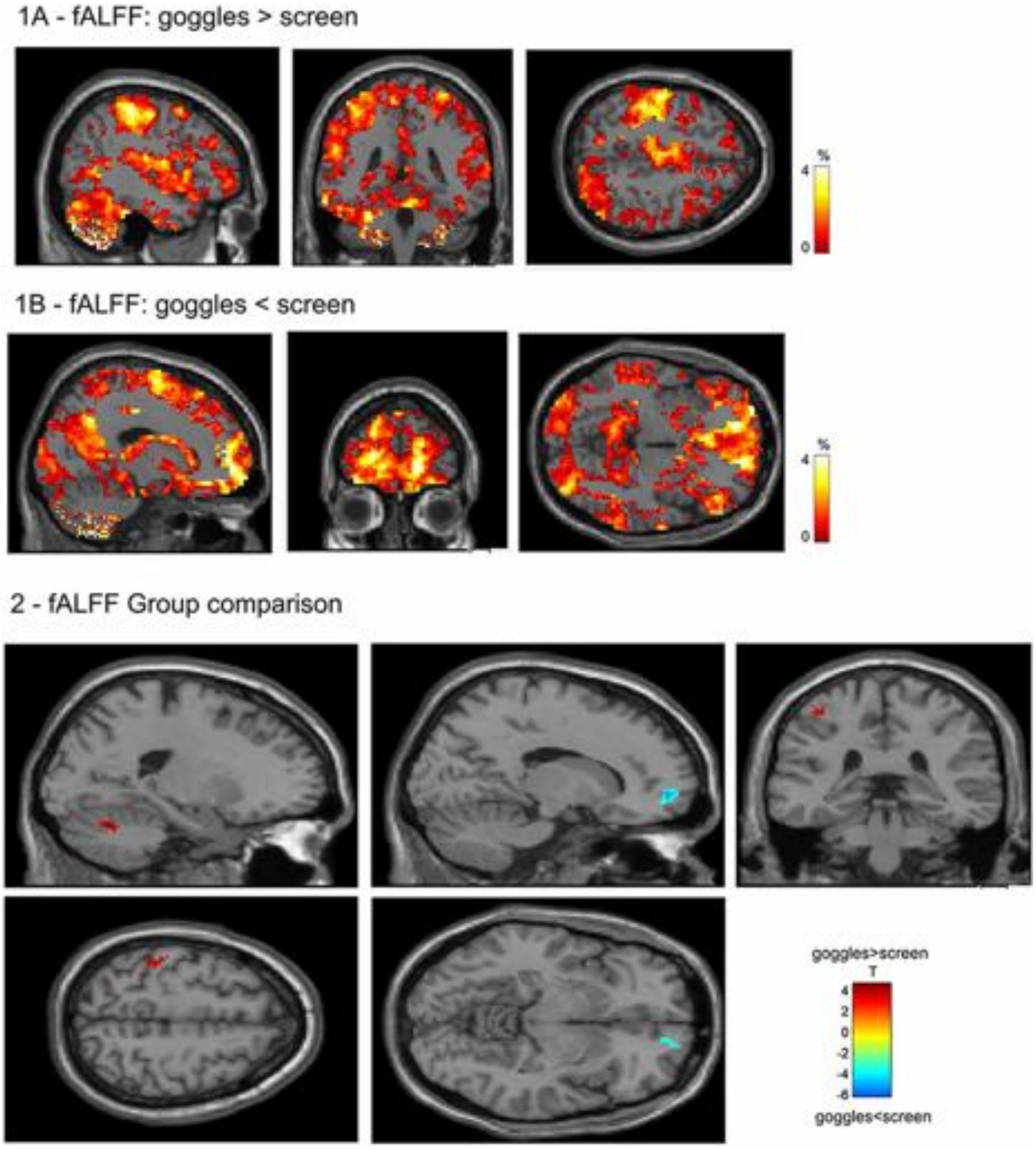


Figure 2.2: 1—Difference of the mean fALFF maps, 1A goggles > screen and 1B goggles < screen. 2—Group differences in fALFF. Higher fALFF (in red) in left cerebellum and left post-central gyrus in the goggles (monoscopic + stereoscopic) condition compared with the screen condition. In the reverse contrast (in blue) there was higher fALFF in right superior orbital frontal cortex.

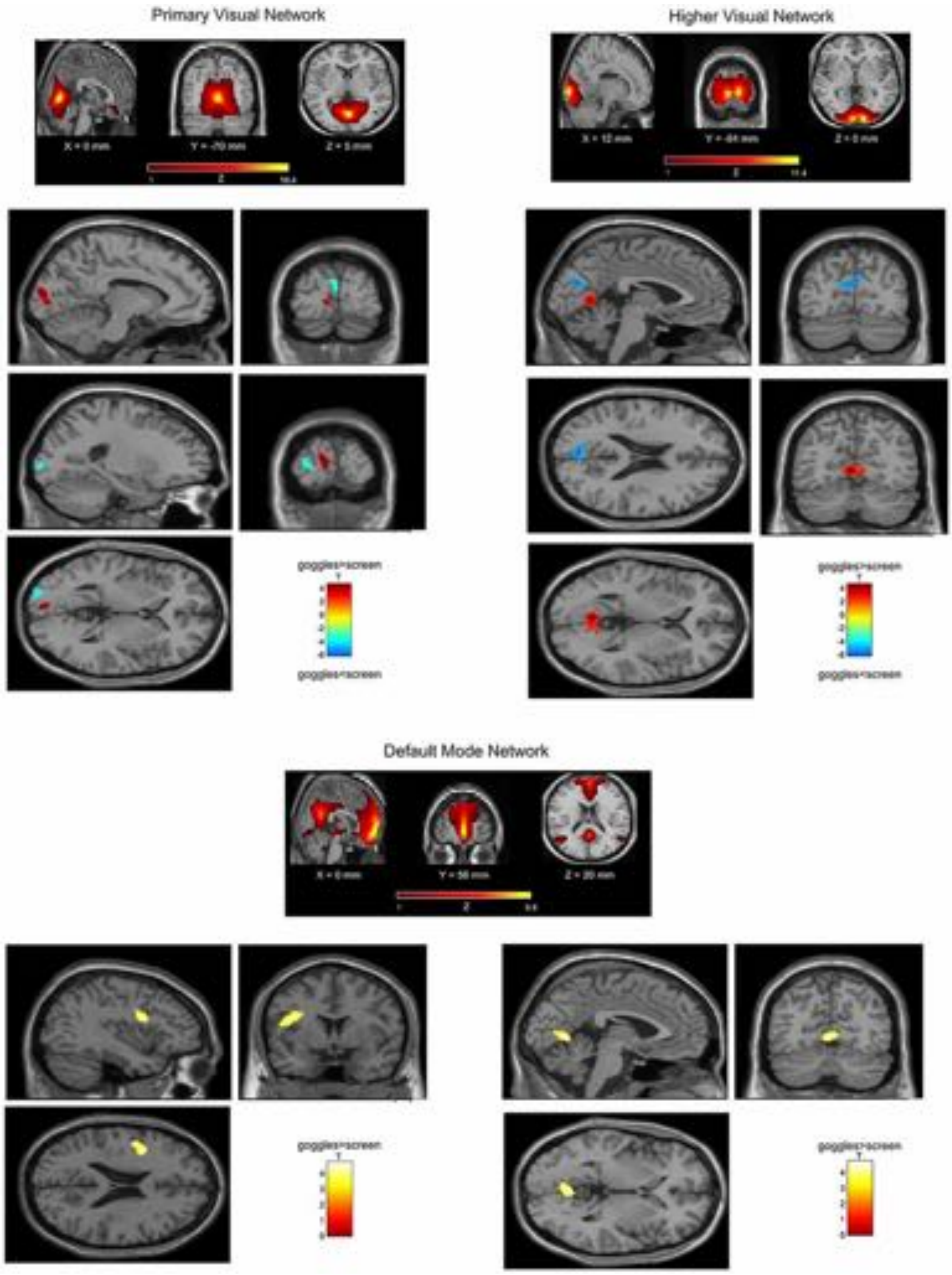


Figure 2.3: Network spatial maps and group differences: DMN and visual. In the primary visual network, there was an increase in connectivity (in red) in the left calcarine and in the higher visual network in the left lingual in the goggles (monoscopic + stereoscopic) condition as compared to the screen condition. A decreased in connectivity (in blue) was seen in the left middle occipital in the primary visual network and in the bilateral cuneus in the higher visual network. In the DMN, we found an increase (in yellow) in the connectivity in the inferior frontal and bilateral lingual for goggles (monoscopic + stereoscopic) as compared to screen condition.

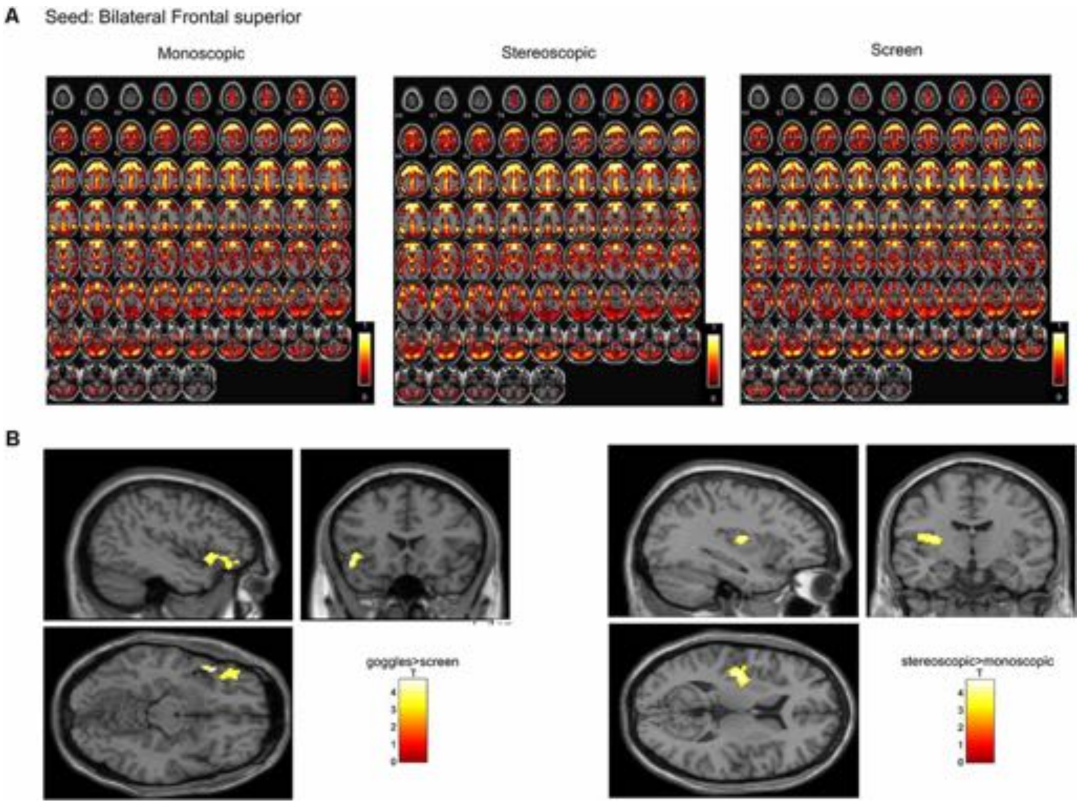


Figure 2.4: (A) Mean SeedFC maps per condition, seed located in the bilateral frontal superior cortex. (B) Left—Group differences in SeedFC in goggles vs. screen condition. There was stronger connectivity between a seed in the bilateral superior frontal cortex and the left superior temporal lobe for the MRI goggle contrast goggles (monoscopic + stereoscopic) as compared to the screen. (B) Right—Group differences in stereoscopic vs. monoscopic condition. The stereoscopic view elicited stronger connectivity between the bilateral frontal cortex and left insula and putamen.

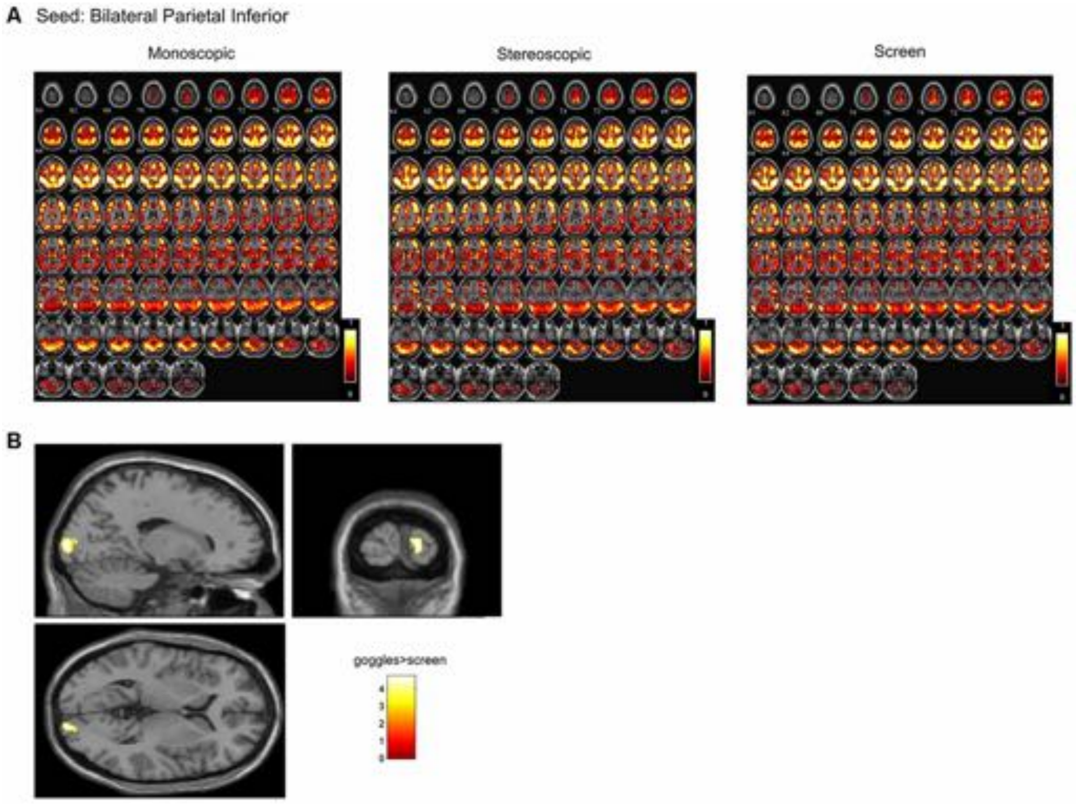


Figure 2.5: (A) Mean SeedFC maps per condition, seed located in the bilateral inferior parietal cortex. (B) Group differences in SeedFC in goggles vs. screen condition. There was stronger connectivity between a seed in the bilateral parietal inferior cortex to right calcarine cortex and to right lingual cortex for the MRI goggle contrast goggles (monoscopic + stereoscopic) as compared to the screen.

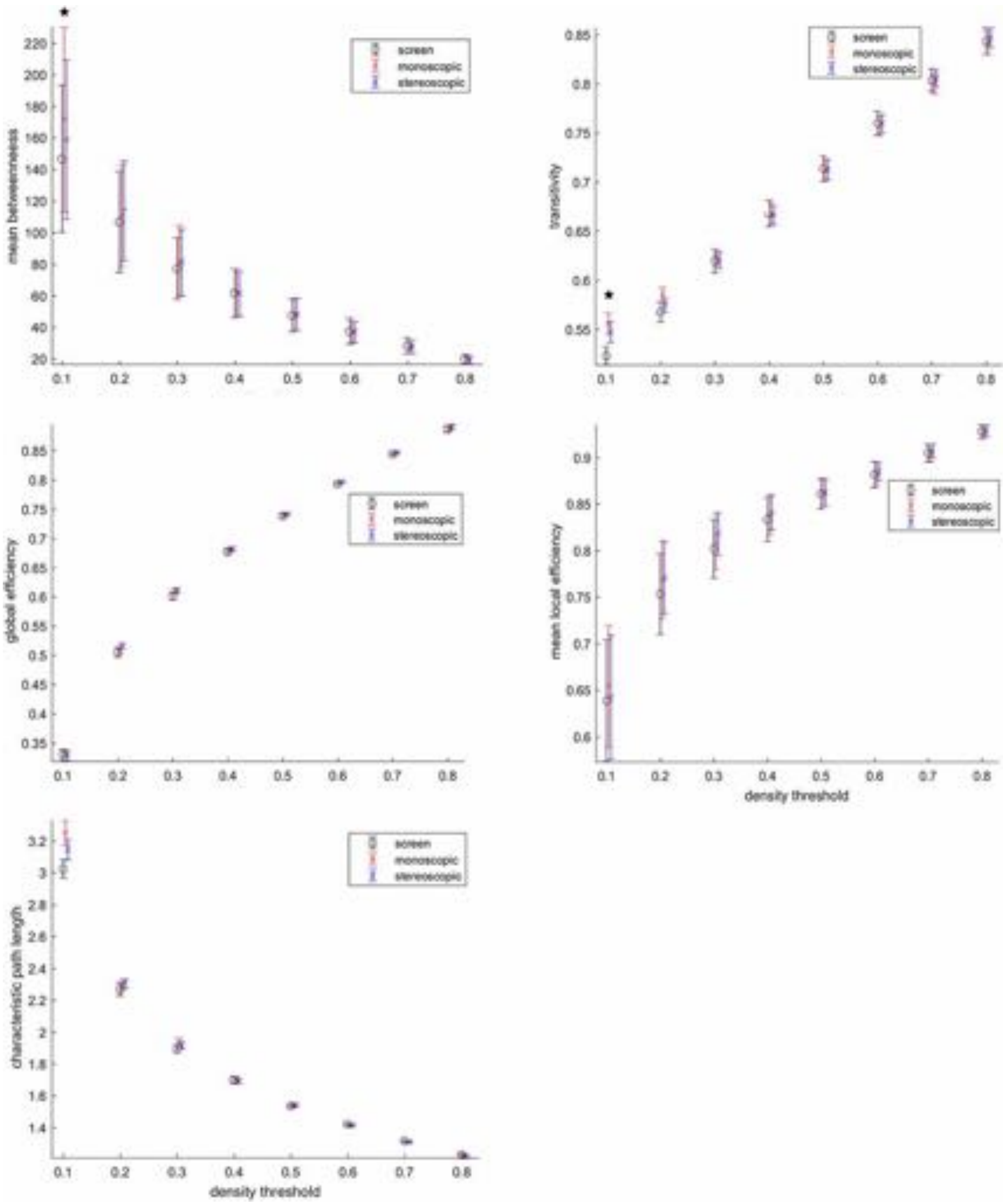


Figure 2.6: Graph analysis. Betweenneess, transitivity, global efficiency, local efficiency and characteristic path length were calculated. Means group differences before multiple comparison correction ($p < .05$ uncorrected). Group differences were not significant after multiple comparison correction.

Chapter 3

3 Improving Depressive Mood with Immersive Virtual Reality Gaming

Video games and VR applications have great potentials for creating alternative and more accessible treatment options for mood-related disorders. Additionally, VR induces a higher sense of presence compared to mobile gaming devices. Whether this higher sense of presence leads to stronger impacts of gaming on depressive mood was the main research question of this chapter. Therefore, we assessed the effects of gaming using VR or a tablet device on depressivity, anxiety, affect, flow, and sense of presence. The results suggest that in comparison to a control and the tablet condition, gaming using VR could significantly reduce depressivity and negative affect. Moreover, VR could increase the positive affect, yet not significantly different from the tablet and control conditions. The sense of flow and presence were, additionally, higher in the VR condition.

3.1 Introduction

Video games have great potentials for mental health interventions, not only because they are engaging and effective [108], but also since they are popular and millions of people play video games worldwide [232]. In the US alone, over 164 million adults equivalent to 65% of American adults play video games [21]. Playing video games is not limited to any specific age, gender, or ethnicity [21]. The rise of affordable smartphone games has also made almost every consumer a potential gamer [232]. In fact, smartphones are the most common devices used for video game play and casual games are the most popular game genres [21]. The video game players believe that games have positive effects, such as relaxation and stress relief on their mental health [21]. Research supports potential benefits of games and gamification for mental health as well [271, 108].

In addition, the potential of using VR for therapeutic interventions has already been shown by numerous studies [293, 75]. However, research ozon the application of VR for treatment of mood disorders and depressivity has been relatively disregarded [117]. Only a few studies have studied this topic, which nonetheless have shown promises for its symptom relief [220].

From a pragmatic point of view, it is important to determine whether platform's level of immersion and its associated feeling of presence are decisive for the extent of gaming effect on mood and depressivity. Yet, such comparison has not been made so far and it is still unknown whether a mobile device such as a tablet suffice or an immersive display such as a VR HMD is needed to achieve the full impact of gaming.

To address this gap, we developed FloVR, a VR-based game designed to induce positive mood and reduce depressivity, and in addition, optimized it for a tablet device. Eighteen participants underwent three conditions: (i) playing FloVR in an immersive VR setup, (ii) playing it on a tablet device, and (iii) reading an article as a control condition. Hereby, we pursued the answers to the following research questions: (Q1) What are the effects of playing FloVR on mood, anxiety, and depressivity? (Q2) Can VR induce a higher sense of presence and flow compared to the tablet and control conditions? If so, (Q3) are the effects on mood, anxiety, and depressivity stronger when the game is played using the VR HMD setup?

3.2 Related Work

Mood improvement can have a long-lasting positive effect on mental health [135]. The so-called mood-congruence effect occurs, when people memorize or recall information in a distorted manner, influenced by their current mood [396]. Being in a negative or positive mood increases the tendency to memorize and recall respectively negative or positive information. This effect is compatible with the Beck's cognitive theory of depression [34], which states that depression is underlying dysfunctional basic assumptions and thought patterns (e.g., "I have to be perfect."), that lead to a negative biased perception of reality. People in a depressed mood memorize and recall depressing memories more likely, thus develop and maintain negative cognitive schemata and a negative perspective on reality [167]. This repeated evocation of mood-congruent memories strengthens the preservation of depressivity.

In this context, video games provide a promising platform for mood improvement (or mood repair) and reduction of anxiety and depression-related symptoms [80, 193, 290, 382, 289, 285, 55, 305, 302, 160]. Indeed, an online survey with 1614 participants [284] suggest that video games are being systematically used for stress recovery. The results of another study that conducted 71 interviews with male and female online gamers from 11 different countries revealed that video games are used to alleviate negative feelings [162]. Also, a qualitative analysis of an online survey [164] showed that gaming helps players relieve stress and deal with their feelings by allowing them to work through their emotions and engage in relatable experiences.

Furthermore, Russoniello and colleagues [307, 306, 304, 305, 106, 107] conducted a series of studies and confirmed the positive effects of playing video games on the psycho- and physiological states of the participants. For instance, they tested the effects of playing video games on mood and stress [307, 306] and reported on consistent brain activities (e.g., decreased left alpha brain waves associated with a decrease in depressive-type behaviors) with increased mood and consistent heart rate variability changes with autonomic nervous system relaxation or decreased physical stress. In addition, they observed significant reduction of clinical depression symptoms [304] as well as reduction of state and trait anxiety symptom severity [106]. These results were obtained through a prescribed regimen of playing video games for at least 30 minutes, three times per week, and for one month. Moreover, they showed, in another study [305], that a single 30-minute session of playing a video game was significantly more effective in reducing treatment-resistant depression symptoms when compared to a second antidepressant medication.

Playing video games can also facilitate achieving a state of flow [54, 290, 193, 80, 107] which has a particularly positive influence on mood and is associated with eudaimonic well-being. The latter is defined as a feeling of autonomy, personal growth, self-acceptance, mastery, positive relatedness, and life purpose [121]. To experience flow, one must be involved in a task, in which the level of difficulty matches the individual's skills for managing the task [84, 83]. These conditions result in an intense and focused concentration on the present moment, a merge of action and awareness, a loss of reflective self-consciousness, a sense of control, a distortion of temporal experience and intrinsic reward [84]. Since some characteristics of video games (e.g., clear goals associated with tasks, presence of immediate feedback, and progress through the game) align with the ones that could induce flow, their players could potentially achieve a state of flow [54, 290, 193, 80, 271]. Once achieved, their attention is dedicated to the game which could prohibit concurrent experience of anxiety to occur [107, 83].

Furthermore, multiple studies have shown that VR-based interventions can reduce depressivity [263, 117, 375, 221]. For instance, Falconer et al. [99] were able to reduce depressivity by increasing participants' self-compassion in VR. Participants were instructed to behave compassionately towards a virtual child in an immersive VR scenario. The participants were then embodied into that virtual child to receive their own recorded compassion afterward. Also, Shah et al. [323] developed a VR stress management program. The program combined psycho-education and VR-based relaxation practices, which led to a significant reduction of depression and anxiety in participants. Additionally, Banos et al. [27] implemented a Cognitive Behavioural Therapy (CBT) program in their VR-world EMMA. They found statistically significant improvements in relaxation and depression compared to a traditional CBT. Furthermore, a couple of guided meditation VR applications have emerged in the past years. For instance, Guided Meditation VR by Cubicle Ninjas [85] or DEEP by Monobanda [245] are designed to reduce users' stress, anxiety and discomfort levels. Research on these applications substantiates positive effects on mental health (e.g. [269, 378, 387]).

However, it is still unknown whether the level of immersion that a gaming setup can deliver, specifically a common mobile gaming device vs. an immersive VR setup, plays a role in the effectiveness of video games for reducing depressivity and improving mood. A comparison between playing a video game in VR versus on a desktop display in terms of experienced flow, immersion, positive emotions, and psychological needs (e.g., challenge and competence) has been made by Pallavicini et al. [260]. Their results showed that playing in VR induced more intense positive emotions and a higher sense of immersion and flow compared to the desktop condition. The performance on the game as well as the the fulfillment of the players' psychological needs (such as challenge) were independent of the display device.

3.3 FloVR

Inspired by Thatgamecompany's game Flower [362], we developed FloVR, a VR-based game designed to induce positive mood and reduce depressivity. In this game, players fly through a nature landscape and make it blossom by pollinating flowers. Visual and auditory elements associated with the positive affect were implemented. FloVR (see Figure 3.1) was developed using the Unity 3D game engine. Since nature has been proven to induce positive feelings, reduce stress, and improve mental health [58, 29], a grassland scene was implemented for the game. The predominantly blue and green colors of the scene elements are also associated with pleasure and neutral levels of arousal [374]. Furthermore, the surface of the terrain contains various smooth and gentle hills, which are associated with gentle and playful moods [274].

(a) (b)

Figure 3.1: Screen-shots of the user's view while flying above the virtual landscape. The colorful halos indicate the positions of the next flowers to be pollinated. Small leaves and dandelions are circling around the user's view.

In FloVR, players fly through this nature scene with the goal of making the landscape blossom. They achieve this by flying near the flowers to virtually pollinate them, which causes more flowers to grow in the surrounding area. The flowers are ordered in such way, that they altogether build a path through the terrain. To highlight them visually and add variety to the scene, the flowers were given different, bright colors (blue, red, orange, yellow, purple and white) and were surrounded by a halo of the same color. Small dandelion seeds come out of the flowers when pollinated and small leaves start to circle around the players' view. The leaves continue circling throughout the game to symbolize the flowers are already pollinated, representing another type of positive feedback to support a feeling of flow.

Background music was included in the game, since it can have a positive impact on emotions [379]. Even though there are individual differences in the perception of music, certain structural elements can have an impact on the emotional response. Especially mode, tempo, pitch and harmony of a song can be crucial elements for its emotional perception [379]. As a major mode is associated with a positive mood [161, 379], the song had to be in a major mode. Similarly, it had to have higher pitches and simple harmonies [291]. As a fast tempo is associated with a higher arousal [161, 102] but can be perceived as uneasy [291], the song had to be rather slow. To find the right soundtrack for the game, three songs were selected based on these criteria and evaluated in a pilot study. Nine participants (mean age=30, SD=14.22, 5 female) had to listen to each song on three different days at around the same time of the day using their phone and fill out the State Trait Anxiety Depression Inventory-State (STADI-S) (see Questionnaires) before and after. The song Silver Blue Light by Kevin MacLeod [1] resulted in the highest reduction of anxiety and depression (measured by the STADI-S) and was thus used for the game.

Finally, in order to examine the role of immersion on mood, we implemented the game for two platforms: PC to be displayed on an HTC Vive HMD (i.e., VR condition) and an Android tablet device. The flight direction in the VE was indicated by either HMD or tablet orientation. The speed of flying was set to low (19.8km/h) for both devices. These simple controls were chosen to make the conditions as comparable as possible and to avoid simulator sickness [186].

3.4 Experiment

We conducted an experiment with three conditions (FloVR on HMD, FloVR on a tablet, and reading a text on a tablet). These conditions were chosen because they differ in their levels of immersion and the sense of presence that they evoke. Thereby, the following hypotheses were formulated:

- H1: playing FloVR, regardless of the type of the medium on which it is played (HMD or tablet), can improve mood and temporarily reduce depression compared to a control condition.
- H2: playing FloVR in VR can provide a higher sense of presence compared to the tablet and the control conditions.
- H3: playing FloVR in VR, due to its higher sense of presence, can improve mood and reduce depressivity more than playing it on a tablet.

3.4.1 Materials

For the VR condition, the HTC Vive HMD and an Intel computer (GForce GTX 780 Ti, Intel Core i7 4930, 16GB RAM) were used, which could render the game at more than 90fps. In the tablet condition, the game was played on a Samsung Galaxy Tab S3 tablet with Android Nougat 7.0. To increase the level of immersion in the VR condition, the Bose QuietComfort 15 noise-canceling headphones were used. As the speakers of the Samsung Galaxy Tab S3 have a lower sound quality than the Bose Headphones, the UE Boom by Logitech was used to play the sound in the tablet condition to decrease the differences in the level of sound quality. Both conditions were played with the same volume of about 50dB. The text of the control

[1] http://incompetech.com

condition was also presented on the Galaxy Tab S3. Its font was set to Times New Roman. The line spacing was set to 1.5 to facilitate finding the next line. The default font size was set to 11 pt.

Participants could choose to be tested at either UKE or the Computer Science Department of Universität Hamburg. They were not allowed to switch between the two locations. The testing rooms of both institutes were of average size (about $14m^2$) and contained few tables, chairs, computers and the camera stations of the HTC Vive. The rooms were quiet and distraction-free. The participants were seated in a revolving chair throughout the whole experiment. The questionnaires were presented to the participant on a computer. The brightness of the screens was kept constant throughout the whole experiment.

Questionnaires

Becks Depression Inventory II (BDI-II) [147]

The inventory includes 21 questions regarding symptoms of depression over the past two weeks. Each question has 4 answers which are ordered in their intensity and evaluated on a scale ranging from 0 to 3. Higher scores indicate more intense symptoms. The BDI-II has a decent test-retest reliability for non-clinical samples and is sensitive to changes [146]. The questionnaire had a good internal consistency ($\alpha = .84$).

State Trait Anxiety Depression Inventory (STADI) [208]

The inventory includes a state and a trait questionnaire, each containing 20 affective statements. Subjects can indicate how much they agree with each statement on a 4-point Likert scale scoring from 1 (Not at all) to 4 (Very much so). Both questionnaires assess the affects Euthymia, Dysthymia, Emotionality and Worry with 5 questions each. These are used to assess the intensity of the higher-order affects Depression, described by Dysthymia and Anhedonia (inverted values of Euthymia) and Anxiety, described by Emotionality and Worry. The state questionnaire is used to measures current feelings of depression and anxiety whilst the trait questionnaire assesses them as more long-living personality traits. This relatively new inventory has demonstrated validity [57]. The internal validity was acceptable for Anxiety ($\alpha = .77$) and good for Depressivity ($\alpha = .81$).

Positive Affect Negative Affect Schedule-Expanded Form (PANAS-X) [386, 296]

The schedule includes 60 different adjectives of affect, whose intensity can be described on a 5-point Likert scale from 1 (Not at all) to 5 (Extremely). The participants were instructed to answer these questions in regard to their current affective state. The questionnaire was used to assess the subjects' current mood, through the variables Positive Affectivity and Negative Affectivity. Even though the term affectivity also refers to emotions, Positive Affectivity and Negative Affectivity are predominantly assessing mood [386]. Emotional states are measured by the lower-order affects Fear, Sadness, Guilt, Hostility, Shyness, Fatigue, Surprise, Joviality, Self-assurance, Attentiveness and Serenity. Studies report a good test-retest reliability [386].

The internal consistency of Positive Affectivity was good ($\alpha = .86$) and acceptable for Negative Affectivity ($\alpha = .73$).

Flow Short Scale (FSS) [287]

The scale includes seven questions regarding components of the flow-experience which can be expressed on a ranging 0 (does not apply) to 6 (does apply). Three questions assess the worry that one could have about failing at accomplishing the task. In addition, three 9-point scales evaluate the self-estimated skill level, difficulty level of the task, and feeling of being overwhelmed. The scale has been validated by several studies [288, 97, 98] and showed a good internal validity of flow ($\alpha = .85$).

Igroup Presence Questionnaire (IPQ) [316]

It contains 14 items on a 7-point Likert scale ranging from 0 to 6 with different scale anchor, meaning that some items have general scale anchors (0: Fully disagree to 6: Fully agree) and some have more precise anchors (e.g., 0: Not consistent and 6: Very consistent). The questionnaire has also four sub-scales: general presence or the sense of being there, spatial presence, involvement and experienced realism. A confirmatory factor analysis has supported the questionnaire's validity [316, 317]. The internal validity was excellent ($\alpha = .92$).

Simulator Sickness Questionnaire (SSQ) [186]

The questionnaire allows subjects to indicate the intensity of 16 symptoms regarding simulator sickness on a 4-point Likert scale from 1 (None) to 4 (Severe). The questionnaire is a standard questionnaire to measure simulator sickness and its validity has been shown [186]. Its internal validity was acceptable ($\alpha = .77$).

The questionnaires were handed out before and after each condition, except of the FSS and IPQ, which were only completed after each condition. All questionnaires were presented in the German version.

3.4.2 Participants

The study was approved by the local ethical committee of the Department of Informatics at the Universität Hamburg. Recruitment was done through a local job market website and an email distributor among the students of the Department. To apply for the study, three online questionnaires had to be filled out: 1) the Mini International Neuropsychiatric Interview (MINI) [155], which is a short Structured Clinical Interview SCID for Diagnostic and Statistical Manual of Mental Disorders, 4th Edition (DSM-IV) and the ICD-10 psychiatric disorders., 2) BDI-II and 3) the trait part of the State-Trait-Anxiety-Depression-Inventory (STADI-T).

Applicants who had an acute Axis I disorder or acute alcohol or drug problems, according to the results of the MINI, were excluded. Furthermore, only participants with slightly elevated BDI-II scores (≥ 4) were tested, to make sure their level of depression could be improved after all. Out of 290 applicants, 43 fulfilled these criteria of which 19 took part in the study. On the first day of the experiment, participants were asked whether they ever had any kind of mental disorder or took psychoactive medication. One participant gave a positive answer and was therefore excluded. The remaining 18 participants were nine females and nine males between the ages of 18 and 55 (M=25.9, SD=10.5). 17 participants were Caucasian and one Turkish. Only one participant had a BDI-II score of 17 (mild depression), the others were below a score of 11 (no to minimal depression). Two participants were taking a contraceptive pill and one participant took L-Thyroxin, a drug against hypothyroidism. Eight of the subjects were computer science students. Four obtained class credit for their participation and the other 15 were paid 25 Euro. Six of the 18 participants had not worn an HMD before. Seven participants almost never played 3D video games, while four played on a monthly basis, three on a weekly basis and four on a daily basis. Six wore glasses or contact lenses and one subject had a red-green color blindness. The color blindness might have had reduced the positive effect of the green colors in the game. However, since this affects both the tablet and the VR condition, it should not have favored any of the hypotheses.

All 18 participants managed to control the game and were able to follow the path of the flowers in both the VR and tablet conditions. They had an average BDI-II score of 6.67 ($min = 4$, $max = 17$, $SD = 3.36$) when applying for the study. The BDI-II mean was lower when assessed by the BDI-II questionnaires given prior to every condition during the actual experiment, however, this difference was not significant ($p = .07$). Throughout all three conditions, the pre-mean of the BDI-II was at 5.22 ($min = 0$, $max = 17$, $SD = 4.26$), with some participants having a BDI-II score lower than the targeted value of at least 4. As assessed by the STADI-T, the mean value of Trait Depression was 11 ($min = 5$, $max = 18$, $SD = 3.86$) and the mean value of Trait Anxiety was 15 ($min = 10$, $max = 29$, $SD = 5.07$).

3.4.3 Procedure

For each of the three conditions participants were tested on a separate day, with the first and last day being no more than two weeks apart. They were allowed to choose the days and the time they wanted to be tested (between 7am and 8pm), however, it had to be the same time for each of the three days.

On the first test-day written informed consent was obtained from all participants. The subjects were informed that the study was interested in the subject's feelings when playing the game and they were provided with a summary of the experiment's procedure. In addition, participants were asked about their demographics, visual disorders, gaming-habits, current medication and possible psychiatric diagnoses. At the end of the last test day, the compensation was paid and detailed information about the study's purpose was provided.

After entering the room and sitting down, the participants were asked to fill out the STADI-S, PANAS-X, BDI and SSQ questionnaires on a computer screen. Thereafter, they either read a glider flight report about gliding over the Frisian coast [2] on a tablet (control condition),

[2] published on https://www.femo-design.de

played FloVR on the tablet (tablet condition) or played FloVR in VR using an HMD (VR condition). A within-subject design was chosen for the study, therefore each participant underwent all three conditions. The order of the conditions was counterbalanced across participants.

Before playing the game in the VR and tablet condition, the participants were given information about the gameplay. They were asked to imagine to be a bee, whose goal is to make the landscape flourish by flying over flowers or trees which are surrounded by a bright halo. That would lead to pollen popping out of the flowers and new flowers, bushes or trees growing all around. They were told to follow the path of the flowers and collect as many flowers as possible, however, they should not mind when missing out a flower. To make sure the participants would not struggle to control the game, the experimenter gave a short demonstration of the controls. In the VR condition, this was simply done by showing how to rotate the head when they wanted to fly in a certain direction. In the tablet condition, participants were told to imagine an arrow that sticks out of the back of the tablet, which indicates the flying direction. Therefore, they would have to rotate the back of the tablet toward the direction they wanted to fly. This could be done by either using their wrists or by moving the tablet within a spherical radius around them. To make sure their arm muscles would not be strained, they were able to place their arms on the armrest or their upper body, except when they were flying upwards. In both the VR and tablet condition, they were allowed to use the revolving chair for the sideways rotation.

After 5:50 minutes both the reading and the gaming time were over. This time was set, as the song of the game lasted exactly that long and as the exposure to nature of about 5 minutes has been shown to result in a rather optimal mood improvement [29]. Furthermore, prolonged exposure to VR may have increased the risk of simulator sickness [186]. Participants would automatically stop flying and a text would show up, asking them to contact the experimenter. By then, the song was over and participants had also reached the last flower due to the continuous flying-speed. In the reading condition, the experimenter instructed them to finish their current sentence and stop reading. Finally, again the STADI-S, PANAS-X, BDI and SSQ were filled out. Additionally, the FSS and the IPQ were completed. When filling out the IPQ for the reading condition, the participants were asked to answer the questions in regard to how well they could delve into the story of the flight report. After completing the last questionnaire, the experiment day was over.

3.5 Results

In the first place, the conditions were examined independently of each other to test for significant changes in those variables that were measured before and after each condition. As some of the variables were not normally distributed according to the Shapiro-Wilk-Test, the Wilcoxon signed-rank test was used to compare the pre- and post measurements. Thereafter, difference scores were calculated and were used to compare all three conditions. For this purpose, the Friedman test was applied on the difference scores or the values of the variables

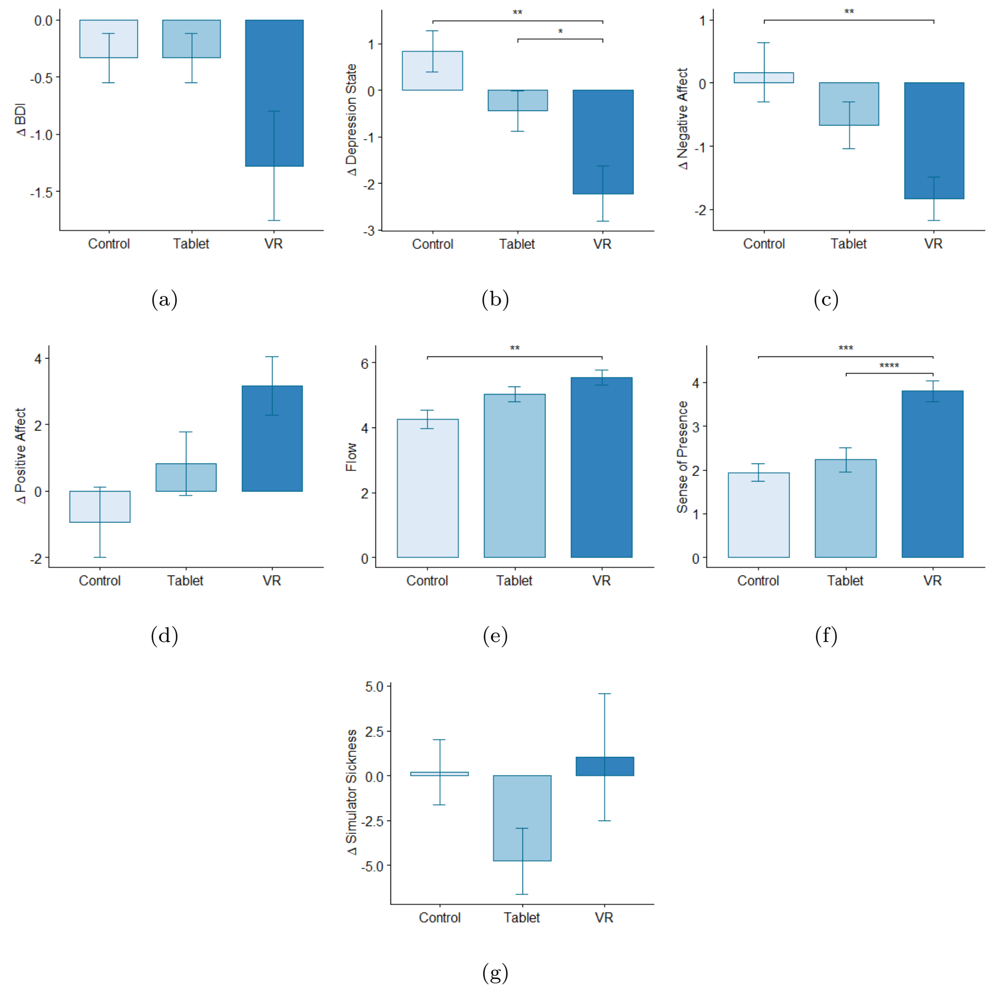


Figure 3.2: Results of the self-assessment questionnaires: (a-d) Difference scores (post-pre) of (a) BDI-II (b) State Depression measured by STADI-S (c-d) Negative and Positive Affects measured by PANAS-X (e) FSS post measurements of Flow (f) IPQ post measurements of the Sense of Presence

that were measured only once and after each condition (e.g., Flow and Sense of Presence). To compare the conditions, pairwise Wilcoxon signed-rank tests with Bonferroni adjustment method were used. Significance was set at an α level of .05. Finally, carry-over effects were assessed by comparing the three possible positions in which the respective condition had been tested, using the Kruskal–Wallis test.

3.5.1 BDI-II

The results of the Wilcoxon signed-rank test suggested that only VR condition could significantly reduce the participants' depression symptoms ($M = -1.28, SD = 2.02, p = .018$). The pre- and post measurements of BDI-II were not significantly different for the Tablet and Control conditions. The Friedman test on the pre- and post difference scores did not show any significant differences between all three conditions ($\chi^2 = 4.844, p = .089$). These scores are shown in Figure 3.2(a). Also, no evidence of carry-over effect could be observed for the BDI-II scores.

3.5.2 STADI-S

The VR condition significantly decreased Depression State ($M = -2.22, SD = 2.51, p = .004$) and its lower-order factor Anhedonia ($M = -1.78, SD = 2.13, p = .007$) as well as Anxiety State ($M = -2.28, SD = 3.34, p = .014$) and its lower-order factor Worry ($M = -1.94, SD = 1.92, p = .004$). The Control condition and the Tablet condition led to significantly decreased values of Anxiety State (Control: $M = -2.11, SD = 2.54, p = .005$/Tablet: $M = -1.5, SD = 2.31, p = .022$) and its lower-order factor Worry (Control: $M = -1.5, SD = 2.12, p = .012$/Tablet: $M = -1.28, SD = 1.84, p = .013$).

Significant differences in the Friedman-Test on the pre- and post difference scores of all three conditions were found for Depression State ($\chi^2 = 18.98, p < .0001$). Pairwise comparisons with Bonferroni correction suggested that the VR condition had significantly higher reductions in Depression State compared to both the Control ($p = .007$) and the Tablet ($p = .035$) condition (see Figure 3.2(b)). Moreover, Anhedonia was the only lower-order effect of the STADI-S that was found to have significant pre-post-differences between the conditions ($\chi^2 = 18.72, p < .0001$). According to the pairwise comparisons VR led to greater reductions of Anhedonia compared to the Control ($p = .012$) and the Tablet ($p = .017$) conditions. Finally, no carry-over effect could be observed for the STADI-S scores.

3.5.3 PANAS-X

Comparing the pre- and post measurements revealed that the VR condition significantly reduced the participants' General Negative Affect ($M = -1.83, SD = 1.47, p = .0006$) including their feelings of Fear ($M = -1.17, SD = 1.25, p = .003$), Hostility ($M = -0.44, SD = .78, p = .026$), Sadness ($M = -.89, SD = 2, p = .037$), and Fatigue ($M = -1.61, SD = 2.73, p = .022$). It increased, on the other hand, their General Positive Affect ($M = 3.17, SD = 3.78, p = .005$) and its associated feelings of Joviality ($M = 3.72, SD = 4.11, p = .004$) and Surprise ($M = 3.94, SD = 3.72, p = .002$). Whereas no significant changes of General Negative and Positive Affect from pre- to post measurements could be found for the Control and Tablet conditions. Only two lower-order affects of the General Positive Affect showed significant changes from pre- to post measurements, in which Self-assurance was reduced by the Control condition ($M = -1.33, SD = 2.35, p = .037$) and Surprise was increased by both Control ($M = 1.39, SD = 2.38, p = .014$) and Tablet ($M = 1.44, SD = 2.68, p = .035$) conditions.

The Friedman test showed significant differences between all three conditions on their pre-post difference scores for the General Negative Affect ($\chi^2 = 11.43, p = .003$), its lower-order factor Hostility ($\chi^2 = 6.5, p = .039$), and the General Positive Affect ($\chi^2 = 6.79, p = .034$). The pairwise comparisons with Bonferroni adjustment method revealed that the VR condition was more effective in reducing the General Negative Affect ($p = .006$) compared to the Control condition (see Figure 3.2(c) and 3.2(d)). Lastly, no carry-over effect could be found for the PANAS-X scores.

With these results, the null hypothesis of the first hypothesis (H1) cannot be rejected since only playing FloVR in VR led to a significant improvement of mood and depressivity.

3.5.4 FSS

Levels of Flow were significantly different between the three conditions ($\chi^2 = 8.82, p = .012$). Pairwise comparisons suggested that the VR condition caused higher levels of Flow (see Figure 3.2(e)) compared to the Control condition ($p = .004$). Also, no carry over effect was observed. Additionally, Flow was negatively correlated with the Depression State in both the Tablet ($p = .007$, $r = -.61$) and the VR ($p = .013$, $r = -.57$) conditions. These results support partially our second hypothesis (H2) since Flow was significantly higher only compared to the Control but not the Tablet Condition.

3.5.5 IPQ

As expected, values of the Sense of Presence were found to be significantly different between the conditions ($\chi^2 = 21.33, p < .0001$). The VR condition resulted in significantly higher Sense of Presence (see Figure 5.5) compared to both the Tablet ($p = .00002$) and the Control condition ($p = .0002$). These results support the second hypothesis (H2). The Tablet and the Control conditions were not significantly different ($p = 1$). Also, no carry-over effect was observed.

3.5.6 SSQ

Comparing the pre- and post measurements of the symptoms of Simulator Sickness revealed that the Tablet condition was the only condition that significantly reduced Simulator Sickness ($p = .032$). Both the Control and the VR conditions did not significantly impact Simulator Sickness. Changes of the Simulator Sickness (see Figure 5.4) were not significantly different between the conditions ($\chi^2 = 2.98, p = .225$). No carry-over effects were observed.

3.6 Discussion

Previous research has successfully utilized VR for different types of anxiety disorders. However, a meta-analysis revealed, that VR therapy for mood disorders and depressivity has been relatively disregarded [117]. To investigate the question of whether VR can be efficiently used for improving depressive mood, the game FloVR was developed and designed to improve mood and reduce depressivity. Subsequently, FloVR's impact on affectivity was tested in VR and on a tablet. The evaluation indicates, that FloVR can be successful in improving overall affectivity and in temporarily reducing depressivity.

It is noticeable, that VR has the highest values in reducing Depressivity, State Depression, State Anxiety, and Negative Affectivity as well as in increasing Positive Affectivity. Furthermore, it has the highest values of Flow and Sense of Presence. For each of these variables, the VR condition is followed by the tablet condition with the second-highest values. The control condition resulted in the lowest values.

Consequently, significant differences between the conditions support the assumption, that the VR condition is most effective in improving overall affectivity, followed by the tablet condition and lastly the control condition. Regarding the Simulator Sickness, the tablet condition led to the highest reduction, followed by the control condition and lastly the VR condition.

The VR condition led to significant improvements in all questionnaires regarding affectivity and resulted in significantly higher improvements in variables that assessed negative affectivity and depressivity compared to the control condition. This finding indicates that VR can be efficiently utilized to improve depressive mood. Apart from State Anxiety, the tablet and the control conditions did not improve any dependent affectivity variable, such as State Depression, Negative Affectivity or Positive Affectivity. Furthermore, no significant difference could be found between the tablet and control condition in any of the questionnaires. Thus, playing FloVR on a tablet does not seem to offer any advantages compared to reading the text. As expected, the VR condition had significantly higher values in Sense of Presence when compared to the tablet and the control condition, which is likely caused by its higher level of immersion. Thus, the positive elements implemented into the game do not seem to be sufficient to actually improve depressive mood but only seem to work in combination with an increased sense of presence.

In comparison to the tablet condition, the VR condition had significantly better results in the reduction of State Depression. In particular, the lower-order factor Anhedonia was significantly reduced. The VR condition did not lead to significantly bigger increases in Positive Affectivity or decreases in Negative Affectivity compared to the tablet condition. However, whilst the VR condition was significant in every affect-questionnaire, the tablet condition only produced significant improvement in State Anxiety. Bearing the above in mind, the VR seems to be more effective for mood improvement than the tablet, although not fully supported by the results of the PANAS-X. Since State Depression was strongly negatively correlated with Flow but not with Presence in both the VR and the tablet condition, the game's positive impact on depressivity seems to be predominantly related to the experience of flow.

As all conditions have similarly decreased values in State Anxiety, factors other than differences between the conditions might have been the reason for this phenomenon. The conditions led to significant decreases in Worry, however not in Emotionality. Thus, the decrease could be due to a reduction in test anxiety. Regardless of the main influencing factor, neither the VR game nor the tablet game seem to be particularly more suitable for reducing anxiety than the control condition.

Furthermore, VR significantly improved the values of Depressivity measured by the BDI-II, which assesses symptoms of depression over the last two weeks. The BDI-II was used to have an indication about whether participants would rate their past weeks differently after the intervention (i.e., mood congruence effect). It should be noted, however, that these differences cannot be directly related to the Mood Congruence Effect and are no more than indicative. In addition, differences in Depressivity between the VR and the control condition were not significant. Even if the VR condition might have affected the emotional evaluation of the past weeks, this effect does not appear to be excessively strong compared to the other conditions. Therefore, no statement can be made about the intensity and specificity of this effect.

3.6.1 Limitations

Even though the tablet and the VR conditions were designed to be as comparable as possible, some limitations might have influenced the results. Firstly, some participants might have been able to predict the hypothesis of the study. Participants were informed that the study is interested in their feelings when playing the game. In combination with the experiment design, the study's interest in improving affectivity with VR might have been revealed. This in turn could have influenced participants to answer the mood questionnaires in favor of the hypothesis. Especially improvements in the BDI-II questionnaire, which assesses mood over the past two weeks, could be an indicator of demand characteristics rather than the Mood Congruence Effect. However, this should also have led to higher reductions in acute anxiety in the VR condition compared to the other conditions. It is unlikely, that participants distorted their responses on depressivity but not on anxiety, as both mood disorders are related [396] and the study did not reveal its particular interest in depression.

As the sound was played from different devices, the different sound quality could have also had an impact on the results. The qualities of the speaker and the headphone were broadly comparable but still, the headphones quality was probably higher. The decision to use different devices was made, as noise-canceling headphones are highly immersive from the auditory perspective. They filter out ambient noises so that hardly any external auditory stimuli can disturb the listener. As this study is interested in the impact of immersion, the tablet condition had to have a non-immersive sound in order to serve as a non-immersive comparison.

The controls followed the same logic, however, some participants found the controls of the tablet a little more difficult than those of VR. Due to the different nature of the devices, these discrepancies could hardly be avoided. In fact, the higher level of immersion in the VR condition might have been a reason for reporting easier and more intuitive controls. The resulting increase in difficulty might have been overwhelming to the subjects and left a negative impact on their mood. However, the self-estimated difficulty of the conditions and self-estimated feelings of being overwhelmed, both measured by the FSS, were not significantly different between the conditions. Additionally, State Anxiety similarly decreased in both conditions, which contradicts a feeling of being overwhelmed in the tablet condition. Furthermore, all participants managed to control the game and follow the path of the flowers with both devices. Because subjects had to hold the tablet in their hands, this burden on their arms and hands could have also influenced the results. However, they were able to rest their arms on the upper body if not flying upwards and no participant complained to have been overly strained.

The statistical approach also need to be addressed. First, the sample size of 18 participants is quite small which led to a relatively low power of the study. Second, the use of multiple statistical tests has increased the chance of rejecting true null hypotheses (Type I errors), which is partly be counteracted by the conservative p-value correction (Bonferroni) that was used. Furthermore, the study has only investigated correlations between the three conditions and participants' changes in affectivity, so no causality can be inferred.

Lastly, given that most of the participants were under 30 years of age and of Western European origin, the results cannot necessarily be generalised to older people or those of other ethnicities.

3.6.2 Future Work

As the study tested short-term effects with healthy participants, more research has to be conducted to ascertain whether VR can efficiently be utilized for depression therapy. The effects of using FloVR on a daily or weekly basis should be investigated in the future. Participants might be able to keep their elevated mood or improve it throughout time. In this case, the hazard of becoming addicted must be investigated. However, they could also get used to the game and the positive effect might disappear. To avoid this, the game could be extended, with different levels to keep up the subject's interest in the game.

Furthermore, subjects affected by depression might react differently to the VR game than healthy participants. Studies have shown that individuals affected by a major depressive disorder are failing to show a positive bias when primed by a stimulus with a positive valence [218]. Participants might be less affected by the positive elements of the game and thus it might be more difficult to improve their mood. On the other hand, most of the participants that took part in this study were already in a rather good mood, with one of them even having a BDI-II score of 0. As it might be more difficult to improve the mood of participants that are already happy, a study with depressive subjects could also lead to better results in mood improvement.

Additionally, the game could be customized for the subjects. For example, the concepts of VR exposure therapy could be transferred to VR depression therapy. Objects that are related to the subject's cause of depression could be embedded into the game. This could help them approach adverse memories and experience them in a new context to eventually overcome them. Lastly, the game needs to be compared to other mood-improvement techniques, such as reading or imagining an emotional story. Some subjects might find it really difficult to delve into their imagination, thus should be more affected by VR.

3.7 Conclusion

To conclude, although this research is limited by a small sample size, the study suggests that FloVR can be used successfully for improving depressive mood. A short exposure time of about six minutes led to an improved mood and reduced feelings of depression. Especially when compared to the control condition, the results show that the VR condition was more effective. These differences are not as clear between the VR and the tablet condition, however, they still indicate that VR can be more effective. Whether the VR game is actually suitable and efficient for application in clinical contexts is not answered by this study. An evaluation of the game by the people diagnosed with mood disorders as well as its long-term effects remains a topic for further research. This study represents one of the first VR studies on improving mental well-being with a focus on improving mood and reducing depressivity. It substantiates that it is worthwhile to further invest in research on VR mood disorders therapy.

4

Chapter 4

Effects of Exposure to Immersive Videos and Photo Slideshows of Forest and Urban Environments

A large number of studies have demonstrated the benefits of natural environments on people's health and well-being. For people who have limited access to nature (e.g., elderly in nursing homes, hospital patients, or jail inmates), virtual representations may provide an alternative to benefit from the illusion of a natural environment. For this purpose and in most previous studies, conventional photos of nature have been used. Immersive VR environments, however, can induce a higher sense of presence compared to conventional photos. Whether this higher sense of presence leads to increased positive impacts of virtual nature exposure is the main research question of this chapter. Therefore, we compared exposure to a forest and an urban virtual environment in terms of their respective impact on mood, stress, physiological reactions, and cognition. The environments were presented via an HMD as (i) conventional photo slideshows or (ii) 360° videos. The results show that the forest environment had a positive effect on cognition and the urban environment disturbed mood regardless of the mode of presentation. In addition, photos of either urban or forest environment were both more effective in reducing physiological arousal compared to immersive 360° videos.

4.1 Introduction

Nowadays, individuals spend more and more time in artificially designed living spaces, in particular, humans spend up to 90 percent indoors [3]. This tendency has led to an isolation of individuals from regular contact with nature which has a negative impact on their mental and physical health. Several studies have demonstrated that such artificial stimulation and being in purely human-generated environments can lead to mental fatigue as well as a loss of vitality and health [182, 350].

These negative effects can be reduced by means of engaging in interactions with nature [185].

There is evidence to suggest that natural environments have a positive influence on human psychology, physiology, and cognition [257, 349, 171]. According to the Attention Restoration Theory (ART), natural environments capture less cognitive resources, and therefore, allow an interruption of attention-grabbing tasks inherent in urban environments and thus, elicit attention restoration and recovery from mental fatigue [180, 144, 361]. Natural elements such as green landscapes and flowing waters have a calming effect on physiological arousal [188, 10]. One of the long-term effects of access to nature is a positive attitude towards life and an increased satisfaction with one's own home, one's own work and generally one's own life [180, 179].

As an instance of natural environments, forests have been studied frequently suggesting their positive effects on human body and mind [53, 169, 278, 262, 355, 16]. These positive effects include, but are not limited to psychological relief, lower stress and depression levels [327, 138, 231, 119, 16, 247] as well as physiological effects such as lower blood pressure, Heart Rate (HR), and salivary cortisol hormone levels [216, 215, 355]. Therefore, forest therapy, also referred to as "forest bathing", is practiced widely, in particular in Asia, to derive substantial benefits from the positive health effects of walking, resting, and interacting with forests [217, 255, 73, 210, 342, 187, 281, 211].

For people with limited access to nature (e.g., elderly in nursing homes, hospital patients, or jail inmates), already the visual representation of nature can relieve stress and improve emotional well-being [178, 219, 209, 231, 124, 367]. Many studies in environmental psychology have used conventional photos to compare natural and urban environments or to demonstrate the positive effects of nature photos [370, 203, 257, 349].

In this context, immersive VR may facilitate some of these characteristics such as the feeling of being in nature during the exposure [338]. By reproducing realistic stimuli and eliciting psychological processes, VR has the potential to increase external validity of the research findings [340]. It can, in addition, provide the experimenter (and potentially therapists) with a systematic control over the natural elements such as weather conditions, vegetation (up to the smallest details such as movements of the grass and leaves on the trees), wildlife, and lighting that is hard or impossible to achieve in real life [89, 62]. Furthermore, therapeutic applications may benefit from the low-cost virtual environments, which can be duplicated and distributed easily, making them usable at a larger scale [293] and make it accessible to individuals in need, e.g., in nursing homes. Thus, VR can complement the research on human perception and behavioral responses to nature stimuli by maximizing the benefits of lab-based (e.g., control over independent variable) and field-based (e.g., realistic stimuli) experiments [340].

For this reason, previous studies have already employed nature exposure in VR. Several studies have compared real physical nature exposure with exposure to 360° videos of nature [61, 259, 67, 330, 408, 72]. For instance, Browning et al. [61] compared real nature exposure and a 360° VR nature video recorded from the same location. In comparison to a physical indoor environment without nature, both real and VR nature exposure were more restorative and increased physiological arousal. Although, only the real exposure to nature outdoors increased mood in a positive direction.

Researchers have also compared exposure to different environments merely in VR. For instance, a study [385] demonstrated that different types of forest environments, presented via 360° videos, can improve mood and relieve stress. Another study [11] revealed that in comparison to a control environment, exposure to 360° videos of nature can reduce physiological arousal and negative affect. Furthermore, in a study by Chung et al. [74] and in comparison to 360° videos of fireworks, exposure to 360° videos of nature improved cognitive functioning and restored involuntary attention of the participants [74]. In comparison to urban environments, Yu et al. [410] could show that exposure to 360° videos of forest or waterfall environment was able to decrease negative emotions such as fatigue and depression. In contrast, levels of fatigue were increased and self-esteem was decreased after exposure to urban environments. Also, in a study by Schutte et al. [319] participants were exposed to a natural and an urban environment using 360° videos. Thereafter, participants reported significantly more restorativeness by exposure to the natural environment compared to the urban environment.

Multiple studies have reported stress recovery elicited by multisensory exposure to nature in VR. For instance, in a study by Annerstedt et al. [15], participants experienced a psychosocial stress (i.e., TSST [189]) in VR followed by an exposure to natural scenes in VR either with or without sound. As a result, recovery from stress was facilitated by exposure to VR nature and was enhanced when the environment was presented with natural sounds. In another study by Hedblom et al. [150] visual stimuli (i.e., 360° photos of urban, park, and forest environments) were accompanied by auditory stimuli (e.g., bird songs for natural environments) and olfactory stimuli (e.g., grass odour for park). Consequently, exposure to natural environments reduced stress levels significantly. Finally, Schebella et al. [312] suggested that multisensory exposure to 360° videos of nature are beneficial to recovery from stress compared to visual-only exposure and that recovery is least effective in a virtual urban environment.

In order to use visual representation of natural environments in experiments or for preventive and/or therapeutic purposes, it is important to know whether the level of immersion and its associated feeling of presence are decisive for the extent of the effect. Different levels of immersion could be, for example: the actual stay in a natural environment, viewing a natural environment through a window, viewing a 360° video of a natural environment (stereoscopic or monoscopic) on a display (such as a smartphone with integrated gyroscope or using an HMD), staying in an artificially generated virtual world or watching a regular video or pictures of nature. To subjectively distinguish between different levels of immersion in VR context, the sense of presence is usually measured. It describes the psychological sense of being in a virtual environment [338] and can have multiple components such as the sense of being physically present in a place (spatial presence), the attention devoted to the virtual environment (experienced involvement), as well as the experienced realism of the environment [316].

Previous studies have examined different levels of immersion based on human's psycho- and physiological responses [157, 266, 25, 123, 194, 111, 45, 364, 265, 71]. For instance, a study [157] suggested that closest to reality psychological responses can be achieved by a 360° panorama and physiological responses by a 3D model of the real environment (an interior shopping environment). Although in the same study, different levels of immersion including a conventional photograph of the environment were employed, they were all compared against the real environment and not against one another. In another study [25] a significant increase in the sense of presence from monoscopic to stereoscopic and from 180°to 360° images was demonstrated. In addition our group, Forlim, et al. [111] previously reported that stereoscopic renderings delivered via an HMD elicit higher functional connectivity in the brain when compared to monoscopic renderings on projection screens or HMDs. Furthermore, Chirico et al. [71] confirmed that immersive videos enhance the intensity of self-reported awe emotion as well as parasympathetic activation compared to 2D screen videos.

However, to the best of our knowledge, a direct comparison between conventional photo presentations and 360° video presentation of nature has not been tested so far. Hence, it is largely unknown whether conventional photo presentations suffice to create the full impact of virtual exposure to nature or whether an immersive display such as 360° video presentation can further increase the positive effects.

The presented experiment in this chapter followed a within-subject design and consisted of a control (in a silent black virtual room with a white screen in the middle showing a fixation cross) and four experimental conditions (see Figure 4.2). Each experimental condition consisted of three parts: (i) a cognitive test (serially subtracting 13 from a given starting number such as 1022) for five minutes, (ii) exposure for six minutes to either an urban (i.e., an old town of northern Germany, see Figure 4.1(b)) or a forest (i.e., a northern German mixed forest, see Figure 4.1(a)) virtual environment presented either using 360° videos or conventional photo slideshows from the same content both displayed via an HMD, and lastly (iii) filling out the questionnaires. The order of the experimental conditions was counterbalanced. The control condition was consistently administered at the beginning and immediately after the baseline measurement (see Methods). During the experiment physiological data was recorded, namely Galvanic Skin Response (GSR) and HR. It is worth noting that the cognitive test in this experiment served two functions simultaneously: it was used to induce stress prior to exposure and at the same time to measure cognitive performance due to the prior exposure phase. That is, the cognitive test measured the cognitive performance after exposure to the previous (and not following) condition. Thus, after the last condition, the cognitive test was administered for the last time measuring the cognitive performance after the last exposure.

4.2 Methods

4.2.1 Participants

Recruitment of participants took place via an email distributor among the students of the Faculty of Computer Science at the Universität Hamburg. In addition, the study was advertised on the campus of the University of Hamburg and the University of Applied Sciences in

(a) (b) (c)

Figure 4.1: Virtual environments: (a) forest environment (b) urban environment (c) control environment

Hamburg and a call in social networks. A total of 35 subjects participated in the study. However, one person had to be excluded due to deuteranopia (green blindness). The remaining 34 subjects (11 female) were between 21 to 34 years old ($M = 27.26, SD = 4.144$). The study was approved by the local psychological ethics committee of the Center for Psycho-social Medicine at the University Hospital Hamburg-Eppendorf and was carried out in accordance with relevant guidelines and regulations.

4.2.2 Materials

We selected a northern German mixed forest as the forest environment (see Figure 4.1(a)). Since our focus was on vegetation, other natural elements such as water and animals or humans were avoided and were not present in this environment. Our urban environment was an old town of northern Germany (see Figure 4.1(b)) which contained no vegetation nor animals or humans. Each 360° video was a six min video consisting of three two min single stationary videos. To shoot the individual videos, the tripod with the camera was moved six meters forward, measured from the center of the tripod. The result is a composition of three stationary single shots in which the tripod, per single shot, was placed firmly in one place. Looking at the final video created the impression of teleportation between these three shots.

In total, three different environments were created for the study. The first environment was a black room with a white screen in its center. In the middle of the white screen was a black fixation cross (Figure 4.1(c)). This environment had no background sound and was used as the control condition. The second environment was identical to the first one, with the difference that the slideshows were played on the screen. The virtual camera in these two environments was one meter away from the virtual screen. The 360° videos were played in the third environment on the inner side of a virtual sphere. Here, the virtual camera was placed in the center of the sphere to create the impression of being inside the 360° VR environments.

The experiment was conducted in a laboratory room. The participants were seated on a firm chair. The position of the chair was fixed to ensure a fixed position in the virtual environment. During the experiment, the subjects wore a HTC Vive Pro HMD as well as Neulog Pulse and Galvanic Skin Response (GSR) sensors. GSR or Skin Conductance Response (SCR) measures the amount of changes in electrical conductivity of the skin (in this study, at the finger of the non-dominant hand) when its glands produce ionic sweat in response to a given stimulus.

Thus, it is considered as an indicator of localized phasic arousal processes and has been interpreted as an indicator of stress in the literature [52]. Additionally, previous research in particular on TSST [189], suggest that both heart rate and heart rate variability can detect the physiological effects of stress on human participants. Therefore, we concluded that both measures apply as our cognitive test was taken from the TSST. In this study, we measured heart rate since it could successfully indicate physiological changes of performing multiple TSST in VR in a previous study [248]. For rendering, system control, and logging an Intel Computer (Core i7 6900K at 3.2 GHz) with an NVIDIA GeForce GTX 1080 graphics card was used. The questionnaires were completed on a MacBook Pro (Retina, 13 inches, model year: end of 2013).

The following questionnaires were employed in this study:

State Trait Anxiety Depression Inventory-State (STADI-S) [38] measures the current state of anxiety and depression of a person. See Chapter 3, Section 3.4.1 for more information about the questionnaire.

Profiles of Mood States (POMS) [322] was used to assess mood. For this purpose, a value for the total mood disturbance was determined. The questionnaire contains keywords and statements that describe different feelings and are scored on a 5-point Likert scale ranging from 1 (Not at all) to 5 (Extremely).

Short Stress State Questionnaire (SSSQ) [309, 152] records the status of engagement, distress and worry after a given task. The questionnaire consists of 24 items rated from 1 (Not at all) to 5 (Very much so).

Perceived Stress Scale (PSS) [78] measures the perception of stress. It contains 10 items ranging from 0 (Never) to 4 (Very often). As the original items refer to the situations during the past month in one's life, for the purpose of this study, the items were modified to measure the momentary perceived stress.

Igroup Presence Questionnaire (IPQ) [316] was used to measure the perceived sense of presence in VR. See Chapter 3, Section 3.4.1 for more information about the questionnaire.

Simulator Sickness Questionnaire (SSQ) [186] measures 16 symptoms that may occur during or after VR exposure. See Chapter 3, Section3.4.1 for more information about the questionnaire.

Procedure

Upon arrival, the subjects were informed about the purpose of the study and their right to interrupt or quit at any time. Thereafter, they signed the informed consent. At baseline (see Figure 4.2), the SSQ-pre was administered and the baseline GSR and HR were measured. For this purpose, the subjects wore the HMD and saw the control environment (i.e., a black virtual room with a white screen showing a fixation cross) for six min. Participants were asked to look at the fixation cross on the white screen and not to move or to speak.

The study had a within-subject design and consisted of a control and four experimental conditions. The order of the experimental conditions was counterbalanced but the control condition was consistently administered at the beginning and immediately after the baseline

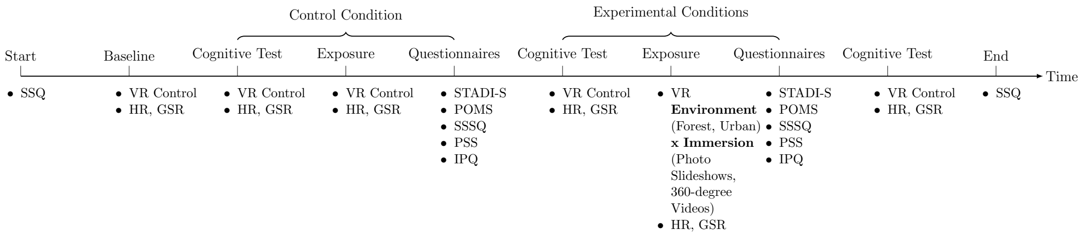


Figure 4.2: Experimental Procedure. VR: Virtual Reality (i.e., the contents were displayed using a head-mounted display); Control: a silent black virtual room with a white screen in the middle showing a fixation cross; SSQ: Simulator Sickness Questionnaire; HR: Heart Rate; GSR: Galvanic Skin Response or Skin Conductance Response (SCR); STADI-S: State Trait Anxiety Depression Inventory-State; POMS: Profiles of Mood States; SSSQ: Short Stress State Questionnaire; PSS: Perceived Stress Scale; IPQ: Igroup Presence Questionnaire.

measurement. Each condition consisted of three parts.

In the first part, the participants were asked to serially subtract 13 from a given starting number (e.g., 1022) while wearing the HMD and seeing the control environment. They were told to generate correct, sequential answers as fast as possible. They did not receive any feedback for giving the correct answers but immediately after giving an incorrect answer they were asked to start from the beginning. This task was taken from the Trier Social Stress Test (TSST) [189], a widely used protocol for inducing moderate psycho-social stress in laboratory settings. TSST consists of a five-minute free speech followed by another five minutes performing a mental arithmetic task. We employed the TSST's mental arithmetic task as the first part of our conditions. We measured the number of correct and incorrect answers, the total number of answers as well as HR and GSR values. This part lasted for five min. When treating the performance in this task as the dependent variable we compared the cognitive performances after the respective exposure to one of the four experimental conditions.

In the second part, the participants were exposed for six min to either an urban or a forest virtual environment presented either using 360° videos or conventional photos taken from the same video in the form of a slideshow. In the control condition, the presented virtual environment did not change during the exposure. In other words, after performing the mental arithmetic task inside the control environment during the first part, the subjects were exposed further to the control environment and did not have any specific task to do. They were, however, allowed to look around in the virtual environment and were instructed not to speak or move. In this part, GSR and HR were measured.

In the third part and after the exposure, the participants took off the HMD and the sensors and filled out the following questionnaires: STADI-S, POMS, SSSQ, PSS, IPQ.

It is important to clarify that the cognitive test in this experiment was designed to test the participants' cognitive performance after exposure to the previous (and not following) virtual environment. For this reason, after the last condition, the cognitive test was administered for the last time. Thereafter the participants filled out the SSQ-post and were compensated with course credits.

4.2.3 Data analysis

Prior to the analysis the physiological signals were smoothed using a low-pass Butterworth filter with the cutoff frequency of 1 Hz. Thereafter, they were normalized using the following formula [230, 149]:

$$\tilde{S} = \frac{s - min(\bar{s})}{max(\bar{s}) - min(\bar{s})} \tag{4.1}$$

Where S is the raw signal, s the smoothed version of S, $\bar{s}$ the signal taken over the entire session, and $\tilde{S}$ the normalized signal.

Subsequently, difference scores were used, with the average values measured during the cognitive test being subtracted from the values of the exposure phase. On these difference scores a two-way repeated-measures ANOVA was performed.

In order to analyse the cognitive performance as well as the questionnaire data, the difference scores were calculated by subtracting the respective value from the control condition from the values of the given experimental subjective or cognitive measure. These differences were not calculated for the IPQ responses as we were interested in the original presence values after each condition and not the changes with respect to the control condition.

It has to be noted that according to the Shapiro-Wilk test, some data were normally distributed (e.g., IPQ Sense of Presence as well as GSR and HR Difference scores) and some were not (e.g., Cognitive Test). Therefore, we decided to report the analysis based on parametric tests in order to not switch between statistical tests. Thus, to test our hypotheses, for all data except the SSQ, two-way repeated-measures ANOVAs were performed, with a significance level at .05. As an effect strength, the partial eta squared (η^2) was reported. Thereby, a value of .01 represents a small effect, .06 a medium effect, and .14 a large effect [77, 204]. The SSQ responses were analysed using a paired t-test with a significance level of .05. Additionally, Cohen's d was reported as the effect size for t-test which is commonly interpreted as small ($d = .2$), medium ($d = .5$), and large ($d = .8$) effects [77].

4.3 Results

To analyse the responses of our mood and stress-related questionnaires (i.e., STADI-S, POMS, SSSQ, and PSS) as well as the cognitive test, differences between experimental conditions and the control measurements were computed. On these difference scores, a two (environment: forest, urban) by two (immersion: 360° videos, conventional photo slideshows) repeated-measures ANOVA was performed.

4.3.1 Questionnaire data

State-Trait Anxiety Depression Inventory-State (STADI-S)

No significant effect of the factors was found neither for depression nor anxiety sub-scales of the state version of the STADI (STADI-S).

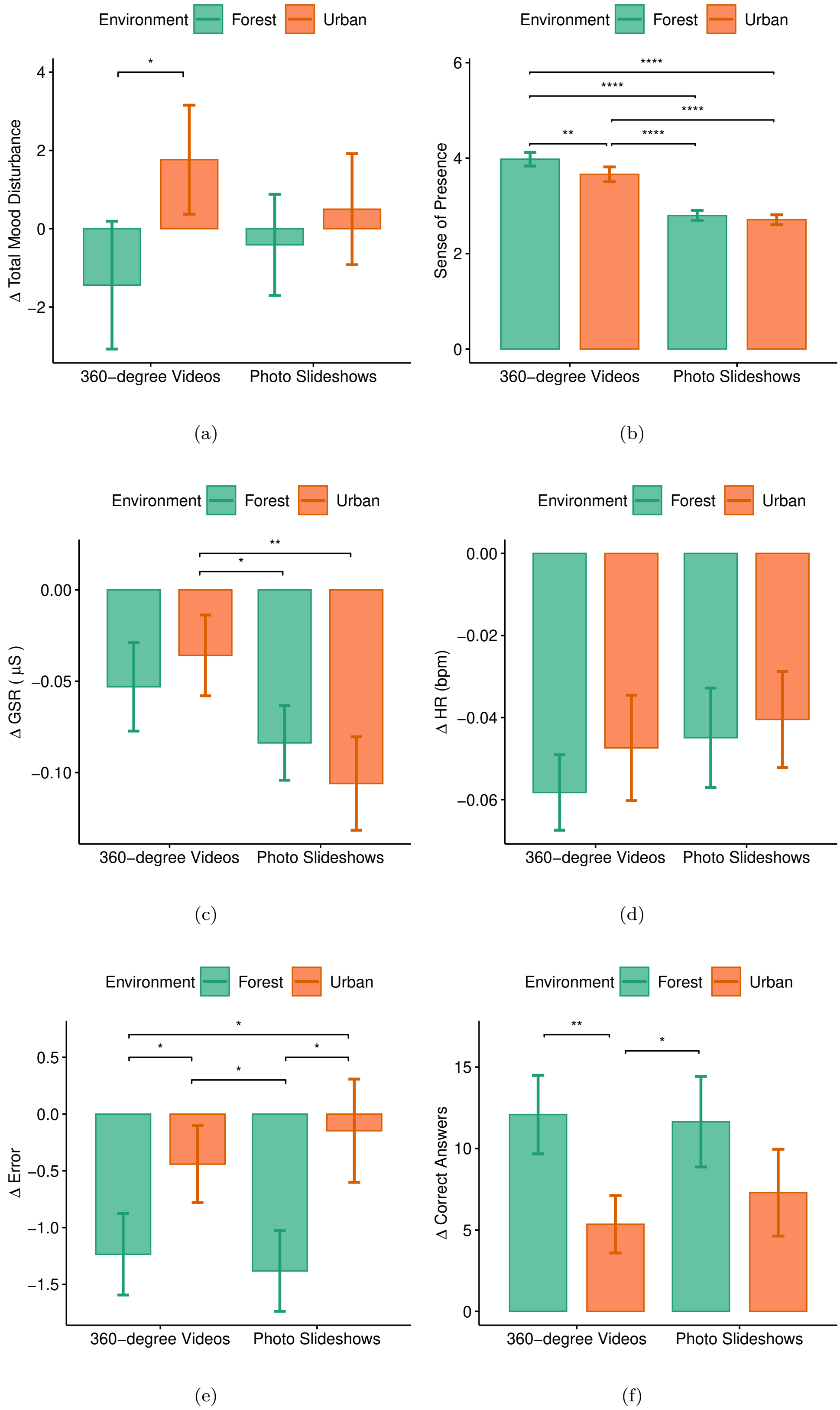


Figure 4.3: (a) Difference scores (exposure-control) of the total mood disturbance measured by POMS (b) IPQ sense of presence (c-d) Different scores (exposure-cognitive test) of the mean (c) GSR and (d) HR values (e-f) Different scores (exposure-control) of the cognitive test (e) errors and (f) consecutive correct answers

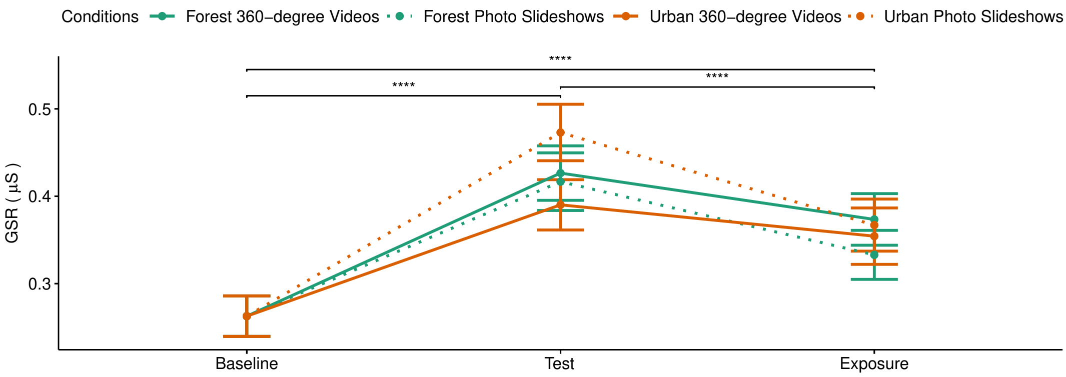


Figure 4.4: GSR measures during different phases of the experiment. The four conditions are distinguished by the line color (green for forest and orange for urban environment) and the line type (dotted for slideshows and solid lines for 360° videos).

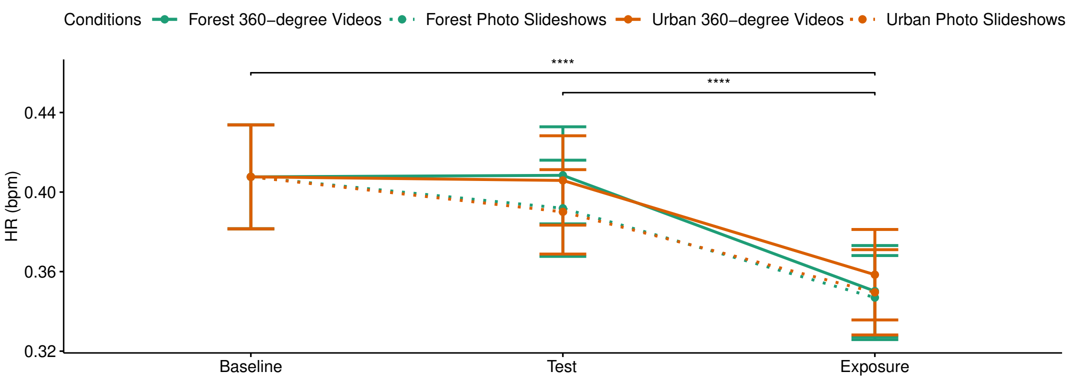


Figure 4.5: HR measures during different phases of the experiment. The four conditions are distinguished by the line color (green for forest and orange for urban environment) and the line type (dotted for slideshows and solid lines for 360° videos).

Profile of Mood States (POMS)

The environment factor (i.e., the type of environment: forest vs. urban) showed a significant negative effect on mood ($F(1, 33) = 5.02, p = .03, \eta_p^2 = .13$). Paired t-tests suggested that exposure to the urban environment disturbed the participants' mood more than the forest environment ($p = .027$) and that this difference was significant between the 360° videos of forest and urban environments ($p = .028$, see Figure 4.3(a)). A main effect of the immersion level or its interaction with the environment factor could not be observed. We also calculated the difference scores (exposure-control) for each sub-scale of POMS and performed the two-way ANOVA. A significant main effect of environment could be found for fatigue only ($F(1, 33) = 5.19, p = .03, \eta_p^2 = .13$). While the 360° videos of forest decreased the feeling of fatigue, pairwise comparisons revealed that this reduction was significantly different from the changes in fatigue (i.e., increase of fatigue) elicited by the 360° videos ($p = .027$) or photo slideshows of the urban environment ($p = .016$). No significant difference was observed between photo slideshows and 360° videos of the forest environment.

Short Stress State Questionnaire (SSSQ)

No significant main or interaction effects of the environment and immersion factors were found for neither of the SSSQ sub-scales namely task engagement, distress, nor worry.

Perceived Stress Scale (PSS)

No significant main or interaction effects of the environment and immersion factors could be found for the PSS score.

Igroup Presence Questionnaire (IPQ)

The IPQ scores of each condition were directly used for the analysis, without the control measurements being subtracted from them. As a result, a significant main effect of immersion level could be observed for the sense of presence ($F(1, 33) = 79.11, p < .001, \eta_p^2 = .706$). The sense of presence for the 360° videos was higher than the slideshow conditions ($p < .001$). The environment factor also had a significant effect on the sense of presence ($F(1, 33) = 13.927, p < .001, \eta_p^2 = .297$) in such a way that the forest environment induced a higher sense of presence compared to the urban environment ($p < .001$). No significant interaction effect was found (see Figure 5.5).

For the sense of being there (or the general presence) sub-scale, we found a significant main effect of immersion level ($F(1, 33) = 62.767, p < .001, \eta_p^2 = .655$), which shows that the mean value of the 360° videos was higher than the mean value of the slideshows ($p < .001$).

For the spatial presence, the main effect of immersion level was significant ($F(1, 33) = 96.371, p < .001, \eta_p^2 = .745$). Here, the mean value of the 360° videos was higher than the mean value of the slideshows ($p < .001$). Moreover, our results show a significant main effect of the factor environment ($F(1, 33) = 11.85, p = .002, \eta_p^2 = .264$) and therefore underline that the mean value of the forest environment was higher than the urban environment ($p < .01$).

Therefore, the 360° video and the forest environment led to a higher sense of spatial presence.

For the involvement sub-scale, the main effects of immersion level ($F(1, 33) = 19.649, p < .001, \eta_p^2 = .373$) as well as the environment ($F(1, 33) = 10.574, p = .003, \eta_p^2 = .243$) were significant. The 360° videos ($p < .001$) and the forest environment ($p < .01$) showed higher involvement values compared to respectively slideshows and the urban environment. In addition, a significant interaction effect of these two factors could be observed ($F(1, 33) = 35.254, p < .001, \eta_p^2 = .517$). Here the highest involvement values were observed in the 360° forest video ($p < .001$).

For the experienced realism, the main effect of the immersion level was significant ($F(1, 33) = 17.006, p < .001, \eta_p^2 = .34$) and the 360° videos had higher values than the slideshows ($p < .001$).

Simulator Sickness Questionnaire (SSQ)

The SSQ was administered two times: once at the beginning of the experiment (i.e., prior to wearing the HMD for the first time) and once in the end (i.e., after the last experimental condition). A paired t-test suggested a significant ($t(33) = 3.67, p < .001, d = .63$) increase of the total simulator sickness score from pre- ($M = 14.08, SD = 16.37$) to post measurements ($M = 30.8, SD = 28.33$). This means that the experiment and its total associated stay in VR increased the symptoms of simulator sickness.

4.3.2 Physiological measures

Considering the cognitive test as a stress induction task prior to exposure, it can be seen in Figure 4.4 that physiological arousal measured by the GSR values were increased during the cognitive test phase prior to the exposure and were decreased during the exposure for all four conditions. A three (experiment phase: baseline, (cognitive) test, exposure) x two (environment: forest, urban) x two (immersion level: 360° videos, photo slideshows) repeated-measures ANOVA showed a significant main effect of experiment phase ($F(2, 66) = 17.76, p < .001, \eta_p^2 = .35$) and a significant interaction of immersion and experiment phase ($F(2, 66) = 3.38, p < .05, \eta_p^2 = .09$). The results of pairwise t-tests showed significant differences ($p < .001$) between all three phases. Thus, the cognitive test could successfully serve its first function to induce stress prior to exposure.

Figure 4.5 depicts the mean HR values during different phases of the experiment for all four conditions. A three-way (experiment phase, environment, immersion level) repeated-measures ANOVA showed a significant main effect of experiment phase ($F(2, 66) = 13.51, p < .001, \eta_p^2 = .29$). Pairwise comparisons suggest that participants experienced the lowest HR values during the exposure compared to the baseline ($p < .001$) and cognitive test phase ($p < .001$). The difference between the baseline and the cognitive test phase was not significant probably due to ceiling effects. During the exposure, however, their HR decreased significantly.

Difference scores were calculated for both physiological measures by subtracting the mean values during the cognitive test phase from the mean values during the exposure phase (see

Figure 4.3(c) and 4.3(d)). A two-way repeated measures ANOVA for the factors environment and level of immersion showed a significant effect of the immersion level for the GSR values ($F(1, 33) = 8.55, p < .01, \eta_p^2 = .21$). A paired t-test suggested that the GSR difference scores were significantly larger ($p < .01$) for the photo slideshow conditions compared to the 360° video conditions. Pairwise comparisons showed that urban ($p < .01$) and forest ($p < .05$) photo slideshows caused larger difference scores compared to 360° videos of the urban environment (see Figure 4.3(c)). The pairwise comparisons did not show any significant difference between the 360° videos of the forest environment and any other conditions. No significant main or interaction effect could be observed for the HR difference scores. That is, all four conditions decreased HR with no significant difference.

4.3.3 Cognitive test

The cognitive test was considered as a dependent variable measuring the cognitive performance after the exposure phase. The environment factor had a significant effect on the errors ($F(1, 33) = 9.52, p = .004, \eta_p^2 = .22$) and the consecutive, correct answers given ($F(1, 33) = 13.56, p < .001, \eta_p^2 = .29$) in the cognitive test acquired after exposure. Paired t-tests revealed that the number of errors (see Figure 4.3(e)) in the forest environment was significantly lower than in the urban environment ($p < .001$) and the correct, consecutively given answers (see Figure 4.3(f)) were higher in the forest environment compared to the urban environment ($p < .001$). No significant main effect of immersion or its interaction with the environment factor could be observed.

4.4 Discussion

We hypothesized that the environment and the immersion level, as well as their interaction, have an influence on mood, stress recovery, and cognitive performance. In particular, we expected that the forest environment would produce a more positive effect than the urban environment. In addition, we hypothesized that more immersive presentations (i.e., 360° videos) create a higher sense of presence and consequently have greater effects, as more realistic environments would lead to realistic behavior and trigger corresponding responses [338].

The effects of exposure to forest or urban environments on mood was measured by means of total mood disturbance. Here, it could be shown that the type of environment was determinant for the mood disturbance as exposure to the urban environment led to a significant mood disturbance whereas exposure to the forest environment resulted in a reduction of mood disturbance. In particular, the feeling of fatigue was increased after exposure to the urban environment regardless of their type of presentation (i.e., 360° videos or photo slideshows) and was reduced by exposure to the 360° videos of forest. This result confirms the findings of previous studies [410, 319, 312, 235, 11, 61]. Despite variations in visual and auditory stimuli of the previously studied urban environments in VR, they all reported mood disturbances; whether the urban environment was a crowded subway station [410], or a shopping mall [312] or a shopping plaza [410], or a small town with buildings lining streets, some road traffic, a pedestrian mall, and the sound of traffic and people talking in the pedestrian mall [319]. We intentionally excluded crowds, cars, and prominent nature elements such as trees. Thus, our VR urban environment comprised of only buildings as visual stimuli and mostly the sound of wind breeze as auditory stimuli. Yet exposure to this environment disturbed our participants' mood. On the other hand, our forest environment was successful in reducing the disturbed mood. Since no significant effect of immersion level on mood could be shown in this experiment, it can be stated that photos of the environment are sufficient to observe its effects on mood. Nevertheless, the feeling of fatigue could be decreased only by exposure to 360° videos of the forest. Therefore, although Browning et al. [61] could show an improved mood for the real exposure to outdoor forest only, our findings proved that VR exposure to forest can be beneficial for inducing positive mood. This is in line with the findings of previous studies that showed exposure to VR nature can improve mood and reduce negative affect, whether the VR nature was 360° videos of rural areas and remote beaches [11] or various types of forest environments [385, 410].

No significant effects were found for STADI-S, PSS, and SSSQ. Since we already showed that the urban environment caused mood disturbance, one could expect that the STADI-S shows a similar effect. However, it should be noted that STADI-S rather covers more clinical aspects such as anxiety and depression, whereas the POMS measures mood changes in a more healthy range. Therefore, a possible explanation would be that the healthy volunteers did indeed experience a disturbance in mood by being exposed to the urban environment. However this mood disturbance was not strong enough to be detected by the scales of the STADI-S.

An increase of the physiological arousal measured by the GSR values could be observed from the baseline to the cognitive test phase which was again decreased by exposure to any of our four experimental conditions. The photo slideshows in this case were more effective in lowering the arousal levels compared to 360° videos. The reason here could be the higher immersion level of 360° videos and their associated sense of presence which has been shown to be positively correlated with physiological arousal [238, 248]. The immersive VR has been also shown effective in inducing emotions (such as the feeling of awe) and enhancing their intensity [71]. Thus compared to non-immersive photos, immersive videos may elicit higher emotional reactions which again results in higher physiological responses. Therefore, although 360° videos were able to reduce the physiological arousal, their higher immersion level prevented this arousal reduction to reach the same level as the non-immersive photo slideshows. Thus, to reduce physiological arousal caused by psycho-social stressors, one should rather use conventional photos of either urban or forest environments.

Besides GSR, we measured participants' HR during the course of experiment. The results showed that the HR was already high at baseline and the cognitive test did not increase it any further. A limitation of this study is that we did not plan any additional resting phase before the baseline measurement started. Perhaps this is the reason why the difference between the HR measurements at the baseline and during the test phase was not significant, probably due to ceiling effects. The exposure to any of our conditions, however, was able to reduce the HR significantly, regardless of the type of environment or the level of immersion. This finding is in line with previous studies [410, 11]. For instance, Yu et al. [410] showed that blood pressure and HR were reduced by exposure to both urban and natural environments with no significant differences between them. Also, Anderson et al. [11] showed that HR variability was reduced during the exposure to natural and indoor VR environments with no clear differences across them. In their study, the GSR values were also decreased for all conditions but this reduction was greater for the natural scenes, similar to the findings of Hedblom et al. [150]. Moreover, Schebella et al. [312] showed that recovery from stress measured by HR values could be achieved by VR exposure to both natural and urban environments with no significant difference between them. Nevertheless, it is important to determine whether and which stimuli are best suited to decrease the induced stress. By analyzing the GSR values, we could show that all our stimuli could reduce the induced stress, but non-immersive stimuli were more effective in doing so.

The hypothesis that exposure to forest environment improves cognition was confirmed by this study. It could be shown that the maximum number of correct answers (number series) in two conditions exposing to the forest environment was higher and the total number of errors was lower. This can be attributed to the positive effect of exposure to the forest environment and cognitive benefits of interacting with nature which has been studied before [180, 179, 39, 257, 144, 143, 74, 185]. For instance, Berman et al. [39] found that viewing pictures of natural scenes can improve cognitive performance compared to urban scenes. Also, Chung et al. [74] showed that 360° videos of nature can restore involuntary attention. In our study however, neither an effect of immersion level nor an interaction with the environment factor could be found. Therefore, it can be concluded that the presentation of forest using both methods namely the photo slideshows and the 360° videos had a positive effect on cognition with no significant difference between them. Thus, to induce a positive effect on cognition, a presentation of forest using the conventional photo slideshows might be enough to produce the full impact of forest exposure.

The cognitive test in this experiment served two functions: as a stress induction task prior to exposure and at the same time as a dependent variable measuring the cognitive performance after exposure. Initially, we had planned to use an additional cognitive test to measure the cognitive performance. However, during the piloting phase, we realized that the total length of the study could overwhelm the participants. Therefore, we used only one cognitive test to keep the length of the experiment limited to a reasonable time. This is a limitation of this study and in future work, these two functions could be disentangled from each other.

The results of the SSQ suggested that the experiment increased the symptoms of simulator sickness. A reason could be that the study was 180 minutes long during which a cognitive task was carried out repeatedly. Therefore, participation in the study could have led to fatigue and may have caused or exacerbated the symptoms of simulator sickness. Therefore, the symptoms can and should not be attributed solely to the exposure to the virtual environments. Moreover, a limitation of this study is that the SSQ was not administered after each condition.

Thus, it cannot be determined whether different levels of immersion or types of environment played a role in inducing simulator sickness. Whether potentially associated simulator sickness prevented the positive effects of nature to occur remains a topic for future research.

The hypothesis that the immersive 360° videos can facilitate the positive effects of nature onto mood, recovery after stress and cognition could not be demonstrated in this experimental setup. Nevertheless, the IPQ results showed that the 360° videos did induce a higher sense of presence compared to the slideshows. Also, the IPQ sub-scale involvement was highest in the 360° video of a forest. Therefore, the question arises whether the provided level of immersion for the 360° videos was sufficient for changing our affective and cognitive measures. The 360° videos of this study were taken in a resolution of 4K and were monoscopic (i.e., the same image was displayed on both lenses). Monoscopic images lack cues of depth perception that affect the sense of spatial perception [250]. Since realistic representations have an impact on immersion [338], the use of a 360° camera with a higher resolution (to render more realistic stimuli such as the movement of single leaves') and stereoscopic display (i.e., different images shown on the respective lenses to create a sense of spatial depth) should be considered in future investigations. It might still be true that with stronger immersion and sense of presence, the reactions of the participants could have been more different between 360° videos and the photo slideshow.

However, in the present setup an advantage of using 360° videos compared to photos could not be determined, the positive influence of natural environments on cognition and reduction of mood disturbance could be observed. As the use of visual representations of natural environments can be a viable option in contexts that offer little access to natural resources. In future work, the underlying elements of the forest environments that cause the more positive impact in contrast to the urban environments should be further studied. It would be possible that it is not the forest in its complexity that is necessary to trigger the observed positive effects, but rather some bottom-up visual features that are commonly found in nature pictures. Previous research has shown that preference ratings of nature pictures can be explained by such lower-level image features [181, 165].

The Prospect-Refuge Theory [17] suggests that humans have preferences for certain environments. According to Appleton [17], humans prefer places that offer a safe and sheltered refuge and at the same time a good view or overview of the surrounding environment. This theory relies on evolutionary approaches, which require a predator to be able to observe a potential prey without being discovered. Accordingly, there is the possibility that there may be a natural preference for the environment of a dense forest over an empty road or a rather open space within a city. Consequently, future studies should take these aspects into account while selecting the virtual environments to be compared.

In this work, the visual stimuli of natural and urban environment were accompanied by the respective auditory stimuli recorded from that environment. In other words, while seeing either 360° videos or photo slideshows of forest environment, our participants could listen to the sound of birds singing in that forest. The selected urban environment for this study was an empty old town in which mostly a soft wind breeze could be heard. As mentioned earlier, our main reason for including the auditory stimuli was to increase the feeling of presence [338, 153]. However, since Annerstedt et al. [15] showed that recovery from stress by exposure to VR nature was enhanced when the environment had nature sounds, one may consider to repeat the present experiment to investigate the effects of pure visual stimuli. Furthermore, in the 360° videos of the forest, subtle movements of the leaves of the trees caused by the wind breeze as well as the changes of the sunlight when shined through the trees were observable. Such subtle changes could indeed not be observed in the urban environment. We consider this as a limitation of this work. Thus, future studies may decide to provide an urban scene with a comparable level of movement as the forest.

In addition, in this experiment, the conventional photo slideshows were displayed using a VR HMD which was required for presenting the 360° videos but not for the photo slideshow. On the one hand, using the HMD for both conditions enabled us to control for unintentional effects of the display medium while comparing the immersion properties of the presented materials (i.e., 360° videos or conventional photos), but on the other hand limited us from generalizing the findings to other display media. As presentation of conventional photo slideshows on a monitor may not produce the same effects as presenting them using an HMD and remains a topic for further research.

In sum, the benefits of interacting with real or virtual nature has been reported in previous studies. Virtual exposure to nature has been administered classically using conventional photos and recently using immersive 360° videos or computer generated models of nature. In this chapter, we aimed to answer the question of whether immersive 360° videos of nature intensify its positive effects on mood, stress recovery, and cognition compared to conventional photos of nature. Our results suggest that indeed exposure to photos of a forest environment suffice to prevent mood disturbance observed in response to urban exposure, reduce physiological arousal, and improve cognition. In addition, photos of either urban or forest environment were both more effective in reducing physiological arousal compared to immersive 360° videos. Thus, in contrast to our priori hypothesis, more immersive presentation of the forest environment could not lead to more positive effects of nature.

Chapter 5

5 Effects of Virtual Audience Size on Social Anxiety during Public Speaking

Prior studies have explored the possibility of inducing Social Anxiety (SA) in VR. Among various existing protocols for this purpose, TSST has been proven to be robust in evoking SA in the majority of participants in both in vivo as well as VR conditions. The TSST consists of giving a speech and performing mental arithmetic calculations each for five minutes in front of three persons. In this chapter, an adaptation of TSST will be presented which aims at investigating the effects of different numbers of VHs (i.e., three, six, or fifteen) on perceived SA. In addition, the results were compared to an in vivo TSST with three real persons in the audience. Twenty four participants took part in this experiment. As a result, physiological arousal could be observed with VR inducing SA yet less than in vivo TSST. Furthermore, some of the subjective measures showed a state of anxiety experienced during the experiment. An effect of the virtual audience size could be observed only in HR as a virtual audience size of three VHs induced the highest HR responses which was significantly different from an audience of size six and fifteen.

5.1 Introduction

Anxiety disorders are the most widespread category of mental disorders, while Social Anxiety Disorder (SAD) also known as social phobia is the most common type [347]. People with SAD avoid social interactions with others. Their concern is saying or doing something that causes embarrassment and humiliation. As a result, they often suffer from low self-esteem, high self-criticism, and depressive symptoms [347].

There are several psychological and pharmacological treatment methods for SAD. ET is one of the most commonly used CBT methods [176, 100]. ET relies on the assumption that SAD is a conditioned (i.e., learned) reaction and exposure provides an opportunity for deconditioning (i.e., corrective learning) of the anxious responses [100]. Therefore, patients are carefully exposed to the anxiety provoking stimuli, i.e., situations, people, or places which trigger their anxiety, in order to enable them to gradually control their anxious reactions.

In this context, VR has enormous potential to facilitate therapy by providing a safe, controlled environment for VRET. ET administered by means of VR can be applied more systematically and in a contextually relevant setting [264]. Real-time feedback, enhanced ecological validity, low-cost environments, which can be duplicated and distributed are some of the benefits of using VRET [293, 292].

Furthermore, it has been suggested that users show higher preference and acceptance of VR versus in vivo exposure. For instance, 81-89% of the subjects of a study on spider phobia [126] and 76% of the participants of another study on specific phobia [125] chose VRET over in vivo exposure. The refusal rates for the in vivo ET (27%) were also significantly higher than the VRET (3%) in the latter study. Therefore, VR provides a preferred alternative and could potentially increase the number of individuals who seek and undergo treatment.

Multiple studies using VRET have demonstrated its success in treating anxiety and specific phobias, which is even reflected in the American Psychological Association's Practice Guideline for the Treatment of Patients with Acute Stress Disorder and Post-traumatic Stress Disorder [36] as well as its Presidential Task Force on Evidence-based Practice in Psychology [258]. In particular, VRET has been used as an effective and evidence-based treatment method for several psychological disorders such as post-traumatic stress disorder [237, 244, 282, 283], fear of flight [391, 299], fear of heights [198, 96] as well as fear of public speaking [12, 50, 177, 295, 191, 140, 383].

ET for the fear of public speaking is commonly practiced by the patient delivering a speech in front of an audience. The audiences in this method are usually trained actors. Thus, by using VR for the treatment, both patients and therapists can profit from the mentioned advantages of VRET. For instance, the therapist can easily change numerous parameters of the simulation such as the number of people in the Virtual Audience (VA), their appearance, or their behavior. Having the same conditions in real life requires hiring and training the same amount of actors for every single exposure. However, this flexibility also introduces new research questions: **Q1)** Does the virtual audience size affect the VRET? and **Q2)** What is the optimal virtual audience size for VRET?

Based on the Social Impact theory (SIT) [205, 206], a variety of changes in physiological states, subjective feelings, cognition, and behaviours can occur in an individual as a result of the (real, implied, or imagined) presence (or actions) of other individuals. Three factors play a role in the social impact felt by the individual: strength (of the source), immediacy (of the event), and number of source persons. According to this theory, as the number of persons increases, the impact felt by the individual will increase, but with a negatively accelerating power function. In other words, an increase from no audience to one person has a much greater social impact compared to an increase in the audience size from 5 to 6 persons [236]. In a study on the effects of audience size during a public speaking task [236], heart rate parameters of highly speech anxious subjects were higher during a presentation in front of an audience with a size of two persons compared to six people. These values were also higher compared to a no audience (size 0) condition.

Most of the VRET applications for treating SAD and fear of public speaking have compared no treatment vs. treatment, in vivo ET vs. VRET, and no audience vs. an audience. The reported audience sizes were mostly kept constant within a study and only differ between studies. To the best of our knowledge only in one study [12] the VA size as well as the VE were varied, but the effect of these changes was not reported.

In this chapter, we present a VR implementation of TSST [189] to analyze the effects of different virtual audience sizes. The TSST is a standardized and widely used protocol for inducing moderate social stress in laboratory settings. In brief, TSST consists of a preparation and a test period for delivering a speech and performing mental arithmetic in front of three persons. In addition to the comparison of the in vivo and VR TSST, we examine the effects of a change in the VA size (3, 6, or 15 virtual humans (VHs)) on the induced stress.

5.2 Related Work

The TSST was first introduced in 1993 by Kirschbaum and colleagues [189] and consists of purely psycho-social stressors. The participants in this protocol have a short preparation time and immediately after that they have to give a free speech and solve mental arithmetic (each for five minutes) in front of three persons. The audience comprises always of males and females who are trained to provide neutral feedback in response to the participants' speech and arithmetic performance. The participants are informed that the audience is trained in monitoring nonverbal behavior. For further voice frequency and video analysis, they are told that their voice and video will be recorded too. This part of the protocol is only for the stress inducing effects as later after the test participants will be informed that none of these analysis will actually be carried out. The free speech part of the test is a simulated job interview during which the participants have to convince the audience that they are the perfect applicant for the open position. If they finish their speech in less than five minutes for the first time, they are told that they still have some time left and they should continue. Whenever the interruption of the speech happens after that, the audience stays quiet for 20 seconds and then asks prepared questions. After the five minutes speech is over, the audience instructs the participants to serially subtract number 13 from 1022 as fast and as accurately as possible. On every wrong calculation, they are instructed to stop and start from 1022 again.

The subsequent studies and meta analysis of the TSST have confirmed its stress induction effect [156, 90, 69, 109]. Consequently, it has become one of the most widely used protocols for inducing stress to study its effect on not only psychological disorders such as SAD [192, 196, 297] but also on learning [172], memory [79, 229, 393], and physical health[184, 243]. The TSST can be used for both children and adults [190, 199, 200, 407] and has been shown to be robust against protocol variations [134].

A meta analysis on VRET applications for anxiety and related disorders [70] has reported on eight successful VRET studies for SAD [140, 191, 295, 12, 177, 50, 383, 44]. Each of these studies have used one or multiple VEs, which corresponded to different cases of SAD such as observation anxiety, intimacy anxiety, scrutiny anxiety, assertiveness anxiety, and performance anxiety. The latter case of SAD has been considered in all these studies. While one study focused on music performance anxiety [44], the rest studied the fear of public speaking in front of a VA. However, the number, appearance, and behaviour of these audiences were different across all these studies. For instance, in one study [12] three different number of VHs (5, 35, 100) were the audience in three different virtual rooms (conference room, classroom, and auditorium), but the effects of the number of VHs in the audience and VE were not reported. Another study [336] could show that speaking in front of a VA, compared to an empty virtual room, could significantly increase the signs of anxiety in people with SAD. Although their VA showed various negative, neutral, and positive behaviours during the test, they did not report on the effects of a change in the behaviour of the audience. This is unfortunate since some studies [270, 335] have suggested that speaking to VHs that provide negative feedback can provoke greater anxiety compared to a neutral or positive audience.

In addition, some studies have monitored participants' physiological responses, while anxiety responses have been assessed mostly through self-report questionnaires. For instance, Kothgassner et al. [195] has examined three types of stress responses (self-reported, autonomic, and endocrine) during a 5 minute public speaking task in an empty virtual lecture hall or in front of 20 virtual or real (in an actual university lecture hall) audience. The assessment of the HR, and salivary cortisol as well as the self-reported state anxiety demonstrated comparable increase in all types of stress responses in both real and virtual social stimuli.

Various VR versions of TSST have been examined in previous studies [154, 173, 246, 301, 384, 412, 325, 105]. Although, the results of all these studies support the use of VR TSST for social stress induction, the number of VHs and their behaviour does not always follow the original TSST protocol. For instance, one pilot study [173] examined an implementation of TSST for VR using a CAVE™ system for 10 healthy male subjects. The VHs in this study performed gentle movements during the test. Although having animated VHs could resemble the behavior of real people in the real-life TSST, the selection of the animations could potentially influence the neutrality of the audience. The mentioned animations of the VHs in this study [173] (i.e., nodding of the head, moving a foot, and turning the face away from or towards the participant) could (un-)intentionally provide positive or negative feedback to the participant. Perhaps the most straight forward implementation of TSST protocol for VR has been suggested by Zimmer et al. [412] in which only minor changes (e.g., shorter preparation time without possibility to take notes) were made in the VR protocol.

5.3 Virtual Trier Social Stress Test

As mentioned before, prior implementations of TSST in VR have varied certain aspects of the test, but often without detailed consideration of their respective effects. In our study, we analyzed the effects of VA size in VR TSST, and compared the results with an in vivo TSST with the original number of juries (i.e., three persons). We implemented three conditions of the TSST for the purpose of this study. The VEs consist of a neutral room (which was not a one-to-one recreation, but similar to the physical room), a table, and three different number of VHs, which are arranged as following: **N3)** three VHs in the first row (Figure 5.1(a)), **N6)** six VHs in the first row (Figure 5.1(b)), and **N15)** fifteen VHs, six in the first row, five in the second row sitting in between the six so that the participant was able to see their faces clearly (not directly behind the VHs in the front row), and four in the third row, again sitting between the five persons seated in the previous row (Figure 5.1(c)).

These sizes for the VA were chosen to represent the original TSST audience size (N3), two times larger (N6), and five times larger (N15) than the original size (N3) and almost double the size of the level in between (N6). The number 15 was carefully chosen to serve an additional function, which was making the virtual room appear full of visible VHs. Moreover, the gender of the VHs was always in a proportion of two males and one female and their age was between 40 and 60 years old. Their clothing was also fixed to half of the male VHs wearing a white medical coat and the rest business suits. These settings were inline with the in vivo juries' age, gender, and clothing. Moreover, VHs presented a neutral facial expression and various neutral postures. VHs randomly changed their sitting positions between conditions. The animations, such as crossing the legs, were designed to increase the lively properties of the VHs but nonetheless not convey positive or negative feedback. Furthermore, they fully payed attention to the participant. That is, their head and eyes followed the participant's head, which gave an impression of having continuous eye contact. Finally, the main VH (i.e., the male VH in the white medical coat in the first row who was presented in all three conditions) welcomed the participants, gave them necessary instructions, and asked prepared questions during the test.

5.4 Experiment

We ran a user study to compare the in vivo and VR TSSTs in terms of the evoked stress and to find out whether the number of VHs can change the participants' perceived SA. Since some studies [318] suggest that the individuals with SAD avoid mutual gaze in social interactions, we hypothesized that the higher their level of SAD, the more the participants would avoid looking at the most stressful elements in the VE. We assumed that the VHs are the most stressful elements and particularly the ones in white medical coats (e.g., the main VH in N3).

Therefore, the following hypotheses were formulated: **H1)** In vivo TSST can evoke SA. **H2)** VR TSST can evoke SA. **H3)** The evoked SA of VR is as high as in vivo. **H4)** SA in VR increases with an increase in the size of VA. **H5)** In VR, the main VH of N3 will be the most avoided VA. **H6)** The avoidance rate towards the main VH is positively correlated with the level of SAD.

This study was approved by the local Ethics Commission of the Department of Informatics of the Faculty of Mathematics, Informatics and Natural Sciences of Universität Hamburg and was performed in the group therapy room of the anxiety and obsessive-compulsive disorder

(a)

(b)

(c)

Figure 5.1: TSST in VR: (a) N3: three VHs (b) N6: six VHs (c) N15: fifteen VHs

station at the University Medical Center Hamburg-Eppendorf.

5.4.1 Participants

The recruitment was accomplished by means of mailing lists, online job markets, and university boards. Since the entire experiment was designed to be conducted in German language, having sufficient German language skill (not necessarily as mother tongue) was required (min=4, max=5, mean=4.92, SD=.28, on a 5-point Likert scale). To be more precise, only two participants rated their German skills 4 and the rest selected 5. The exclusion criteria were any history of psychiatric disorders, consumption of any medication which may influence the perception of stress, and age below 18 or above 55 years. Most of the participants (18 people, 72%) did not have any vision correction, 5 people had glasses, and 2 contact lens corrections. Furthermore, 16 participants (64%) had prior VR experience. None of them had any balance disorders nor smoked (since nicotine consumption can influence the stress responsiveness [189]). All applicants were also screened using the MINI [155].

Moreover, we used three questionnaires to estimate the SA level of the participants. The first questionnaire was the Social Phobia Diagnostic Questionnaire (SPDQ) [253]. Following the DSM-IV model for social phobia, SPDQ has 25 items including several yes/no questions and symptom rating 5-point Likert scales. The SPDQ total score can range from 0 to 27. The eligibility criteria for this study was SPDQ total score ≥ 1 and our participants had an average SPDQ of 4.86 (SD=3.25). 24 people (age=28.17 (SD=7.54), 18 females) out of 316 applicants were eligible to take part in the study.

While SPDQ is a diagnostic assessment tool based on DSM-IV, Liebowitz Social Anxiety Scale (LSAS) [223] is designed as a clinician-rated measure of a person's level of fear and avoidance in social situations. To be in line with the extensive body of literature based on LSAS, it was administered as a complementary measure for the SAD level. This questionnaire has 24 items, each of which describes a social situation which can evoke anxiety and avoidance. The SA in this questionnaire was calculated by summing up the rated fear and avoidance for all 24 items (0-144). The mean LSAS score of our participants was 31.22 (SD=19.42).

STADI-T was also used to estimate the participants' tendency to perceive stress. See Chapter 3, Section 3.4.1 for more information about the questionnaire. The mean value of all participants was 34.39 (SD=7.46).

5.4.2 Materials

The VR TSST application was developed using Unity 3D game engine and C# programming language. The VHs were composed and animated using Adobe Fuse CC and Mixamo. Facial blend shapes were created using Blender to create visemes and were synced with the main VH's speech using the Oculus Lipsync plug-in. The speeches were generated using an online Text-to-Speech service[1]. The quality of the generated sounds was approved by the clinical experts of the team (i.e., it was not considered artificial nor a source of distraction). For

[1] http://www.fromtexttospeech.com/, language: German, voice: Michael, speed: fast

rendering, system control, and logging an Intel Computer with Core i7 7700K processor, 32GB DDR4-RAM, and MSI GeForce GTX1080Ti 11GB Founders Edition graphics card was used. The experimenter used a monitor, a wireless mouse, and a keyboard to run the Wizard of Oz setup (i.e., a setup which is operated by the experimenter but seems to be autonomous, e. g., when the main VH reacts to the participant and speaks, the experimenter decides whether and what he should say but the participant is not aware of this).

To measure stress, we employed several physiological measures and questionnaires. Physiological measures included salivary cortisol, GSR, and HR. In addition, we tracked participants' head direction in VR by casting a ray from the virtual camera to record the number of times it hit the VHs. This is particularly helpful for distinguishing those VHs, at which the user looked or avoided looking during the test. The distance between the participant and the VHs as well as the juries in the in vivo condition was fixed at 2.5 meters. Thus, the real and virtual audiences were in the participant's social zone [136].

The participant wore an HTC Vive HMD, a headphone, Neulog GSR (worn on the non-dominant hand), and a SCHILLER's medilog® AR12 plus Electrocardiogram (ECG) sensor. The questionnaires were completed on a laptop (Intel(R) Core(TM) i5 2.5GHz processor, 8 GB RAM, and an Nvidia GeForce GTX 1050 graphics card) placed close to the participant on a small rolling table that could easily be placed in front of the participant. In the in vivo condition, this laptop was also used for recording from the GSR and Pulse sensors. The cortisol samples were taken using the Salivette® cotton swabs with 20mg citric acid. To take a saliva sample a cotton swab was chewed for exactly 1 minute by the participant. For this purpose, the participants were asked to not eat or drink except water for at least 30 minutes prior to the study.

In addition, we measured the participants' emotions using STADI-S and Self-Assessment Manikin (SAM) questionnaire [56], which is a non-verbal pictorial tool to assess affective dimensions of pleasure, arousal, and dominance.

After each TSST, the participants were given a self-made graph paper which asked them to state the degree of their fear on a scale of 0 (no fear) to 10 (extreme fear) for each minute of the TSST. This measure is common in the therapeutic context to rate therapeutic sessions. We did not receive any complaints when our participants understood and completed the task. We calculated the total fear score of each condition by summing up the values of all 10 minutes. This produced a scored between 0 to 100.

Moreover, we designed a questionnaire, namely State Social Anxiety (SSA), for the purpose of this study to address detailed aspects of the social anxiety state of participants during the speech part of the TSST. It consists of 8 items, two of which have respectively 16 and 7 sub-items. All items and their sub-items were to be answered on a 5-point Likert scale which range from 0 (not at all) to 4 (very strongly). The total score of this questionnaire can be calculated by summing the responses to all questions which can range from 0-116.

The Co-Presence and Social Presence questionnaire [273] was also used to measure participants' perceived feeling of VHs' social presence. It has 15 statements to measure four sub-dimensions of social presence using 5-point Likert scales (1=strongly disagree - 5=strongly agree).

The VR simulator sickness was measured by means of Kennedy's SSQ [186]. The Kennedy's procedure was followed for scoring the SSQ. That is, the sum of items 1, 6-9, and 15-16 giving S1 were multiplied by 9.54 to calculate the Nausea's value; the sum of items 1-5, 9, and 11 (S2) was multiplied by 7.58 resulted in the Oculomotor's value; and the sum of items 5, 8, and 10-14 (S3) was multiplied by 13.92 for the Disorientation's value. The total SSQ score was then computed by summing up S1-S3 and multiplying it by 3.74. Besides, we used IPQ [316] to measure the experienced sense of presence in VR. It has to be mentioned that the German version of all questionnaires (including IPQ) were used in this experiment.

5.4.3 Methods

We used a within-subject experimental design with three TSSTs in VR (N3, N6, and N15) and one in vivo TSST. All participants signed an informed consent form prior to the study and underwent all conditions. The in vivo and VR TSSTs were conducted on two separate sessions whose day of the week and time of the day were kept the same for each participant. For instance, P1 could have the in vivo session on Tuesday at 9 a.m. and the VR TSST on Tuesday at 9 a.m. one week before or after the in vivo session. The order of VR and in vivo TSST were fully counterbalanced. Thus, half of the participants started with the in vivo first and the other half with VR first. The order of VR TSSTs were as well fully counterbalanced within the VR session. The procedure of the VR and in vivo TSSTs were as following:

- **Baseline**: the participant is welcomed, fills out the consent form, gives the first cortisol sample, gets sensors attached, and fills out the following questionnaires in this order: demographic (only in the VR session), SSQ (only in the VR session), SAM, STADI-S.

- **Preparation**: in VR, an overlay text gives a short explanation of the participant's tasks. It is also stated in this introductory text that the VHs will concentrate on the non-verbal behaviors of the participants and that during the test their video and voice will be recorded for later analysis. The experimenter gives this information to the participant in the in vivo condition. Then they receive a job title, a pen, a paper and 5 minutes time to prepare themselves for the given job interview. The job titles used were police officer, teacher, travel tour guide, and salesperson. These titles were selected as it was assumed that they could potentially induce SA due to their need for interaction with other people and were selected based on the assumption that everyone knows enough about the job to apply for it. The order of the job titles were counterbalanced as well. A cortisol sample was taken and two questionnaires, i.e., SAM and STADI-S were filled out in the end of the preparation time.

- **TSST**: The test consists of two parts:

  - *Presentation:* depending on the given order to the user, the VR application shows three (N3), six (N6), or fifteen (N15) VHs. In the beginning, the main VH (or main jury member in the in vivo condition) asks the participants to deliver their speech during the next 5 minutes. If they stop talking for the first time, he will encourage them to continue and states that there is still some time left. If they cannot continue any further, the main VH/jury will put any of the eighteen prepared questions (e.g., Why haven't you found a position yet?). The questions were chosen at any time by the experimenter through the Wizard of Oz setup in the VR or by the main jury in the in vivo condition and were tried to match the content of the given speech. There were no substantial differences in the selected questions for each participant as the number of questions was limited and they were distributed among all three presentations (N3, N6, and N15). If needed, the VH/jury could encourage the participant to speak louder or indicate that he is not able to answer their questions (if the participant asks a question).

- *Math:* after 5 minutes of presentation, the main VH or jury asks the participant to subtract serially the number 13 from 1022, 1333, 1555, or 1777 respectively in N3, N6, N15, or in vivo condition (i.e., the starting number is fixed per condition). The participants do not receive any feedback when they give the right answers, but immediately after giving an incorrect answer they are asked to start from the beginning. This phase lasted for 5 minutes. After each TSST a cortisol sample was taken and the following questionnaires were filled out in this sequence: SAM, SSQ (only after the last TSST (i.e. TSST3) in the VR session), STADI-S, SSA, Co-Presence and Social Presence (only in the VR session), self-estimated fear, and IPQ (only after TSST3 in the VR session).

- **Rest**: the participants were allowed to rest for 5 minutes after each TSST and for 20 minutes in the end of the session. After that a cortisol sample was taken.

The in vivo session lasted for approximately 1 hour and 20 minutes and the VR session for 2 hours and 15 minutes. The entire length of the experiment was therefore about 3 hours and 35 minutes. After the study, the participants were compensated with course credits (if they were students) and up to 60 Euro.

5.4.4 Results

An α level of .05 was used for all statistical tests. The normality of the data was tested using the Shapiro-Wilk test. In case of violation of the sphericity assumption by Mauchly's test, degrees of freedom were corrected using Greenhouse-Geisser estimates of sphericity. The p-value of the pairwise comparisons were corrected using the Bonferroni method (referred to p.adj in the rest of the chapter).

Physiological Measures

Heart Rate (HR)

The raw ECG signals were manually denoised using medilog® Darwin V2 software. The HR data were normally distributed based on the Shapiro-Wilk test. The repeated-measures ANOVA on HR values showed a significant difference between different stages of the in vivo TSST ($F(1.88, 43.34) = 49.91, p < .001, \eta^2_G = .32$). The pairwise t-test comparisons

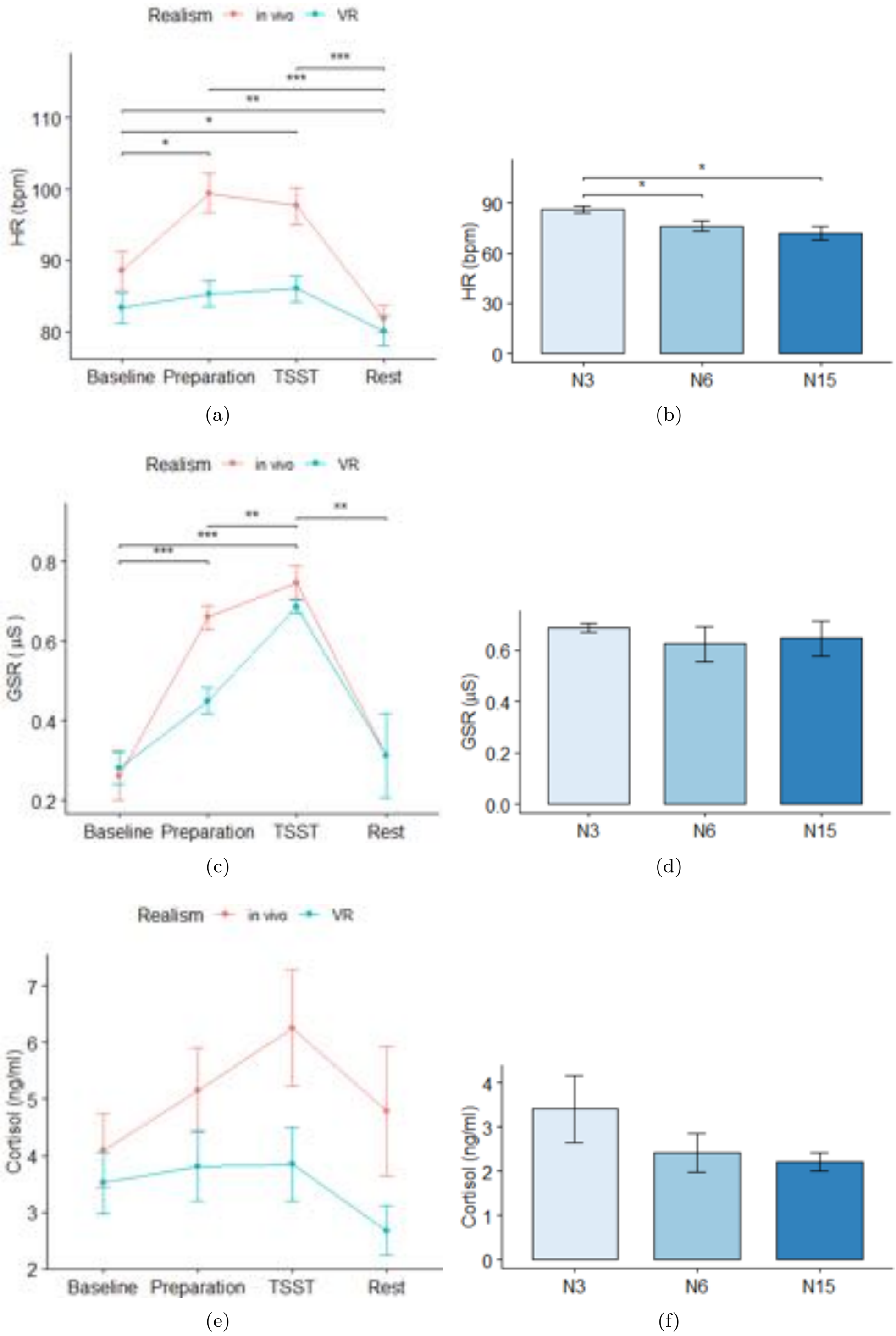


Figure 5.2: Physiological measures (a,b) HR (c,d) GSR (e,f) Cortisol. (a,c,e): realism effect (in vivo vs. VR-N3); (b,d,f) VA size effect in VR

revealed that the HR during the in vivo TSST was significantly higher than the baseline ($p < .001, p.adj < .001$) and the resting state ($p < .001, p.adj < .001$) which supports H1. The post-hoc power analysis with a repeated-measures correlation of .77 showed the power of .99.

The repeated-measures ANOVA showed also significant differences between different stages of N3 ($F(1.49, 34.25) = 20.13, p < .001, \eta_G^2 = .08$), N6 ($F(2.1, 48.5) = 20.4, p < .001, \eta_G^2 = .06$), and N15 ($F(2.14, 49.17) = 29.22, p < .001, \eta_G^2 = .1$). The pairwise comparisons showed that the resting state was significantly lower than N3 ($p < .001, p.adj < .001$), N6 ($p < .001, p.adj < .001$), and N15 ($p < .001, p.adj = .001$) which support H2.

The comparison between the in vivo TSST and N3 for all participants suggested a significant main effect of realism ($F(1, 23) = 55.63, p < .001, \eta_G^2 = .15$) and an interaction effect between realism and TSST stages ($F(1.8, 41.8) = 25.1, p < .001, \eta_G^2 = .06$). The pairwise comparisons also revealed that HR during in vivo TSST was significantly higher than N3 ($p = .002, p.adj = .002$). Similarly, when comparing the in vivo with N3 as the first TSST in VR (see Figure 5.2(a)) a significant difference between VR and in vivo ($F(1, 7) = 39.55, p < .001, \eta_G^2 = .32$) was found. There was a significant interaction effect of the realism and TSST stages ($F(1.48, 10.42) = 9.17, p < .001, \eta_G^2 = .15$) as well. The post-hoc power analysis suggested a power of .6 for this part of the test. The post-hoc pairwise t-tests indicated that the HR during in vivo TSST was significantly higher than N3 ($p < .001, p.adj < .001$). Thus, H3 cannot be supported by the HR data. The HR of those participants who experienced VR for the first time in our experiment did not differ significantly from the rest of the participant, considering N3 as the first TSST in VR ($p = .582, p.adj = .58$) or regardless of its order of presentation during VR session ($p = .079, p.adj = .079$).

To examine H4, we first compared N3, N6, and N15 regardless of their order of presentation to the participants. The repeated-measures ANOVA showed a significant difference between them ($F(2, 46) = 27.27, p < .001, \eta_G^2 = .1$). Although the original p-value of the pairwise comparisons suggested a difference between N3 and N15 ($p = .03$), the Bonferroni corrected p-value did not show a significant difference ($p.adj = .09$). Thus, we looked into the HR values during N3, N6, and N15 when they were the first, second, and third TSST of the VR session. For the first TSST (see Figure 5.2(b)), ANOVA suggested a significant effect of the VA size on HR ($F(2, 21) = 5.33, p = .013, \eta_G^2 = .34$). The post-hoc unpaired t-tests also showed that the HR during N3 was significantly higher than N15 ($p = .007, p.adj = .02$) and N6 ($p = .02, p.adj = .03$). We could not find a significant effect of VA size for the second TSST. For the third TSST again ANOVA suggested a significant difference between VA sizes ($F(2, 21) = 4.78, p = .019, \eta_G^2 = .31$). The post-hoc tests showed that N3 was significantly higher than N6 ($p = .016, p.adj = .049$). Although a trend towards higher N3 compared to N15 during TSST3 could be observed, the differences were not statistically significant ($p = .03, p.adj = .069$). The post-hoc power analysis showed a power of .75 for this part. Thus, H4 could not be confirmed, yet an effect on the opposite direction (i.e., a decrease of SA with an increase in VA size) could be observed.

Galvanic Skin Response (GSR)

The GSR data were sampled at 60 Hz and smoothed using a low-pass Butterworth filter with the cutoff frequency at 1 Hz and normalized using the following formula [230, 149]:

$$\tilde{G} = \frac{g - min(\bar{g})}{max(\bar{g}) - min(\bar{g})} \quad (5.1)$$

Where G is the raw GSR signal, g the smoothed version of G, $\bar{g}$ the signal taken over the entire session, and $\tilde{G}$ the normalized signal. $\tilde{G}$ was used for the entire analysis whose normality could not be assumed based on the Shapiro-Wilk test.

The Friedman test suggested a significant difference between different stages of the in vivo TSST ($\chi^2 = 49.1$, $p < .001$). The pairwise Wilcoxon tests showed that the GSR during the in vivo TSST was significantly higher than baseline ($p < .001, p.adj < .001$), preparation ($p = .002, p.adj = .017$), and resting state ($p < .001, p.adj < .001$). Thus, H1 could be confirmed.

The Friedman test also showed significant differences for different stages of the N3 ($\chi^2 = 25.25$, $p < .001$), N6 ($\chi^2 = 27.55$, $p < .001$), and N15 ($\chi^2 = 31.55$, $p < .001$). The pairwise Wilcoxon tests revealed the VR TSSTs were significantly higher than their baseline (N3: $p < .001, p.adj < .001$, N6: $p < .001, p.adj < .001$, N15: $p < .001, p.adj = .003$) and the resting (N3: $p = .001, p.adj = .006$, N6: $p = .001, p.adj = .006$, N15: $p = .001, p.adj = .006$) states. Thus, H2 can be supported by GSR data.

The comparison between in vivo TSST and N3 for all participants using the Friedman test suggested a significant difference between the conditions ($\chi^2 = 94.6$, $p < .001$). The Wilcoxon test suggested that the GSR values were significantly higher ($p < .001, p.adj = .009$) during the in vivo exposure compared to the N3 for all participants. When looking at the N3 when it was the first TSST of the VR session (see Figure 5.2(c)), the GSR values during the test did not differ significantly ($p = .313, p.adj = .31$). Thus, H3 could not be supported for all participants but rather for those who experienced N3 as the first VR TSST. In other words, N3 as the first TSST evoked as high SA as in vivo TSST. Besides, the GSR of the VR first users did not differ significantly from the rest of the participant, considering N3 as the first TSST in VR ($p = .643, p.adj = .64$).

The comparison between N3, N6, and N15 did not show any significant effect of VA size when considering all TSSTs using the Friedman test ($\chi^2 = .75$, $p = .69$) or only the first one in VR using the Kruskal-Wallis test ($\chi^2 = 1.22$, $p = .54$) (see Figure 5.2(d)). Thus, H4 could not be supported by GSR data.

Salivary Cortisol

The salivary samples were normally centrifuged directly, unless the experiment of the day was finished after the laboratory's working hours. In this case, they were stored in a refrigerator with a temperature of 5°C to be analyzed on the following day. The samples were centrifuged in a sigma type 4 K 15 centrifuge for 10 minutes at 4000 rpm (revolutions per minute) at a temperature of 4°C. The centrifugate is stored in a -80°C freezer and the samples were stored at this temperature after centrifugation for max. 3 weeks. The analysis of the cortisol kits (from DRG, Marburg) was measured on the RIA star of Packard (now Perkin Elmer). Quality controls from Bio Rad (name: Lypocheck Immunoassay Plus, level 1-3) were used and treated as the saliva samples to be analyzed. The normality of the cortisol data could not be assumed using the Shapiro Wilk test. The Grubbs test on these data detected one data point in VR baseline (value=10) and two data points in VR Rest (values: 19.7 and 3.8) as outliers which were removed for the analysis.

Based on the Friedman test, the cortisol values were significantly different during different stages of the in vivo exposure ($\chi^2 = 12.04$, $p = .007$). The post-hoc test revealed that the TSST was significantly higher than the preparation ($p = .002, p.adj = .01$) and the rest ($p = .002, p.adj = .02$) which support H1.

The Friedman test also showed that different stages of the N3 ($\chi^2 = 28.36$, $p < .001$), N6 ($\chi^2 = 25.92$, $p < .001$), and N15 ($\chi^2 = 28.78$, $p < .001$) were significantly different. The pairwise Wilcoxon tests suggests that the resting state is significantly lower than N3 ($p = .002, p.adj = .007$), N6 ($p = .006, p.adj = .017$), and N15 ($p < .001, p.adj < .001$) which partially support H2.

The comparison between in vivo and N3 for all participants using the Friedman test showed a significant difference between conditions ($\chi^2 = 57.25$, $p < .001$). The pairwise comparisons using Wilcoxon test revealed that cortisol values were significantly higher during in vivo TSST compared to N3 for all participants ($p = .002, p.adj = .037$) as well as those who experienced it as the first TSST of the VR session ($p = .007, p.adj = .007$) (see Figure 5.2(e)). Thus, H3 could not be supported. Similar to HR and GSR, the cortisol values of the VR first users did not differ significantly from the rest of the participant ($p = .736, p.adj = .74$).

We could not find a significant effect of VA size on cortisol data using the Friedman test when considering all TSSTs ($\chi^2 = 1.3$, $p = .52$) or only the first one in VR using the Kruskal-Wallis test ($\chi^2 = 1.27$, $p = .53$). However, as it can be seen in Figure 5.2(f), a trend towards higher cortisol for smaller VA size (i.e., N3) can be observed.

Head Direction

Looking at the frequency of the ray hits across all VR conditions showed that the participants' heads were directed towards the main VH 59.54% of the times, after that comes the female (18.53%) and male (4.23%) VH of the N3 condition (who are also present in N6 and N15 conditions). Thus, H5 could not be confirmed. In order to examine H6, we concentrated only on the N3 condition. To do so, we calculated the avoidance rate using the following formula:

$$avoidance_{VHs} = \frac{total_s - looked_{VHs}}{total_s} \tag{5.2}$$

Where $avoidance_{VHs}$ is the avoidance per VH and per subject, $total_s$ for each subject is the total number of times which the ray hit any VH, and $looked_{VHs}$ is the number of times the subject has looked at the specific VH. Since the Shapiro-Wilk test did not show the normality of the data, a Spearman's rank correlation test was performed between the avoidance rates and the subjects' total SPDQ score. The results did not show a significant correlation between these values ($rho = -.004, p = .97$). Additionally, LSAS ($rho = .04, p = .71$) values were not significantly correlated with the avoidance rates. Furthermore, the attention/avoidance rates to/from the male or female VHs as well as the SPDQ scores were not significantly different across female and male participants. Thus, H6 could not be confirmed.

Mental Arithmetic Performance

The performance of the participants in the mental arithmetic part of the VR TSSTs were recorded. The total number of attempts as well as the error and success rates (total number of incorrect or correct answers divided by the total number of attempts) are shown in Table 5.1. There were no significant differences for the total number of attempts, error, and success rates between conditions.

Questionnaires

Self-Assessment Manikin (SAM)

By depicting the mean responses of all participants (see Table 5.2) in a pleasure x arousal space and based on the Russell's circumplex model of affect [303], we can conclude that our participants were distressed during in vivo and VR TSSTs. The results of the Friedman test suggested a significant difference between in vivo and VR TSSTs and their different stages for arousal ($\chi^2 = 38.99$, $p < .001$) and pleasure ($\chi^2 = 49.23$, $p < .001$). However, the pairwise comparisons indicated that the arousal and pleasure values of the VR and in vivo TSSTs were not significantly different from one another. Besides, we could not find a significant difference between dominance values ($\chi^2 = 16.6$, $p = .055$).

State Trait Anxiety Depression Inventory-State (STADI-S)

The mean values of STADI-S for all participants can be seen in Table 5.2. These values represent a high anxiety in all conditions. The Friedman test ($\chi^2 = 44.23$, $p < .001$) suggested a significant difference between the conditions but the pairwise comparisons were not significantly different.

Self-Estimated Fear

Table 5.2 shows the total fear scores of all participants per condition. The Friedman test was used which suggested a significant difference between the conditions ($\chi^2 = 12.43$, $p = .006$) but the post-hoc test did not show any significant differences. We also looked into the two parts of the TSST, namely the 5 minutes speech and 5 minutes arithmetic solving. The Wilcoxon test did not show any significant differences between the fear reported in these two parts. We ran also the Friedman test separately for each part. Only for the speech part, a difference between the conditions ($\chi^2 = 21.18$, $p < .001$) were found but the post-hoc test

Table 5.1: Mental Arithmetic Performance. Mean (SD) of error and success rates as well as the total number of attempts per condition.

Condition	Error Rate	Success Rate	Total Attempts
N3	.08 (.06)	.92 (.06)	50.46 (19.98)
N6	.1 (.09)	.9 (.09)	53.2 (23.62)
N15	.14 (.24)	.86 (.24)	52.46 (23.48)

did not show any significant difference.

State Social Anxiety (SSA)

Table 5.2 shows the total SSA score of all participants per condition. The Friedman test did not show any significant difference between conditions.

Co-Presence and Social Presence (Co-SP)

As it can be seen in Table 5.3 only the (Co-)Presence of the VHs scale was rated above the medium in all conditions. This suggests that the participants were aware of the presence of the VHs in the VE. The Friedman test did not show any significant difference between the conditions for any of the four factors of this questionnaire.

Simulator Sickness Questionnaire (SSQ)

We looked into the three distinct symptoms of the simulator sickness namely Oculomotor, Disorientation, and Nausea and could not find any significant difference between their pre- and post-measurements (see Table 5.4) using the Wilcoxon test. We found, however, a significant increase in the SSQ total score (TS) (Z = -2.02, p = .04) from pre- to post-measurements. Nevertheless, the mean score of 21.19 measured after the experiment indicates still a relatively low simulator sickness. Besides, no incidents or symptoms of simulator sickness were reported by the participants.

Igroup presence questionnaire (IPQ)

Table 5.5 shows the measurements of the General Presence (PRES), Spatial Presence (SP), Involvement (INV), Experienced Realism (REAL), and the IPQ total score. These values are in line with the similar measurements for VR using HMD [279, 198]. Since Price et al. [279] suggested that IPQ total presence scores are associated with the fear experienced in VR, we ran multiple Spearman's rank correlation tests between IPQ scores and the self reported fear values. No significant correlation was observed. In addition, we looked into the correlation of the GSR data (suggesting an objective measure for arousal) with the IPQ scores. A significant positive correlation between N6 and INV (rho=.527, p = .008) was found.

Table 5.2: M (SD) of SAM (pleasure (P), arousal (A), dominance (D)), STADI-S, self-estimated fear (Fear), and SSA.

Questionnaire	in vivo	N3	N6	N15
SAM: P	3.54 (1.41)	4.25 (1.36)	4.08 (1.5)	3.88 (1.19)
SAM: A	6 (2.09)	6.54 (1.84)	6.42 (1.69)	6.58 (1.74)
SAM: D	4.96 (1.43)	4.83 (1.52)	5 (1.67)	4.96 (1.57)
STADI-S	33.46 (7.59)	32.79 (9.04)	34.83 (9.28)	34.13 (8.06)
Fear	41.92 (21.66)	31.04 (24.7)	32.08 (23.2)	32.17 (22.52)
SSA	28 (14.78)	21.42 (17.61)	21.38 (18.04)	18.5 (14.73)

Table 5.3: Co-Presence and Social Presence, M (SD)

Scale	N3	N6	N15
Presenter's Reaction to VHs	2.33 (1.28)	2.11 (1.15)	2.28 (1.14)
Perceived VHs' Reaction	2.02 (1.17)	1.98 (1.05)	2.04 (1.06)
Impression of Interaction Possibilities	2.15 (1.16)	2.01 (1.26)	2.11 (1.15)
(Co-)Presence of the VHs	2.9 (1.58)	2.78 (1.59)	2.83 (1.48)

Table 5.4: SSQ

Scale	Pre: M (SD)	Post: M (SD)
Nausea	11.13 (12.15)	20.27 (23.92)
Oculomotor	12.95 (15.72)	20.21 (14.94)
Disorientation	6.38 (17.88)	12.18 (15.53)
Total Score	12.47 (15.43)	21.19 (18.71)

Table 5.5: IPQ

Scale	M (SD)
PRES	4.21 (1.89)
SP	19.88 (5.03)
INV	13.71 (2.26)
REAL	12.5 (1.9)
TS	50.29 (8.52)

5.4.5 Discussion

We assessed anxiety responses to three versions of TSST for VR (with 3, 6, or 15 VHs) and for an in vivo TSST with three persons. Several objective and subjective measures were employed for this purpose whose results are summarized in this section.

The HR analysis showed a significant increase from the in vivo baseline to the TSST and then a significant decrease from the TSST to the resting state. In VR, these changes could be observed in a more moderate pace, which resulted in a significant decrease from the TSST to the resting state. Moreover, N3 induced the highest values for HR during the entire VR session, which was significantly higher than N6 and N15 during the first VR TSST. These results confirms findings from previous work (e.g. [236]), which have shown that an audience of a smaller size can induce stronger SA responses in particular visible in the HR.

A significant arousal could be also observed in the GSR values during both VR and in vivo TSSTs, where VR was not significantly different from the in vivo TSST. The audience size did not affect the GSR values significantly, as they were high for all three sizes (N3, N6, and N15) during the first TSST in VR.

The salivary cortisol values showed a significant increase from the preparation to the in vivo TSST and a decrease to the resting state. The in vivo cortisol values were also higher than in VR. Moreover, the changes in the number of VHs did not significantly affect the cortisol values. Concerning cortisol, it is worth noting that some studies have noted the biological differences of SAD patients. For instance, Furlan et al. [122] reported on hypothalamic-pituitary-adrenal axis abnormalities among a group of SAD patients. This abnormality has been observed in the form of a decrease in the salivary cortisol in response to a psychologically stressful task (i.e., a speech task). Similarly, Klumbies et al. [192] classified their participants into two groups of responders and non-responders while analysing the salivary cortisol responses. Therefore, our salivary cortisol measures may not be directly comparable to data from other VR TSST studies on healthy participants.

The VR TSSTs offered the possibility to track the participants' heads. Although eye-tracking is a more reliable method for measuring the visual attention, it is still possible to approximate the gaze by tracking the head. The results of this analysis suggest that participants' heads were most of the times directed towards the main VH. Neither the VHs' avoidance rates nor the attention rates (the frequency of head directions towards the specific VHs divided by the total number of recorded head directions) were significantly correlated with the SAD severity of the participants (i.e., total SPDQ score). Thus, we conclude that the participants were equally and mostly looking at and likely paying attention to the main VH and neither avoidance nor attention rates in head direction were informative measures in the present study.

Moreover, the TSST offers a cognitive variable by means of the assessment of the participants' mental arithmetic performance. Surprisingly, the success rates for the given mental task were quite high (86-92%). That is, across all three VR TSSTs, the participants made only a few errors (8-14% of the times), which may indicate a ceiling effect that may have obscured differences between conditions. Furthermore, there was no significant correlation between the SPDQ scores and the error rates. Future studies may consider to use a time limit to respond in order to make the task more difficult and therewith avoid ceiling effects.

Besides the above mentioned objective measures, subjective questionnaires were evaluated. The results of the SAM suggested that the participants were distressed during VR and in vivo TSSTs. The results of the STADI-S also indicated a state of low anxiety for all conditions with no significant difference between them. The participants, moreover, did not rate the TSSTs as extremely stressful. As they estimated their fear during the tests on a scale of 0-100 on average 41.92 for in vivo, 31.04 for N3, 32.08 for N6, and 32.17 for N15. In order to assess subjectively different aspects of the speech part of the TSST, we developed the State Social Anxiety questionnaire. The mean responses to this questionnaire are also rather low as they range from 18.5 (for N15) to 28 (for in vivo) on a scale of 0-116. It has to be mentioned that a factor analysis has to be conducted in a separate study to assess different items of this newly created questionnaire. The results of the co-presence and social presence questionnaire which was used to measure the perceived feeling of VHs' presence in the VE suggests that the participants were aware of the presence of the VHs with no significant difference across different VR TSSTs.

In contrast to our priori hypothesis, we did not find that larger audiences in VR increased the stress effects in participants. In contrast, we found that smaller audiences increased the stress responses (in particular in HR). This finding contradicts the previously mentioned SIT [205, 206], that predicts that the number of persons should increase the impact felt by the individual with a negatively accelerating power function. Our results, however, are in line with Asch's theory [18, 19] which concludes that a group size of three is sufficient to feel the full impact of the group. While SIT does not foresee a limit to additional impact of larger group sizes (i.e., there is always an impact of additional group members) and Asch's theory does not predict any additional effect beyond a group size of three, the Social Influence Model (SIM) [358] suggests that the group size is important but up to some limit. Therefore, the second and third member of the group have greater additional impact than the first member, the fourth member has less social influence that the third member and once a group reaches the size of five, the individual impacts are not felt anymore and therefore any additional increase beyond this size has no additional impact. Furthermore, a meta-review [47] of the theories and studies on the relationship between the group size and conformity argues that the task and setting play an important role in forming this relationship. Thus, we can conclude that for delivering a speech and performing mental arithmetic calculations (TSST) in VR, three VHs had an impact and induced SA. Yet, additional VHs could not induce further SA beyond this as the group size of six and fifteen were large enough for VHs to lose their individual impacts. It remains open if this finding is associated to the VR conditions, and if a change of the audience size in the real TSST would have resulted in a similar effect. Thus, a clear limitation of the current study is the lack of inclusion of six and fifteen real audience. Limited range of group size is indeed a common limitation of group size studies. Most studies have employed a group size of two to four [47] and none other than Asch [18, 19] have used a group size of greater than nine. Therefore, a replication of the experiments with varied virtual as well as real audience size remains a topic for further research. However, the results suggest that perception of VHs in VR is quite different from perception of people in the real world. Certain specifics of VR technology might have an impact on the perception of VHs. For instance, the still limited field of view or resolution of current HMDs might degrade the perception of VHs. In particular, more distant VHs would appear with less realism since less pixels are available to display their behaviour or facial expressions. This might have a direct influence on the participants perceived social presence of those VHs. However, future work needs to be conducted to better understand the effects of certain VR characteristics on the perception of VHs and the perceived SA.

5.5 Conclusion

In this chapter, we presented an adaptation of TSST for in vivo and VR to investigate the effects of a change in the number of VHs on perceived SA. The analysis of different physiological and subjective measures revealed different aspects of the exposure to a psycho-socially stressful situation. Physiological arousal could be observed with VR inducing SA yet less than in vivo TSST. STADI-S and SAM questionnaires also showed a state of anxiety experienced during the experiment. An effect of the change in the number of VHs could be observed only in the HR as the average HR in response to three VHs was higher than the exposure to six and fifteen VHs. In future work, one could replicate the experiments, but also vary the audience size in the real-world condition. Furthermore, we plan to explore the effects of certain VR characteristics on the perception of VHs and the perceived SA since those aspects have a direct impact on the interaction with VHs in VR in general.

6 Chapter 6 Virtual Reality for Individuals with Occasional Paranoid Thoughts

Individuals with paranoia often experience a high level of self-criticism and negative emotions. Guided Compassion Focused (CF) imagery has shown to be successful in reducing these negative emotions and paranoid thoughts. However, some individuals have difficulties with CF imagery. By enabling a sense of presence, immersive VEs can overcome these limitations and induce specific emotional responses to support the development of self-compassionate feelings. In this chapter, an immersive CF (CF-VR) was compared with a controlled VR condition in a student sample of N = 21 participants with slightly elevated symptoms of paranoia. A virtual mission on the moon was designed and implemented to induce self-compassionate feelings with the help of interacting with a space nebula that represented the power of compassion. The results show that the CF-VR intervention was well accepted and effective in reducing state paranoid thoughts. Worry decreased significantly within the CF-VR group, while self-compassion increased.

6.1 Introduction

Individuals with paranoia exhibit emotional disturbances such as strong fluctuations in anxiety and depressive symptoms along with self-esteem instability and high levels of self-criticism [363, 389, 114, 163]. We need more effective treatments for paranoia, as effect sizes for classical CBT interventions are only small to medium [239, 376]. A recent meta-analysis suggested that interventions focusing on causal mechanisms (causal interventionist approach) bring about larger changes in delusions than classical CBT interventions [239]. As it makes sense to assume an over-activated threat system in paranoia, interventions that target connectedness, kindness, and soothing might be particularly helpful for these individuals.

In CF therapy it is hypothesized that almost all psychological disorders can be attributed to a general imbalance of three central emotion regulation systems hosting threat, drive and excitement, and soothing [131]. All these systems serve important functions: the threat system identifies dangers and initiates according stress responses, the drive and excitement system regulates approach behavior towards rewarding stimuli and activities (including social rank competition), and the soothing system promotes bonding and feelings of (social) safeness, warmth and connectedness. In most psychopathologies, the soothing system seems to be hypofunctional, hypothetically often due to lack of secure attachment relationships or even trauma in important developmental phases. Hence CF therapy aims at strengthening this system using different therapeutic strategies, such as developing sympathy and kindness towards oneself - especially in the face of suffering or failure. Among many other techniques, it draws upon meditation practices and imagery to evoke compassion. According to Gilbert [132] compassion is part of a caring, nurturing and sharing social mentality directed both at the self and others, and this mentality seems to be underdeveloped in paranoid individuals. Therefore, the practice of Compassion-Focused Imagery (CFI), such as imagining a caring and nurturing being conveying compassion towards the self, might be very useful for this patient group. However, previous studies of guided CFI in patients affected by psychosis [234, 20] have shown that some individuals had difficulties with the imagery. For example, undesirable and even scary images can appear during treatment (e.g. one patient generated a ‘ugly, repulsive, super-human’[234]). Nevertheless, in sub-clinical populations, CFI has been shown to reduce paranoid ideation and negative emotions [224].

VR might help to overcome limitations of guided imagery by providing controlled virtual environments that support the development of compassionate feelings while avoiding adverse patient imagery. VR interventions for psychotherapy are known to have several beneficial effects. For instance, VRET has comparable effects as in vivo ET for anxiety disorders [277]. Much less is known about the beneficial effects of VR in the treatment of psychosis. A first pilot trial showed that VR can reduce delusion conviction and anxiety in real-life social situations with large effects [275, 113]. Besides the reduction of negative affect via exposure, it seems to be possible to evoke positive, warm feelings using VR. A single session pilot trial in depressed subjects showed that receiving compassionate gestures generated by the adult-self towards a child-self lead to significant reductions in depression severity and self-criticism, as well as to a significant increase in self-compassion from baseline to 4-week follow-up [99].

For our intervention, we loosely drew upon a CFI exercise (compassionate colors) from the Compassion-Focused Therapy book for practitioners [131]. As a background scenario we chose space, because it can evoke an overview effect (putting one’s existence into a larger perspective), and it elicits feelings of awe, connectedness with humanity, and self-transcendence [406]. The latter phenomena have a striking conceptual overlap with Neff’s definition of self-compassion [252, 251], which involves three major components: common-humanity (one’s own destiny and suffering being part of the broader human experience), mindfulness (vs. over-identification with problems), and kindness (vs. being judgmental). Hence, we combined a space scenario, where individuals were sent by a mission control (audio via headphones) to the moon (see Figure 6.1(a)) in order to explore a mysterious nebula (Figure 6.1(b)) that harbors the power of compassion for oneself and humankind (CF-VR). Participants were gradually guided into opening for feelings of kindness, warmth, wisdom, courage, and strength, while also interacting with light-points within the nebula that intensified their glow by touch (sym-

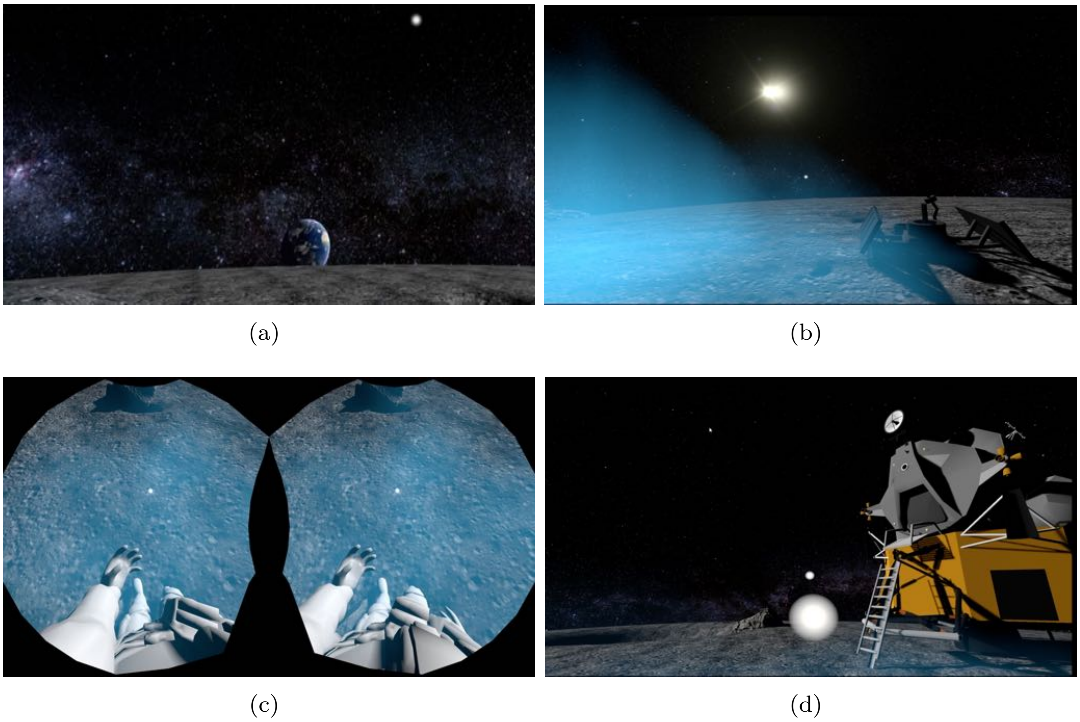


Figure 6.1: (a) The view of earth (b) Blue colored nebula appearing on the moon (c) The participants' virtual body (d) The Apollo 11 lunar module

bolizing the increasing power of compassion). The exact same scenario but with a neutral audio, that instead focused on objective qualities of the nebula and the environment (similar to mindfulness practice), was used as control condition. Both audios and scenarios were about 10 minutes long, and the pausing and pace of the evolvement of the nebula on the moon, alongside with the according instructions, were equal for both conditions.

The study presented in this chapter aimed at piloting the CF-VR vs. control VR in a student sample of N = 21 participants with slightly elevated symptom levels of paranoia. Those were individuals who indicated to experience paranoid thoughts at least once per month on the frequency scale of the paranoia checklist [116]. Our main outcomes for this pilot trial were paranoia and self-compassion. Furthermore, we examined changes in emotions (both positive and negative).

6.2 Methods and Materials

6.2.1 The Space-Compassion VR application

In an effort to make the space experience more vivid, the space environment was created as an animated game object with rotating planets (with clouds), day/nighttime animations of the planet earth, and the milky way background. The user was placed on the moon surface, close to the Apollo 11 (see Figure 6.1(d)) lunar module (taken from the official NASA website) to create the impression of being able to take refuge or leave the moon, and still being connected to other people. The VR experience was individualized for each participant by asking them to choose a color they associate with wisdom, warmth, and non-judgement qualities to be used as the color of the self-compassion nebula. To make the nebula interactive, the users were able to manipulate the intensity of its glowing particles. To create a sense of embodiment and agency, the participants were given a virtual body (an astronaut's suit, see Figure 6.1(c)) which was synchronized with that of the user using an HTC Vive Pro VR HMD, two controllers (for the hands), and two trackers (for the feet). The Space-Compassion VR application was developed with the Unity 3D Engine version 2018.3.11f1.

6.2.2 Procedure

This study was approved by the local Psychological Ethics Committee of the Center for Psycho-social Medicine at UKE. The study was performed in the VR Lab at the UKE. The recruitment was accomplished by means of mailing lists, university boards, and online advertisements. Prior to the experiment, all participants underwent an online screening (T0) to check the eligibility criteria (i.e. at least one paranoid thought per month as measured with the paranoia checklist [116]. Furthermore, demographic data was assessed, including age, sex, employment status, education, history of mental disorders, central nervous medication, and visual impairment (the latter lead to exclusion). Current or past mental disorder was no exclusion criterion. A minimum age of 18 was required, but there was no upper age limit. About one week after the online screening, the laboratory assessment took place. During the laboratory session, all outcomes (paranoia, self-compassion, emotions) were assessed three times: (T1) first directly at the beginning of the session, (T2) after participants were randomized and received a psycho-educative video, depending upon their assignment this was either a neutral video about space (e.g., what is a nebula?) or a video introducing the concept of compassion (e.g., why compassion is important for mental health?), and (T3) third after the VR exposure (CF-VR vs. control VR). At the end of the intervention, we asked participants to rate their perceived personal benefit from the intervention.

6.2.3 Participants

Mean age within the CF-VR group ($n = 12$) was 28.6 years ($SD = 13.7$). Forty-two percent were female, average years of school was 12.2 ($SD = 0.83$), speaking for a high school education according to German standards. In the control VR group ($n = 9$), mean age was 22.1 ($SD = 2.1$) and 56% were female. The mean of school years was 12.1 ($SD = 0.33$). There were no significant differences between the groups in any of these demographic variables ($p > .05$). In total four participants reported a diagnosed current or past mental disorder. Three of them were in the CF-VR group (one with two past diagnoses: depression and eating disorder; one with past post-traumatic stress disorder; one with past mental disorder - not specified and current media addiction). One was in the control VR condition (past anxiety and eating disorder).

6.2.4 Questionnaires

Paranoia

We used a brief version of the German paranoia checklist with three items that has been found to be highly change sensitive [313] (items: 'I need to be on my guard against others', 'People deliberately try to irritate me', 'Strangers and friends look at me critically'), rated on three dimensions: 1) applicability/ predominance of this thought at the moment [0 = 'completely disagree', 11 = 'fully agree'], 2) distress [0 = 'doesn't bother me', 11 = 'bothers me a lot'], and conviction [0='not convinced', 11= 'completely convinced']. Cronbach's alpha was good to excellent (applicability: $\alpha = .85$; distress: $\alpha = .85$; conviction: $\alpha = .90$).

Self-compassion

We used a brief experimental version of the original Self-Compassion Scale (SCS) [251] as used by Ascone and colleagues [20]. The original SCS measures self-compassionate feelings on three dimensions: self-kindness, common humanity, and mindfulness. The brief version contains one item from each of the subscales (i.e., self-kindness: 'I'm tolerant of my own flaws and inadequacies', common humanity: 'I try to see my failings as part of the human condition', and mindfulness: 'When I'm feeling down I try to approach my feelings with curiosity and openness'). Internal consistency (Cronbach's alpha) was acceptable ($\alpha = .64$).

Emotions

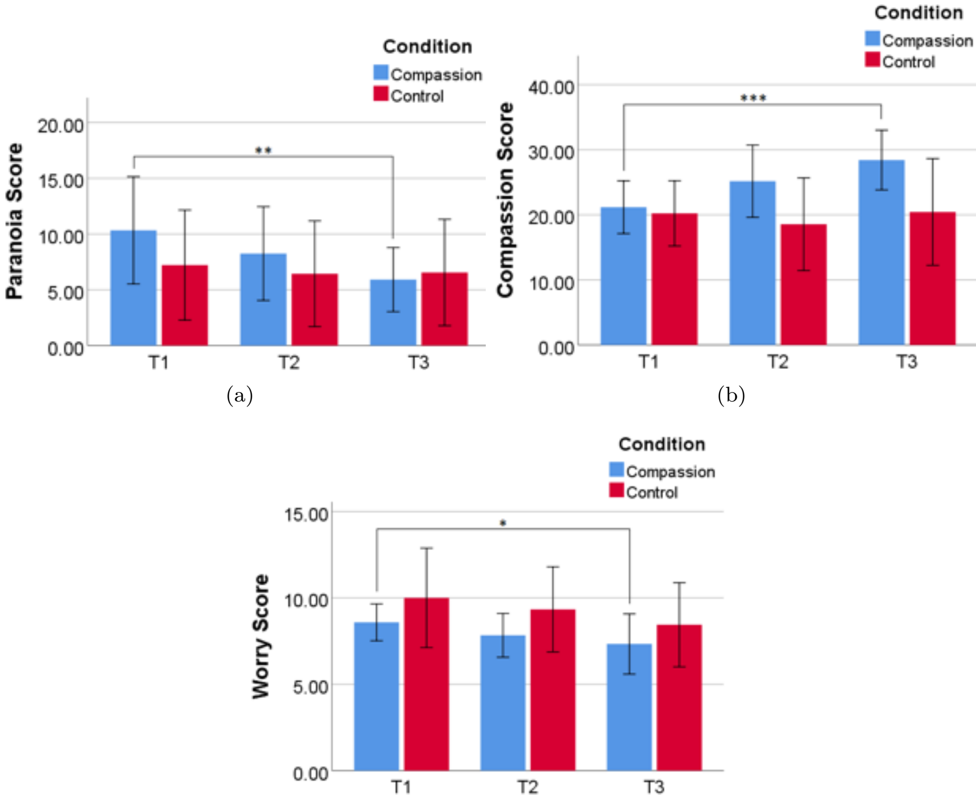
In order to assess different types of emotions, we used the state part of the STADI [207]. See Chapter 3, Section 3.4.1 for more information about this questionnaire.

Self-rated benefit: psycho-education and VR intervention

We assessed the personal benefit individuals experienced during the intervention using a 5-star-rating schema (1 star = rather not helpful; 5 stars = very helpful). The same rating format was used to assess the perceived benefit from the psycho-education (about compassion/ space, depending on the group assignment).

6.2.5 Statistical Analyses

As assumptions for parametric testing were violated, we used non-parametric within-group difference analyses applying the Friedmann test, in order to analyze overall changes over time within the CF-VR or control VR group. A significant omnibus test was followed up by Bonferroni-adjusted exploratory post-hoc pairwise comparisons between the different assessment points (T1, T2, T3). To compare perceived personal benefit ratings for both conditions we computed the non-parametric Mann-Whitney-U-Test, with a two-tailed significance level of .05.



(c)

Figure 6.2: (a) Paranoid symptoms (b) Self-compassion (c) Worry

6.3 Results

6.3.1 Paranoia

The following pattern of results (see Figure 6.2(a)) was identified within the CF-VR group. The applicability/ predominance of paranoid thoughts significantly decreased within the CF-VR group (χ^2 = 14.06, $p < .001$). Looking at pair-wise post-hoc comparisons for all three laboratory assessment time-points, this main effect was attributable to a significant change from T1 to T3 ($Z = -3.06$, $p = .007$), suggesting that only a combination of psycho-education and CF-VR related to a significant symptom change. No other pairwise-tests reached significance ($p > .05$) There were no significant effects for paranoia distress ($\chi^2 = 0.28$, $p = .871$) or conviction ($\chi^2 = 2.14$, $p = .343$). There were no significant effects within the control VR group (paranoia – applicability/ predominance: $\chi^2 = 2.38$, $p = .305$; paranoia – distress: $\chi^2 = 4.00, p = .135$; paranoia – conviction: $\chi^2 = 1.37$, $p = .504$).

6.3.2 Self-compassion

There was a significant increase in self-compassion (see Figure 6.2(b)) $\chi^2 = 16.93$ ($p < .001$). Post-hoc analyses revealed that the increase took place between T1 and T3 ($Z = 3.89$, $p <$

.001), while there were no significant changes between any other time-points ($p > .05$), again suggesting that a combination of psycho-education and the CF-VR drove the effect. There was no significant effect for self-compassion in the control VR group ($\chi^2 = 3.16, p = .206$)

6.3.3 Emotions

There was no significant effect within the CF-VR group concerning nervousness ($\chi^2 = 5.44$, $p = .066$). For worry (see Figure 6.2(c)), a significant result was obtained ($\chi^2 = 6.70$, $p = .035$), which was attributable to changes from T1 to T3 ($Z = 2.45$, $p = .043$). No other differences between time points were identified for worry. The variable euthymia did not change significantly ($\chi^2 = 2.67$, $p = .264$), nor did dysthymia ($\chi^2 = 1.27$, $p = .529$). Within the control condition, there were no significant effects on any of the emotion subscales (nervousness: $\chi^2 = 3.92$, $p = .141$; worry: $\chi^2 = 1.50$, $p = .472$; euthymia: $\chi^2 = 0.83$, $p = .661$; dysthymia: $\chi^2 = 1.40$, $p = .497$).

Self-rated benefit: psycho-education and VR intervention

Overall, the ratings indicated medium to high perceived personal benefit from the psycho-education (CF: $Mdn = 4.00$, $SD = 1.00$; control: $Mdn = 3.00$, $SD = 0.88$) and VR intervention (CF-VR: $Mdn = 3.50$, $SD = 0.78$; control VR: $Mdn = 4.00$, $SD = 0.97$). There were no significant group differences (perceived benefit psycho-education: $U = 33.0$, $p = .119$; perceived benefit intervention: $U = 48.0$, $p = .650$).

6.4 Discussion

In sum, the CF-VR (including the psycho-educative part) was associated with changes in state paranoia (decrease), self-compassion (increase), and worry (decrease), while no changes were identified within the control VR. However, all observed changes also included the psycho-educative part on self-compassion, speaking for the importance of explaining the rationale for the intervention to participants in order to evoke significant change. The absence of effects on the paranoia distress and conviction subscales raises the question of clinical relevance of the observed decreases in state paranoia. There are, in our view, two possible explanations for this. First, as our sample was only sub-symptomatic, it is possible that there were bottom effects. Second, it is also possible that distress and conviction are harder to tackle, and that repeated training may be necessary to evoke according change of the affective quality of paranoid thoughts. The effect on self-compassion shows that our intervention indeed alters the target mechanism. The decrease in worry appears promising, as research suggests heightened levels of anxiety and perpetuating worry as important pathogenic mechanisms for paranoia (e.g. [112, 115, 353]). Future analyses should focus on potential mediation effects in order to infer conclusions about mechanisms of change. Of interest, both interventions including psycho-education were rated as at least moderately helpful, speaking for good acceptance.

A clear limitation of our study is the small sample size, which is why we refrained from parametric tests. Replication of the findings in a larger sample, while contrasting both groups (i.e. repeated measures ANOVA) is a necessary next step. Representativeness of the sample is also questionable, as it mainly consisted of highly educated, young individuals (students). Nonetheless, it is known that student populations are often more delusion-prone than the general population [225] and are in the typical age-range for a first manifestation of psychosis. Inclusion of the standard CFI and a wait-list as additional control conditions in order to examine the advantages of CF-VR remains a topic for further research. In addition, standardized psychiatric diagnostics should be included in future trials.

Based on the qualitative data from our study, the intervention was well accepted and the topic was deemed relevant: "I liked the fact that through the course of the study I learned something about myself and human behavior in general and that it conveyed a positive view upon myself.". However, the VR experience could be improved by including more interactions and more customization (e.g., choice of the guide's voice, color attributes, etc.).

We conclude that our novel space compassion-focused VR has potential to ameliorate paranoid symptoms and worry, and improve compassion. Repeated training in the VR (vs. the standard CFI and/or control VR), larger samples, and replication studies in patients with persecutory delusions appear to be reasonable and promising next steps.

7

Chapter 7

Towards Gamified Alcohol Use Disorder Therapy in Virtual Reality

In this chapter, a PACT (people, activities, contexts, and technologies) analysis for Alcohol Use Disorder (AUD) therapy in VR is provided. Based on this analysis, a gamified VR-based AUD therapy application was developed which consists of three mini-games and employs Approach Avoidance Training (AAT) and Cue-Exposure Therapy (CET) methods. The games were realized in the context of a virtual supermarket, which is considered as a relapse-risky environment. The aim is to help AUD patients practice avoiding alcohol first in a VR-based simulation and later in a real supermarket. This application and its evaluation by clinical experts and AUD patients during two focus groups will be presented in this chapter. In addition, in preparation for a long-term clinical study, a usability study was conducted with 13 healthy participants. The results show that the VR game was enjoyed, increased the motivation, and fewer errors were made than in the comparable non-gamified application.

7.1 Introduction

AUD is characterized by compulsive use of alcohol despite knowing its harmful physiological and psychiatric consequences [81, 93]. According to the 2015 National Survey on Drug Use and Health (NSDUH) [7], 15.1 million adults above 18 years old and an estimated 623,000 adolescents ages 12–17 fulfilled the criteria of AUD in the USA. High alcohol consumption on a regular basis is considered risky and can cause serious health problems. In fact, more than 85,0000 people die from alcohol-related causes annually, making alcohol the third leading preventable cause of death in the US, but other countries are affected by a high alcohol consumption as well [197].

In Germany, AUD is recognized as the most common subcategory of the mental and behavioral disorders among adults [197]. Approximately half a million hospital stays in Germany resulted from alcohol in 2012. In 2015, over 100.000 men and 36.000 women were diagnosed with AUD. The highest rates for risky alcohol consumption were among 55-64 and 18-24 years old men and 45-54 years old women. Underage drinking is also quite common. In 2015, acute intoxication was reported for both males and females under 18 with the highest rate reported for 15-17 years old males.

However, only about 17% of AUD patients in Germany seek therapy [197]. In the US, less than 7% of the adults and less than 6% of the adolescents receive treatment [7]. The biggest obstacles for starting therapy are the lack of awareness of the problem, shame, and the desire to solve the problem alone. In the end, only 2-4% of the patients complete a rehabilitation program [197].

Therefore, it is important to not only increase the awareness about this problem but also improve the current therapy methods and create innovative alcohol therapy methods, which are more accessible and effective compared to the traditional ones available today.

Different therapy methods such as CBT can help patients overcome their AUD. CET is one of these methods in which patients are confronted with certain stimuli which trigger their craving for alcohol. The goal is to repress salience of those stimuli and to prevent or reduce the dysfunctional approaching behavior [228]. CET is a special form of ET. In ET, which has been successfully applied for anxiety disorders [228], patients are carefully exposed to specific stimuli which trigger their symptoms. Overtime, this will help them to gradually control their reactions. Similarly in CET, patients are exposed to certain stimuli (or cues) to trigger their craving (disturbed approach behavior). Exposure to cues increases the craving, which can also be observed physiologically e.g. in a change of the cardiac frequency, electrodermal activity, or the salivation. CET can be used for treatment of various types of substance (e.g., alcohol, nicotine, cocaine, etc.) addiction as well as other psychological disorders such as bulimia, binge eating, shopping addiction, and pathological gambling. For treatment of AUD, alcohol related cues are presented and the patient is instructed to let the craving arise and acknowledge it. The patient gives regularly feedback on his or her subjective rating of the intensity of the craving. A training session is completed once the intensity is considerably decreased. In later therapy sessions CET can be combined with a training of coping skills, so the patient learns how to decline alcohol offers [228].

Another CBT method is AAT by which the patients learn to avoid alcohol by repeatedly performing avoidance behaviors towards alcohol. For example, in an explicit AAT, patients use a joystick to push the images of alcoholic beverages on a computer screen away and pull the images of non-alcoholic beverages towards themselves [324]. Originally AAT was developed as a diagnostic instrument and as a variant of an Implicit Association Test (IAT). An IAT measures the relative strength of associations between concepts [395]. The task is to sort the stimuli into four categories but only using two possible responses. Two of the four concepts are called *attributes* (usually positive-negative) and the other two are the *targets*. The task is to react as quickly as possible to which category (attribute or target) the stimulus belongs. A stimulus can be an image or word. The measured effect is the different response time for the two possible combinations of attributes and targets. The AAT was implemented as an approach-avoidance-IAT and its results show strong connections between alcohol and approach bias in heavy drinkers [395]. Wiers et al. [394] conducted a clinical AAT training study with 214 alcohol dependent inpatients, who were assigned to either training or control conditions. The control groups either did not receive training or received sham training. The participants in the training group had to push a joystick to make an avoidance movement when pictures of alcohol were presented and to make an approach movement to pictures of soft-drinks (pull a joystick). The results show that four brief sessions (15 minutes) on consecutive days changed implicit approach responses to alcohol. Moreover, a year later the treatment outcomes for patients in the training group was better (i. e., 16% less relapse rate).

These methods, however, are associated with some limitations, which have to be addressed while designing innovative therapy methods. For example, therapy sessions using the CET method with the goal to cope with alcohol cues in every day scenarios such as grocery shopping can be expensive and time-consuming. Furthermore, the AAT may be perceived as rather tedious and may not yet be as enjoyable as it could be.

This shortcoming can be addressed by implementing these behavioral therapy methods in VR. In a VE, craving can be successfully induced and has been shown to be more effective than traditional 2D cues as photographs [159]. Since the VE can easily be customized and adjusted to fit the patient's needs. In addition, the patient might prefer a VR to an in-vivo therapy setting [159]. Furthermore, in a VR session, the patients wear an HMD and can immerse themselves in another setting, whereas in reality they are still in a "safe" environment, in which alcohol is not available. Another advantage is that in VR the stimuli (or cues) can be interactive.

Finally, gamification can increase the intrinsic motivation of users [310]. Different gamification elements satisfy the key elements that create motivation such as the feeling of competence, experiencing autonomy, and social relatedness [308]. Moreover, recent studies showed that learning effects can be increased through gaming [133], and even more interestingly that gaming on a regular basis can have a positive impact on the structure of the brain [202, 201].

Some scientists have already reported applying VR for AUD therapy [48, 213, 240]. For example, Brodnick et al. [48] analyzed the subjective craving for alcohol in different VEs. The results of a controlled experiment with 40 AUD patients showed that craving for alcohol was increased in VEs in which alcoholic beverages were present (e.g., party). These results can be helpful for VR-CET. Another study [213] applied VR-CET to eight members of an Alcoholics Anonymous group. The VEs were a Japanese-style pub and a western-style bar. The results of training for eight 30min-sessions showed a decrease in the subjectively reported craving. Lastly, Metcalf et al. [240] developed a cue refusal VR-video game based on Kinect and Xbox with the goal to support cigarette and alcohol recovery. The player had to hit or kick the addiction cue images which flew towards them. The results of an experiment with 61 participants recovering from alcohol or tobacco addiction showed that on average, reported substance use decreased during the intervention period. Moreover, AUD participants in recovery showed a statistically significant increase in self-efficacy, attitude, and behavior during the intervention and a decrease in alcohol use by 75% after the study.

This chapter presents a VR-based game that was designed to enhance AUD therapy. In order to create an optimal user experience for such a VR therapeutic applications, it is essential to understand the user groups and the way they interact with the system [37]. Therefore analyzing the People, Activities, Contexts, and Technologies (PACT) beforehand can help understand the needs and elicit requirements. Thus, in the first section, a PACT analysis for AUD therapy in VR will be presented. Thereafter, the development and evaluation of the VR-based game will be described. The VE is a supermarket, which is not only a place which patients will have to visit on a regular basis in their everyday life, but also bears a high risk for a relapse. Therefore, the patients need to be prepared for this scenario. The VR game consists of different mini-games each of which includes one behavioral therapy method. The main goal of the game is to find and buy items that are written on a shopping list. Before paying for the items, the player has to earn money. Earning money can be done through sorting items into a shelf (CET) and clearing a shelf by placing the alcoholic beverages into a trash bin and the non-alcoholic beverages in a shopping-cart (AAT). In the end, the outcomes of two focus groups with clinical experts and AUD patients as well as a preliminary usability study will be presented.

7.2 PACT Analysis

The PACT analysis framework considers the contexts in which people use different technologies to perform activities [298]. Therefore, it can help understand the users' needs and elicit requirements prior to the implementation of a gamified AUD therapy application in VR. In consultation with the clinical experts at the University Hospital Hamburg-Eppendorf, CET and AAT methods were chosen to be employed in the application, since these methods are widely used in therapy sessions. Moreover, we decided to realize the application in a virtual supermarket as supermarkets are relapse-risky environments. Based on this concept, a PACT Analysis was done which is presented in this section.

7.2.1 People

The user group of a VR-based AUD therapy application is rather heterogeneous as AUD patients can have different demographic characteristics, levels of addiction, and motivation for seeking treatment. Nevertheless, they might have some aspects in common. For example, their behavior might be more impulsive and alcohol-driven [81]. Also, the level of attention when alcohol is in sight, might be lower. Because of the impulsive behavior, short term rewards will be preferred over long term rewards [81]. Not only alcohol is perceived as rewarding, but so is money [33]. Furthermore, alcohol consumption alters the brain structure in a way that impedes behavioral control, encouraging more alcohol consumption and neurodegeneration (death of neurons) [81]. Therefore, not only does alcohol destroy brain cells, but also inhibits neurogenesis (creation of new neurons) [81], resulting in a smaller brain size in AUD-patients [81]. Moreover, if they pursue a therapy for their AUD, we might be able to assume that they have a higher motivation to fight their addiction and change their behavior. Of a VR-based AUD therapeutic application, they might be a frequent user for a certain amount of time. In the beginning of a therapy program in a rehab facility, they might use it on a daily-basis.

Once discharged, the frequency of usage is up to their condition. Most likely they would not have the freedom of choosing a therapeutic application from a range of various systems, since the treatment plan is decided on by the therapists.

7.2.2 Activities

The main features used to characterize activities are the temporal aspects, whether the system is used cooperatively or alone, how complex the interaction is, the displayed content itself, and if it's safety critical. The temporal aspects cover the frequency of usage, the time pressure and whether it's used continuously or interrupted frequently [37]. The objective of the application is to support the user in learning to avoid alcohol in a shopping context to reduce the risk of a relapse. The application is rather complex, as it employs two therapeutic methods namely AAT and CET. It consists of serial tasks but one main objective. Each time the user interacts with the entire application, the story will be played continuously through from start to end. In the beginning of a therapy, the application will be used on a daily basis and later on less frequently. The length of the game can vary depending on the speed of the user. The application is used individually and not cooperatively, but it could be interesting and motivating to see the scores or achievements of other users in the rehab center in a ranked table, which is not possible due to the General Data Protection Regulation. The displayed content is a supermarket environment consisting of various products, shelves, shopping carts etc. It is important to let users personalize the application with their favorite drink.

7.2.3 Contexts

According to Dey and Abowd [6], any information describing a situation is context. This includes the users of an application, where the interaction takes place and the application itself. For the purpose of a PACT analysis, this can be divided into different factors: the physical environment, the social context, the organizational context. In the beginning of a therapy, users are at a rehab facility, where they interact with the application in a room which is equipped with VR systems. Due to the cost factors, it is safe to assume that the physical space for each user within that room will be limited. Multiple users can interact with the application simultaneously, as long as each user has a separate computer and VR HMD. The group is most likely supervised by a therapist. After being discharged, the users can continue interacting with the system at home once the required technology is installed.

7.2.4 Technologies

Input and output possibilities, communication and content support vary in technologies, resulting in different user experiences. To create a user experience that is as immersive as possible, the application should be implemented as a VR application. The data input is through motion: The user moves in the virtual environment by actual movement and can interact with objects by moving arms and then pressing buttons on the controllers. A suitable hardware for instance is the HTC Vive. In this VR application 3D models are needed. These models should be as realistic as possible, so that the application feels realistic and triggers craving in the patients. Various different 3D Models are needed, ranging from bottles of alcohol to shelves. Additionally, a supermarket background noise can be helpful to strengthen the feeling of presence. A HMD with ear- or headphones is needed. The application can be started by patients themselves or their therapist.

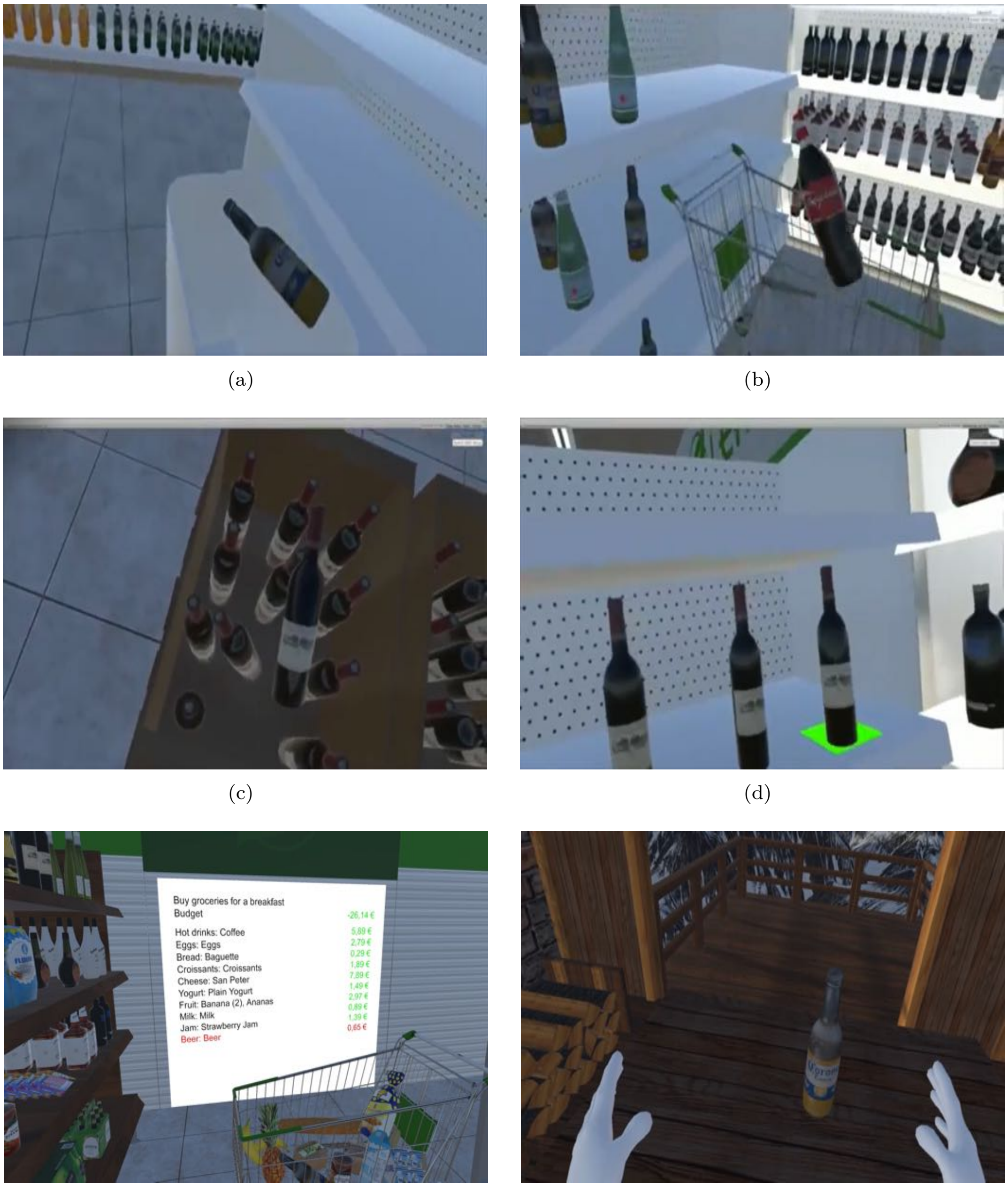
7.3 Design and Implementation

We designed and implemented three mini-games for CET and AAT as well as a non-gamified AAT application to be played in VR. All VEs were implemented using Unity3D game engine and were customized for HTC vive HMD and controllers. Furthermore, players did not have to move in VR as all games could be played while standing still at one place. The three mini-games were featured in a virtual supermarket with a wide range of grocery items, a typical supermarket background noise, and virtual customers walking around the supermarket. The overall game's goal was to find and buy the correct items of a shopping list. But before being able to buy any item, the player had to earn money by playing two mini-games of which one was inspired by AAT and the other by CET. All interactions with the virtual items in these VEs were accomplished using the HTC vive controllers (e.g., grabbing an object by pressing the trigger button).

The AAT game (see Figure 7.1 (a) and (b)) was played in front of an unsorted shelf in the alcohol section of the virtual supermarket. The shelf contains alcoholic and non-alcoholic beverages. The goal was to sort out the items by putting the non-alcoholic beverages into the shopping-cart and throwing the alcoholic ones into a trash bin. When an item was sorted correctly, the player earned 0.5€, the target container was colored in green, and a positive audio feedback is played when a non-alcoholic beverage was successfully thrown into it. Misplacement of an item (wrong container) resulted in losing 0.5€ and receiving negative auditory and visual (wrong container appears in red) feedback. Moreover, there was a time limit, so that the player was offered only a certain amount of time to play this game.

Next, the CET mini-game (see Figure 7.1(c) and (d)) started with the goal of exposing the players for a longer time with alcoholic beverages. To do so, they had to take one bottle at a time out of a bottle crate next to them and place it in a randomly assigned position in an empty shelf. Taking one bottle at the time implied that the player cannot interact with both hands at the same time. This was implemented to increase the focus onto each single bottle. Each placement resulted in receiving 0.5€. Since the bottles could not be placed anywhere else, losing money was not possible in this mini-game. Also similar to the AAT mini-game, the player was only granted a certain amount of time to play.

The shopping game (see Figure 7.1(e)) started by the player receiving a shopping-list whose items had to be purchased using the money which has been earned while playing the AAT and CET games. The player is supposed to buy only the items on this list while most of them do not refer to a specific product, but rather a group of items (e.g., fruit, hot drinks, bread, etc.) leaving some freedom of choice (e.g., type of bread). The items in the shelves



(a) (b) (c) (d) (e) (f)

Figure 7.1: Proposed gamified (a-e) and non-gamified (f) AUD therapy VR applications. The games (a-e) were realized in the context of a virtual supermarket and the non-gamified application in a mountain cabin. (a) AAT mini-game (G-AAT): sorting alcoholic beverages into a trash bin and (b) non-alcoholic beverages into a shopping-cart. (c) CET mini-game: taking alcoholic beverages from a bottle crate and (d) placing them on the marked positions. (e) Shopping mini-game: the items on a shopping list should be found and put into a shopping-cart. The names of the correct items were shown in green and incorrect ones (i.e., alcoholic beverages) in red on the shopping list. (d) Non-gamified AAT (N-AAT): alcoholic beverages were supposed to be pushed away and non-alcoholic beverages were to be pulled closer to the participant

or fridges can be grabbed and placed in a shopping-cart or taken out of it and back in their original place. If the player puts an item from the shopping-list into the shopping-cart, the shopping-cart was colored in green and a positive auditory feedback was played. If the item was not from the list, the shopping-cart was colored in red and a negative sound was played to give an immediate feedback on the player's action. The shopping-list was also updated accordingly to give the player an overview of the purchased items and their prices. If the player puts an item into the cart, which was not on the list, the item appears in red in the list. The purchase cannot be completed as long as incorrect items are in the cart. To make it more challenging, some alcoholic beverages were located at random positions in the shelves.

In addition, a non-gamified AAT VR application (see Figure 7.1(f)) was developed to be compared with the AAT mini-game. The interaction in this environment was designed to be similar to the AAT implementation of Wiers et al. for AUD therapy [394]. Since we could not find a comparable classical version for the other two mini-games (i.e., CET and shopping) we decided on a non-gamified implementation only for the AAT mini-game. This version was situated in a mountain cabin with a minimal interior design to keep the VE neutral. The bottles appeared at the center of a table. The controllers in this VE appeared as two virtual hands with white gloves. The instructions in this game were to push the alcoholic beverages away and pull the non-alcoholic beverages closer. Since the bottles were not implemented to be grabbed, the participant was only able to interact with the bottles by touching and then moving them either forwards or backwards. For each bottle, the participant had two seconds to react. After a short break of two seconds, a new bottle appeared at the center of the table.

7.4 Usability Study

In preparation for a long-term clinical study, we performed a pilot user study, which was approved by the local ethical committee of the Computer Science Department. The aim of this study was to evaluate the usability of three conditions: (i) Non-Gamified AAT (N-AAT), (ii) Gamified AAT (G-AAT), and the (iii) Whole Game (WG, containing all three mini-games: CET, AAT, and shopping). In addition, the comparison between N-AAT and G-AAT could help us understand the effects of the gamification on performance (i. e., the number of errors made while sorting the alcoholic and non-alcoholic beverages), motivation, and enjoyment of the training within each condition. Moreover, the comparison between G-AAT and WG could give us an insight into the experienced level of enjoyment by playing multiple mini-games which have a story behind them (i.e., earn money by sorting to buy your items on the shopping list). Thus, the following hypotheses were formulated:

- H_1: G-AAT is more motivating than N-AAT (due to the gamification element)
- H_2: G-AAT produces less errors than N-AAT in sorting alcoholic and non-alcoholic beverages
- H_3: WG is enjoyed more than G-AAT (due to change introduced by the collection of mini-games and the story behind it, i. e., earn money and do the shopping. This mission

is missing in G-AAT alone, as it is not clear for what purpose the money is collected.)

7.4.1 Participants

13 healthy participants (9 male and 4 female) between 22 and 35 years of age (avg. 25.76) took part in this study. Half of the participants did not need any visual corrections and of the other half, half used glasses and the other half contact lenses to correct their vision. Most participants have experienced VR before. 61.5% were students of the local Department of Computer Science, who received course credit for their participation. All participants signed a consent form prior to the study. They were also free to have breaks or quit the the study at any time.

7.4.2 Procedure

The study took place in a laboratory room of approx. 16m^2 with a dim light. The VR tracking space was approximately 3m × 3m. For rendering, system control, and logging an Intel computer running Windows 10 (graphics card: GeForce GTX 780 Ti; processor: Intel Core i7; RAM: 16GB) was used. In addition to the HTC Vive HMD, participants wore DT 770 Pro headphones for sound and noise-canceling. They also received an introduction to the whole study and instructions for each condition. The order of the conditions was randomized. Questionnaires were given before and after the whole study and after each condition. In addition to the questionnaire, during G-AAT and N-AAT, the data on the total number of sorted bottles and committed errors were logged. After a short training phase, the experiment was started.

7.4.3 Results

In order to examine our hypotheses, we used several questionnaires whose results will be presented in this section. Given that the Shapiro-Wilk test indicated non-normally distributed data, all differences were tested for significance using the non-parametric Wilcoxon Signed-Rank Test with an α level of .05. If the variations between pairs were insufficient ($n < 10$), the critical value for W was used instead of Z and p to evaluate the significance.

First, we used the System Usability Scale (SUS) [60], which is a standardized survey to evaluate the usability of a system. It consists of 10 Likert Scale items, where different statements can be answered from 1 (Strongly Agree) to 5 (Strongly Disagree). The SUS-score for each condition was calculated as: N-AAT=87.3, G-AAT=93.5, and WG=86.7. A SUS-score can rank from 0-100 and a score above 68 is considered to be above average, meaning that all conditions received above average usability scores [8]. A Wilcoxon Signed-Rank Test ($Z = -.41$, $p = .68$) showed no significant differences in the SUS-scores between N-AAT and G-AAT or between G-AAT and WG ($Z = -.15$, $p = .88$). Tests at the individual item level showed only for the first statement *I think I would like to use this system frequently* a significant preference of G-AAT over N-AAT ($W = 0.00$, critical value $= 3.00$).

We also used AttrakDiff [145] for evaluating usability as well as the user experience. The survey consists of three semantic differential questions each containing 9 or 10 word pairs. With the help of a seven point scale, the participant were asked to choose which word of the word pair was more appropriate to describe the system on four dimensions Pragmatic Quality (PQ), Hedonic Quality-Identity (HQ-I), Hedonic Quality-Stimulation (HQ-S), and Attractiveness. The word-pairs and the mean scores for the three conditions are displayed in Figure 7.2. The individual items reveal that the participants found G-AAT to be significantly more inventive ($W = -.45$, critical value = 9.00), more creative ($Z = -2.5$, $p = .01$), bolder ($W = 0.00$ critical value = 0.00), more innovative ($W = 0.00$, critical value = 2.00), more captivating ($W = 0.00$, critical value = 5.00), more challenging ($Z = -2.93$, $p = .003$) and more novel ($Z = -2.8$, $p = .005$) than N-AAT. Comparing G-AAT to WG showed that WG was perceived as significantly more human ($Z = -2.9$, $p = .003$) and more simple ($Z = -2.45$, $p = .01$) than G-AAT. And G-AAT was perceived as more innovative ($W = 0.00$, critical value = 8.00) than WG.

With the help of AttrakDiff, the PQ, HQ-I, HQ-S, and Attractiveness could be calculated for each condition (see Figure 7.3). PQ describes the usability of a system and how achievable its goals were. Comparing directly N-AAT and G-AAT showed that the participants found N-AAT significantly better at PQ ($Z = -2.43$, $p = .02$). HQ-I shows how much the users were able to identify with the product. The differences between N-AAT and G-AAT on this dimension were not significant ($Z = -.45$, $p = .65$). HQ-S reveals how much they were stimulated by the system and how much they felt the system could support the user in improving. On this dimension the direct comparison between N-AAT and G-AAT showed that the participants were significantly more stimulated by G-AAT than N-AAT ($Z = -3.18$, $p = .002$). Attractiveness describes the overall attractiveness of the system based on the perceived quality. For Attractiveness no significant difference was found when comparing N-AAT and G-AAT ($Z = -.45$, $p = .65$). The comparison of G-AAT and WG only showed a significant difference in PQ ($Z = -2.9$, $p = .004$).

Furthermore, the NASA Task Load Index (NASA-TLX) [141] was used in this study to evaluate the workload of each condition. This questionnaire is a multidimensional scale containing questions about different aspects. The participants can rate on a 100 point scale in steps of 5 the mental, physical and temporal demand and also their performance, effort and frustration. Since the targeted patient group may be limited in their cognitive or even physical abilities, it is important to evaluate whether or not the developed VR application were too demanding or difficult. The results can be seen in Figure 7.4. Here the average score and standard deviation is displayed for each item and experimental condition. When comparing G-AAT and N-AAT, multiple significant effects were found: G-AAT has a significantly higher mental demand ($Z = -2.53$, $p = .01$), a significantly higher physical demand ($Z = -3.06$, $p = .002$), a significantly higher temporal demand ($Z = -2.31$, $p = .02$) and needs significantly more effort to be put into to accomplish the desired level of performance ($Z = -3.06$, $p = .002$). The overall demand was significantly higher in G-AAT ($Z = -2.82$, $p = .005$). For performance ($Z = -.04$, $p = .97$) and frustration ($W = 15.00$, critical value = 3.00) no significant differences were found. Comparing G-AAT to WG showed that G-AAT was

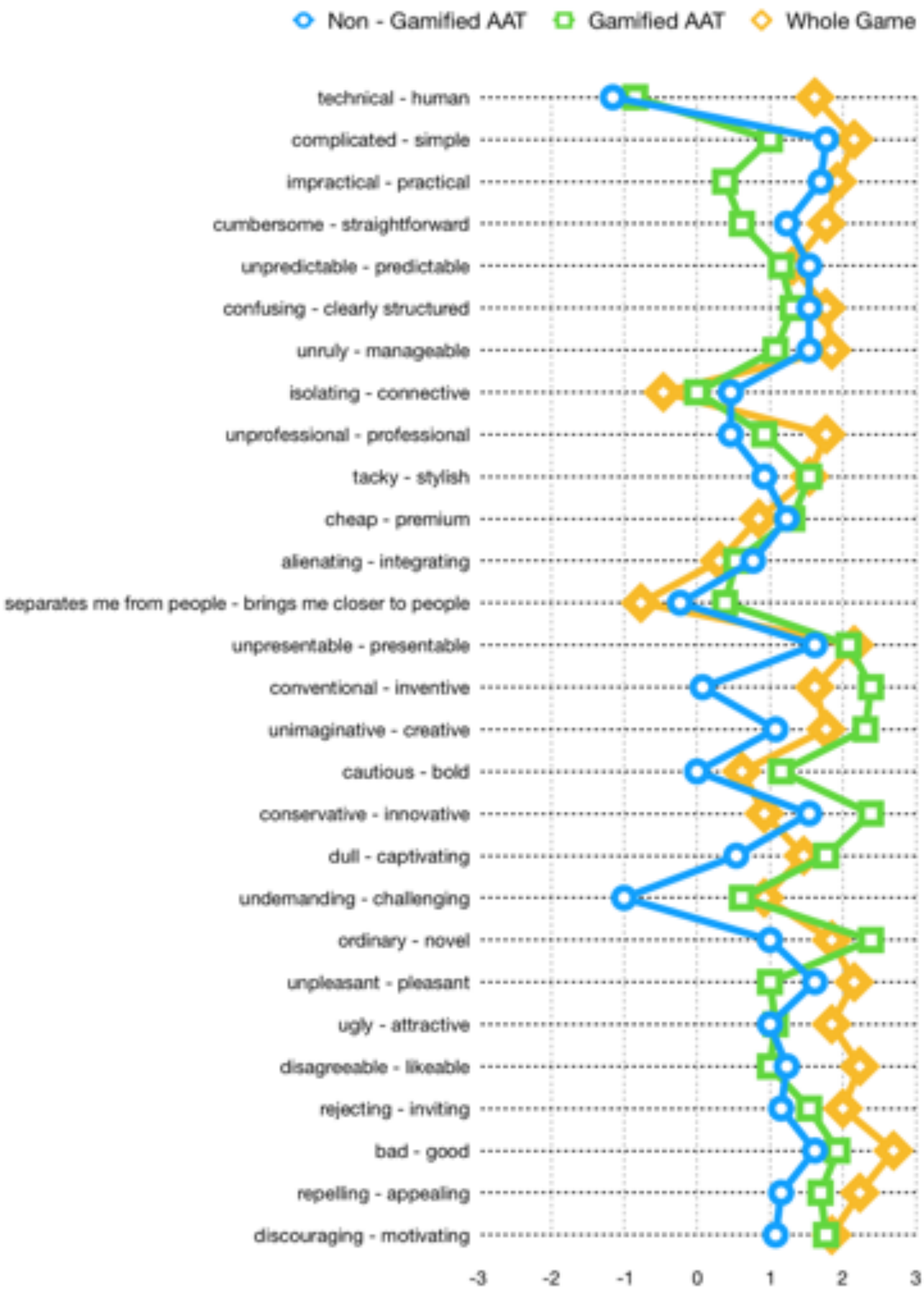


Figure 7.2: AttrakDiff word-pair items and their average scores per condition

significantly more mentally demanding ($W = 4.5$, critical value $= 9.00$). The other aspects did not show any significant difference.

Within the last subjective questionnaire, we asked the participants to rate the perceived

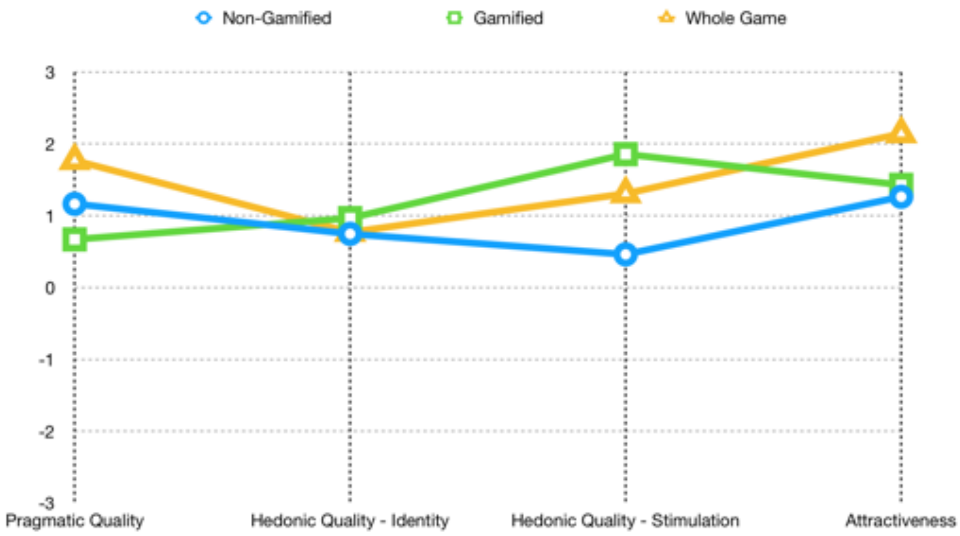


Figure 7.3: AttrakDiff four dimensions: PQ, HQ-I, HQ-S and Attractiveness for all three conditions. When comparing both AAT systems, N-AAT scored significantly higher at PQ and G-AAT at HQ-S.

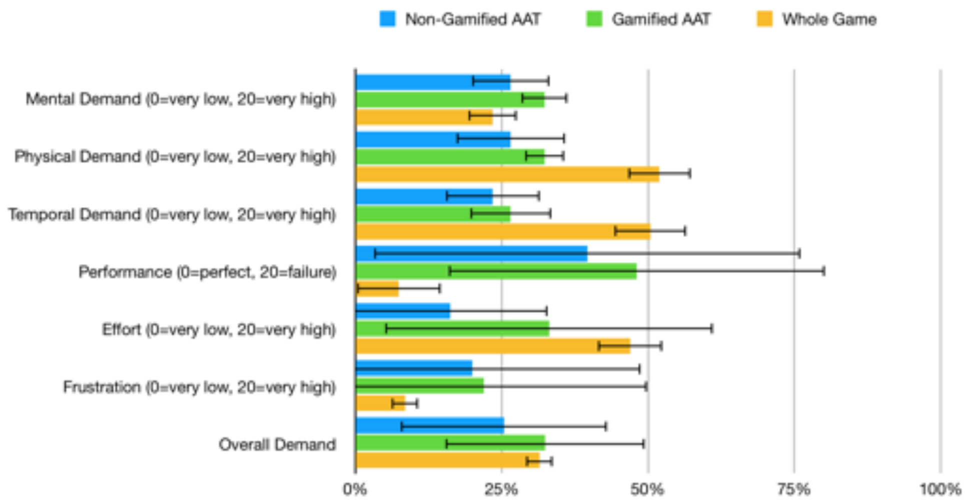


Figure 7.4: NASA-TLX average score and standard deviation per condition.

enjoyment, motivation and realism in a 7-point Likert scale; ranging from *Not at all* to *A lot*. The statements were *How much did you enjoy [the last condition]?*, *How motivated were you to perform well?* and *How realistic did you find the [the last condition]?*. As it can be seen in Figure 7.5, the participants significantly enjoyed the G-AAT more ($W = 0.00$, critical value = 5.00) and felt significantly more motivated to perform well ($W = 0.00$, critical value = 0.00) compared with the N-AAT. No significant effect was found in the enjoyment and motivation when comparing G-AAT to WG. The question about realism showed no significant effects. Thus, H_1 can and H_3 are not supported by these results.

Finally, the logged data revealed that the participants managed to push or pull approx. 44 bottles on average in the N-AAT (max: 55, min: 41, SD = 3.43) of which approx. 1.76

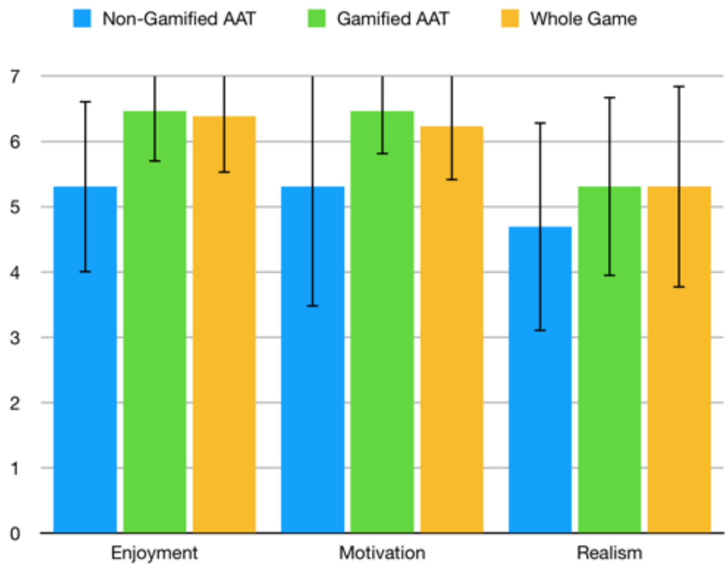


Figure 7.5: In this chart, the scores for enjoyment, motivation and realism are displayed. The results are based on a 7-point Likert scale.

were incorrectly pushed or pulled (max: 8, min: 0, SD = 2.09). Whereas in the G-AAT the participants achieved on average a complete count of 142 bottles (max: 220, min: 89, SD =39.6) with only 0.8 errors on average (max: 4, min: 0, SD = 1.2). The effect proved to be significant ($Z = -2.85, p = .004$). The participants also managed to significantly sort more bottles in G-AAT than in N-AAT ($Z = -3.18$, $p = .001$). This finding is in support of H_2.

7.4.4 Additional Comments

The participants enjoyed G-AAT and found it to be a straightforward and simple mini-game. Multiple participants emphasized how much they enjoyed throwing bottles. They also felt motivated by the time pressure. They appreciated that despite the time pressure, they were able to set the pace. Additionally, they suggested the following improvements: The height of the shelves was not optimal for tall participants, making them have to bend down rather low. Some participants would have found it more realistic if more virtual humans were shopping in the supermarket.

The general feedback about the N-AAT was less positive compared to G-AAT. It was perceived as more monotone and not mentally challenging enough. The participants wished for an acceleration, creating more time pressure (they had constantly two seconds to react per bottle). Some complained about the missing sound and auditory feedback.

Finally, the participants gave combined feedback on the WG. They generally enjoyed the money-aspect and that they were able to earn more money, if they played faster and that they had an influence on the spending: different items varied in costs. They enjoyed the freedom of choice and the progress they made. The different feedback types (the visual hint on the bottle, the summarized earned money, and the auditory feedback) accompanying each movement. The feedback was perceived as motivating especially in the G-AAT and CET. Of all the three mini-games, 7 participants liked the G-AAT the most and 4 participants the Shopping game. No participant favored the CET game. The participants suggested the following improvements: They wished for an option to quit the G-AAT and CET game if they thought they have earned enough money to complete the Shopping game. They found the placement of the alcoholic beverages in the Shopping game slightly irritating, because they were always placed between other items on the shelves.

7.5 Focus Groups with Patients and Clinical Experts

Members of different clinics attended the focus group meeting: The Universtiy Hospital Hamburg-Eppendorf, the Rehab Hospital Hansenbarg, and the Clinic Bremen-Ost. In the meeting, each participant tried the three VR mini-games. They generally liked the AAT and Shopping mini-games best. During the discussions for a potential clinical study design, a preference towards the AAT mini-game emerged, since it is highly comparable with the 2D implementation of the AAT and therefore the effect of VR and gamification can be validly analyzed. Prior to discussing further changes, it was decided that the Rehab Hospital will conduct the clinical study with their AUD patients. They would install four VR spaces, where a randomly selected group of patients will use the gamified VR AAT instead of the 2D AAT.

It was further discussed that even though the AAT game itself is fun and motivating, the gamification elements might distract from the alcohol. First, there should be more visual cues. These visual cues should be real and actual advertisements that the patients are confronted with in reality. It is sufficient if these cues were static - meaning that they would not need to be customizable to fit the patient's alcohol preference. There should be three different advertisements: one for beer which will appeal to the beer addicts, one for vodka which will appeal to the addicts of strong alcohol, and one for sparkling wine which will also appeal to the wine drinkers. These advertisements should always be clearly visible.

In addition to the visual cues, there should be auditory cues in the game. It is not unusual to hear the current offers of the supermarket through speakers. The auditory cues integrate well into the supermarket environment while (hopefully) also increasing the focus on alcohol. To make it less predictable and more realistic, the offers should not only be about alcohol, but also about other products. This way the patients do not learn that as soon as they hear a voice, there will be an alcoholic cue.

Furthermore, it was discussed that it is important to ask participants about their current craving. This should be done at the very beginning and end and three times in between. This way, the craving development can be understood better. The question should be displayed and answered in VR, so that the patient does not have to take the HMD off and can stay immersed and focused. Typically, the craving inquiry consists of a single visual analog scale with the question: *How strongly would you like to drink alcohol?* Patients can set their current craving using a slider, where at the one end the answer is *Not at all* and on the other *The most I've ever wanted*. This data needs to be logged and saved. Some minor changes were also requested. For instance, the trash bin should look more like a traditional trash bin and more negatively. In addition to the earnings, the number of completes and errors should also be displayed.

In a second step, the VR AAT was tried out by two AUD patients from the University Hospital Hamburg-Eppendorf, who agreed to participate in the focus group. Several therapists were also present to support the patients in case of need. Upon arrival, the patients were first introduced to the topic and the overall objective. They were instructed to immediately abort the experiment if they felt unwell either due to cyber sickness, craving, or any other reason. Then they tried the game for approximately 5 minutes and gave their feedback afterwards.

Initially, the first patient tried the AAT game with beer, which was not his favorite alcohol. This way, he could first try it out more safely. He found the game easy to use and was performing rather well. However, he did not quite feel as if he were in a supermarket. Even though the atmosphere was there when looking around, the interactable shelf was not realistic enough, because it is highly unlikely to have a shelf in a well-organized supermarket that has randomly positioned different kinds of beverages in one shelf. He felt that it might be also more realistic if the supermarket and alcohol section were a little more squished. While placing the alcohol bottles into the trash bin, he noticed it would be more realistic if the trash bin included metal lattice and if there was a rattling sound. When he tried the game with his favorite alcohol (Vodka), he immediately felt his craving for alcohol arousing.

The second patient was physically more fit and was directly playing the game with her favorite alcohol. Her favorite alcohol did not induce craving, but she mentioned that this would have been different in the beginning of her therapy. She really enjoyed the game and being able to throw the alcohol away. She emphasized that this has been a liberating feeling. She, too, found the game easy to use and thought that the supermarket was realistic, but far too clean. She used to buy her alcohol at discounters, where there was more dirt on the floor and different objects standing in the way. Also, she would have preferred cans over bottles, since she had mostly bought her favorite alcohol in cans. She did not notice right from the beginning that she had earned money and did not really pay attention to the time. It was discussed that it could have been more noticeable if the money feedback was not positioned on the bottle, but rather above the container (trash bin or shopping cart).

While meeting the patients, an additional option was requested: the therapist should be able to stop the application in a way that the patient is not alarmed. There should be a prompt in VR stating that the patient should stop playing and take off the HMD. An additional goal of the meeting with AUD patients was to find out whether a session of 10-15 minutes would be appropriate. However, due to the extremely high temperatures we decided to not physically overwhelm the patients and therefore let them play the game for only 5 minutes.

7.6 Discussions and Conclusion

In this chapter, We presented a PACT analysis for virtual AUD therapy and described the design, development, and evaluation of three mini-games (G-AAT, CET, shopping) and a non-gamified application (N-AAT) for AUD therapy in VR. They were designed based on classical therapy methods such as AAT and CET, which are associated with some limitations such as the lack of context and motivation for long-term training. That is the patients might find these therapy methods hard and tedious. A gamified VR can not only add the context and user preferences, but also motivation to continue to the intervention program. Furthermore, in preparation for a long-term clinical study, we conducted a user study with healthy participants to evaluate the usability and motivation for training. Before starting such a long-term study, which is associated with more challenges such as access to the AUD-patients and therapists for the supervision, it was necessary to check whether anything needed to be changed. Moreover, it is plausible that the AUD-patients might perceive the VR-settings differently, meaning that the collected data needs to be treated carefully and may not generalize to all populations.

Our results suggest that all conditions (N-AAT, G-AAT, WG) had a high usability score (SUS scores were above average). Moreover, in comparison with N-AAT, G-AAT was significantly more demanding (mentally, physically and temporally), but also more motivating. Furthermore, the participants made significantly less errors in the G-AAT than N-AAT. When directly comparing G-AAT and WG it is important to note that G-AAT was a part of WG. However more than half of our participants liked the G-AAT the best of all three mini-games presented within WG, while rather disliking the CET mini-game. Since many participants also enjoyed the Shopping mini-game, it could be assumed that the motivation and enjoyment would be higher when the final game used in a clinical context would only consist of the G-AAT and Shopping mini-game. This will also provide the game with a clear goal through the narrative (earning money to buy groceries).

In addition, this application was evaluated during two focus groups with clinical experts and AUD patients. As a results, the Rehab Clinic approved to conduct a clinical study with VR AAT and compare it to the 2D AAT (using a joystick and a computer screen). For this purpose, some changes such as including auditory and more visual cues as well as a craving prompt need to be further developed.

Finally, further improvements may be relevant for creating effective gamified AUD therapy in VR. Customization based on the patient needs can be named as an example. Each patient has a different context in which he/she has a high risk of relapse. This context could be the own living room or a certain situation with friends, where the patient might face group pressure. Including the context may increase the effectiveness of modern therapy methods. To create an even more immersive experience and to induce even more craving, olfactory could also be implemented and incorporated into the gamified application.

In short, gamified therapy in VR has the potential to revolutionize the recovery process of AUD patients by increasing their intrinsic motivation to abstain and allowing them to train at anytime and anywhere.

8

Chapter 8

Exploring Acceptability of Virtual Coaches for Home-based Balance Training in an Aging Population

Balance training has been shown to be effective in reducing risks of falling, which is a major concern for older adults. Usually, exercise programs are individually prescribed and monitored by physiotherapist or medical experts. Unfortunately, supervision and motivation of older adults during home-based exercises cannot be provided on a large scale, in particular, considering an ageing population. AR in combination with virtual coaches could provide a reasonable solution to this challenge.

We present a first investigation of the acceptance of an AR coaching system for balance training, which can be performed at home. In a human-centered design approach we developed several mock-ups and prototypes, and evaluated them with 76 older adults. The results suggest that older adults found the system encouraging and stimulating. The virtual coach was perceived as an alive, calm, intelligent, and friendly human. However, usability of the entire AR system showed a significant negative correlation with participants' age.

8.1 Introduction

Falls are a major concern among older adults as approximately one out of three people above 65 and one out of two older adults above 80 years old fall annually [286]. The consequent injuries due to falls can vary from a scratch to hip fractures and the hospital stays due to falls can be long and even last for the rest of the patient's life. As a matter of fact, more than 50% of all injury-related hospitalizations among older adults is due to falls [320]. In addition, post-fall syndrome can affect the quality of life and cause further restrictions such as fear of falling, immobilization, depression, and loss of autonomy [249]. The injuries caused by falls can also be fatal. They account, indeed, for the largest percentage of the deaths caused by unintentional injuries, which are the seventh leading cause of death among older adults [64].

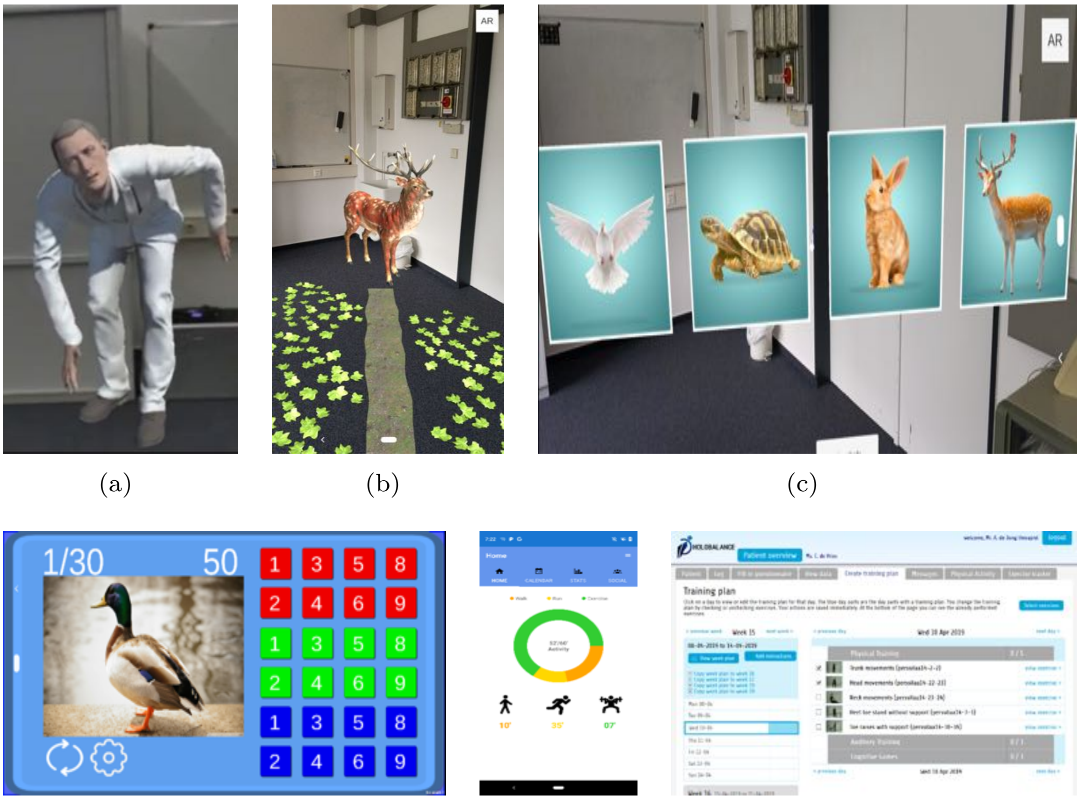
These damages can be reduced through practicing recommended preventive interventions. Several health organizations such as the joint American and British Geriatric Society (ABGS) [261] and the National Institute of Clinical Excellence (NICE) UK have reported a set of clinical guidelines for assessment and prevention of falls in older adults [110]. According to these guidelines, customized exercise programs for strength, balance, gait, and coordination training are effective in reducing falls [328]. They are recommended to be prescribed individually and monitored by a trained professional. In one study, for instance, home-based exercise programs were successful in reducing falls and fall injuries as much as 35-45% and equally effective for older women and men [294]. These programs included walking, muscle strengthening, and balance training, which were individually prescribed by physiotherapists (in the first two trials) and trained nurses (in the second two trials) during home visits. However, limited availability of the experts to supervise every single exercise session per individual introduces yet another restriction to such preventive interventions.

The system proposed in this chapter is designed to support clinical experts in prescribing and monitoring fall preventive exercises. It provides individualized exercise programs for home-based balance training based on the NICE guidelines. It covers sitting, standing, and walking exercises each in three progression levels [267]. For the older adults, the system consists of a holographic virtual coach that supervises the training at home.

A holographic representation of a virtual coach displayed via an AR HMD has several advantages over flat visualizations on monitors. The stereoscopic view of an AR display allows the users to quickly identify 3D poses in space, while also being able to look at the virtual coach from any angle. It is also possible to stand next to the virtual coach and see the instructions from the same perspective, reducing the cognitive load of mental rotations. In addition, the virtual coach and the user are able to walk towards each other, which enables the user to observe changes from a distance in 3D perspective, which is not possible in a flat screen. Furthermore, an optical see-through AR HMD is beneficial for our application scenario compared to fully-immersive VR HMDs.

This is due to the reason that AR HMDs allow the users to see the real environment with its all physical objects and obstacles. This property brings three main advantages into this context. First, it provides a psychological feeling of safety due to being familiar with the surrounding real environment (i. e. one's home). Second, it ensures a physical safety as in case the older user feels that a fall is about to happen (i. e. a near fall), nearby physical objects (e. g. a firm chair or wall) are in sight and can be grabbed to prevent the fall from happening. Third and with the same logic, the physical obstacles as well as any dynamic change in the real environment (e. g. a person crossing their exercise area) could be avoided during the training.

Furthermore, as we showed in Chapter 2, stereoscopic renderings delivered via an HMD elicits higher functional connectivity in the brain when compared to monoscopic renderings on projection screens or HMDs. Hence, an AR-based training session has enormous potential to improve brain connectivity on a functional as well as a structural level, which is particularly useful for treatment of balance disorders and their complex multisensory mechanisms.



(a) (b) (c)

(d) (e) (f)

Figure 8.1: Different modules of the proposed AR balance training system: (a) Balance Physiotherapy Hologram (BPH), (b-c) Cognitive Training and Exercise Games (CTEG), (d) Auditory Training Tool (ATT), (e) Activity Planning App (APA), and (f) Dashboard Module (DM).

Furthermore, the presence of the virtual coach in the physical surrounding may facilitate the performance of the motor exercises (also known as social facilitation effect [351]). Thus, the combination of stereoscopic AR and social facilitation effects opens up new vistas in the area of technology-based therapy and rehabilitation programs.

Moreover, studies have shown that falls are associated with lower cognitive performance [118]. Walking and talking, for instance, challenges the ability of some older adults in allocating attention to and maintaining performance of both tasks at the same time [326, 397]. Consequently, difficulty in performing walking and talking at once puts older adults at a higher risk of falling [381, 32]. However, simultaneous training of cognitive and physical abilities, in particular primary multi-modal exercise programs in combination with multi-sensory secondary tasks can improve cognitive and motor-cognitive dual task performance of older adults [118, 357]. Based on these evidences, the proposed system delivers, in addition, cognitive training and exercise games as well as an auditory training application for improving auditory attention and memory.

Further modules are motion capture and wearable sensors which enable monitoring of the correct performance of the exercises and an activity planning application for self-monitoring physical and cognitive activity performance and achieved goals. For the health care professionals, the system provides a dashboard which demonstrates the collected and analysed data from the older adults' homes and enables them to adjust the training program per individual.

In order to provide a system with optimal usability, the human-centered design approach was followed in this work. Through an iterative development circle, the requirements were analysed, different versions of the prototypes were implemented and evaluated by older adults. After each iteration, new insights were evolved which led to re-prototyping of the concepts.

In this work, we pursued the answers to the following research questions:

- **Q1**: How acceptable and usable is our AR-based balance and cognitive training solution for older adults?
- **Q2**: How is the virtual coach perceived?

Our evaluations show that the answer to Q1 depends on their age: the usability for 60-79 is OK to Good and for 80+ is the worst imaginable [26]. And for Q2, the virtual coach appears socially present in the physical room like an alive and intelligent human. Further contributions are (i) conceptual description of the AR-based balance training system and (ii) explanation of iterative implementations during the human-centered development process.

8.2 Related Work

Technology-based fall prevention interventions have been deployed in different contexts [137]. In the context of fall detection, for instance, Du et al. [94] proposed a robot control system for remote fall risk assessment in home environments. Silva et al. [329] presented a game to assess the quality of the older adults' locomotion using the data from accelerometers built in a smartphone. Another smartphone-based fall detection system was proposed by Abbate et al. [5]. Their system monitored the movements of older adults and in case of detection of a fall notified the caregivers. Computer vision techniques have also been used to automatically detect falls of older adults in their homes [411]. Furthermore, Ogonowski et al. [256] proposed a Kinect based interactive TV system for fall prediction and prevention. Their system made use of personalized physical exercise programs, gamification, and wearable sensors (senior mobility monitor device) and showed the feasibility of integrating such systems into the daily life of older adults.

There have been a few prior studies, which have explored if and how older adults accept such fall detection systems. For instance, Wu and Munteanu [405] involved older adults in the design and development of a fall risk assessment wearable device (in form of a belt). They also employed a field study to investigate the acceptance of the final device. Their findings suggested that the combination of contextual information for fall risk assessment and practical fall prevention instructions can improve the acceptance of such assistive technologies by older adults. Tyagi et al. [368] found out that the patient attributes can determine the adoption of technology-based interventions such as Tele-Rehabilitation (TR). In their observations, for instance, those patients who preferred TR were relatively younger than those who chose going to a day rehabilitation center. Thus, they recommended to include introductory videos in TR programs and provide technical support for older patients.

In addition, Harte et. al. [142] suggested a three-phase human-centered design methodology to enhance the usability and User Experience (UX) of health systems for older adults. The first phase of their methodology contains the construction of a context of use document to report use cases, mock-ups, and user feedback. The second phase suggests an expert usability inspection, and the third phase emphasizes regular UX testings to improve the final prototype. In the end, they reported a successful implementation of their methodology for the design and development of a system for fall detection for older adults.

In the context of decreasing the intrinsic risk factors of fall prevention intervention, prior studies have made use of different types of technology to retain balance and improve functional abilities of older adults [174, 127, 360, 104, 87, 92, 129, 139, 139, 103, 280, 331, 332, 341, 344, 359, 377]. For instance, Hardy et al. [139] created an exercise game to encourage the older adults to perform balance and gain training.

Games can as well mitigate cognitive impairments of older adults [183, 272, 315]. For instance, Schoene et al. showed that playing a step game at home for a period of eight weeks can improve physical and cognitive abilities of older adults. Besides, playing video games in the context of a programmed activity can be valuable for older adults in residential care. It promotes positive self-views by reintroducing challenge and fun in late life for independent older adults. However, these benefits could be inhibited by the degree of age-related changes and impairments for vulnerable older adults [128]. In addition, to enhance the mobility of older adults, Felberbaum et al. [101] identified three main categories of technological requirements. The first category is based on social aspects, which provides social companionship. *Co-walking with a virtual companion* is one of the recommendations in this category which was mentioned by the interviewees and was seen as a way to overcome the feeling of loneliness and making the physical activity more enjoyable for the older adults. The second category of requirements refers to the feedback given to the users to encourage mobility (e. g. setting goals and target destinations). The third is (remote) monitoring (e. g. automatic fall detection) to ensure a safe and at the same time independent mobility. We incorporated these three categories in our AR-based virtual coach.

Virtual coaches have been employed to assist older adults in their everyday life [366], for example, to commit to their medications [268] or to stay physically active [41, 9]. Albaina et. al. [9] proposed a virtual coach in form of an animated flower to motivate older adults to walk. The users were involved in the design and development process through two focus groups and one field study. Their results suggested that the elderly users enjoyed interacting with the flower virtual coach and liked to use it for a longer time. Although their virtual coach was not proven to be critical for motivating people, it was shown to improve acceptability of the system. A human-like virtual coach displayed on a tablet in combination with a pedometer could as well increase the amount of walking among older adults in a 2-months period [41].

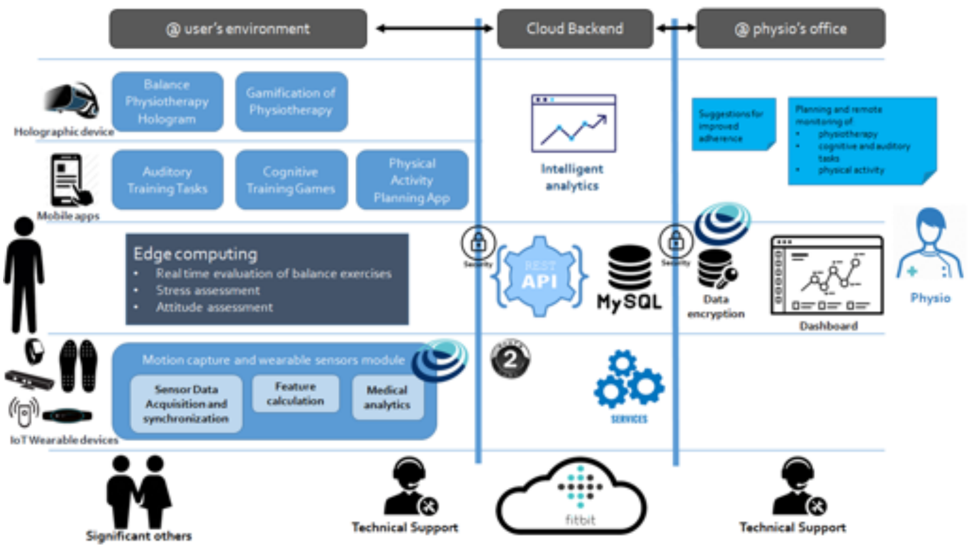


Figure 8.2: System Architecture.

The users also expressed a high level of satisfaction from the virtual coach who could in their opinion help them to walk more.

Moreover, AR technology has been used by older adults in previous studies [409, 311, 222]. For example, Bianco et al. [40] proposed a tablet-based AR application for fall preventive home modifications such as installing more handrails at desired locations at home. Also, training with a projection-based 3D AR game has been shown to be effective in improving mental rotation ability of older adults [212]. However, despite the potentials and advantages that AR can offer to older adults, it has not been used for balance training yet.

8.3 Concept

Based on interviews and focus groups with our cooperation partners from the medical domain in the Holobalance project [2], the following five technical modules, required to provide a home-based AR balance training system, were defined:

- **Balance Physiotherapy Hologram (BPH)**, which is an AR virtual coach that presents instruction, demonstrates correct performance of the balance exercises, and gives feedback to the users about their performance (see Figure 8.1(a)).

- **Cognitive Training and Exercise Games (CTEG)** provides holographic cognitive training games as well as gamified balance exercises (so-called *exergames*). Figure 8.1(b) shows the first progression of the walking exergame in which the user is asked to follow the path on the ground and walk towards the deer. Figure 8.1(c) shows one of the cognitive games. In this game, the user should remember the sequence of the cards and recall it after it has been shuffled.

- **Motion Capture and Wearable Sensors (MCWS)** to track the user's body movements, for instance, by depth cameras or pressure sensors. A heart rate sensor in addition measures the users' cardiac responses.

- **Auditory Training Tool (ATT)** aims at training speech in noise detection as well as auditory memory. Figure 8.1(d) shows the speech in noise detection task whose user interface shows animal pictures and colored numbers. Following an audio instruction (such as *"Show the duck where the blue five is."*), the user should click on the correct colored number (e. g. blue 5 button). The instructions contain different background noises (e. g. cafeteria type noise) to increase the difficulty of this task, which adapts on every question asked and according to the user's correct/correct answers.
- **Activity Planning App (APA)**, which allows the users to see their performance by means of charts, pictograms, and badges (e. g. when the goals are achieved) (see Figure 8.1(e)). Moreover, the app contains a calendar which shows the training program as well as a social networking feature to communicate with other users. Further information on APA can be found here [14].
- **Dashboard Module (DM)**, which is exclusively designed to be used by the clinical experts, which diagnose the older adults and prescribe personalized exercise programs (see Figure 8.1(f)).

Figure 8.2 depicts the architecture of the entire system. It is composed of the home-located devices (MCWS), cloud storage, and processing infrastructure. The data collected from the sensors are fused and analysed on the Edge Computer, which provides real-time feedback to the user articulated by the virtual coach. The results are then sent to the cloud to be monitored by the therapist. Based on the user's performance, the therapist can adapt the exercise program. Further information can be found in our paper dedicated to the architecture [365].

8.4 Method

In this work, we followed the human-centered design approach to understand the context of use, specify user requirements, produce mock-ups and prototypes, and evaluate them against the requirements through semi-structured focus groups and interviews. This process was iterated two times and across three countries. Table 8.1 summarizes the number of iterations and focus groups in different locations.

8.4.1 Iteration I

The initial requirements were defined through interviews and focus groups with our cooperation clinical partners. Based on these requirements the first versions of the prototypes and mock-ups were developed. These prototypes were then presented to the older adults in London to familiarize them with our concept and to collect their initial requirements. The BPH was, in this phase, displayed via a Meta 2 HMD and the motions of the users were tracked by

Table 8.1: An overview of the semi-structured focus groups performed in the UK, Germany, and Greece

Iteration	Focus Group	Location	Participants
I	I	London (UK)	N = 5
	II	London (UK)	N = 47
II	III	Freiburg (Germany)	N = 8
	IV	Athens (Greece)	N = 16

(a) (b) (c) (d) (e)

Figure 8.3: Profiles of the virtual coach (a) Cartoony Female (b) Cartoony Male (c) Realistic Female (d) Realistic male (e) Realistic polygonal 3D reconstruction

a Kinect 2 sensor. The CTEG, ATT, and APA were ready to present to the users just before the second focus group. By then, the CTEG and ATT were functional prototypes, but APA was still a paper mock-up.

Focus Group with Older Adults I

In our first study to collect feedback of the older adults, five seniors (3 female) with an average age of 74.8 years old (SD = 6.14, min=69, max=84, Mdn=72) took part in a focus group in London. Three participants (60%) did not have a balance disorder and 80% (4 persons) did not have prior experience with HMDs. All five participants used the then available system, which consisted of a Meta 2 HMD. While using the system, the older adults saw five profiles of the virtual coach (see 8.3).

The different visual alternatives were identified together with the clinical experts of our project to explore commonly used representations in commercial exergames (male and female cartoon avatars, Figure 8.3(a) and Figure 8.3(b)), in AAA games (realistic male and female avatars, Figure 8.3(c) and Figure 8.3(d)) or in high-end telepresence applications (polygonal 3D reconstruction of the experimenter created using RGBD cameras of a Kinect 2,

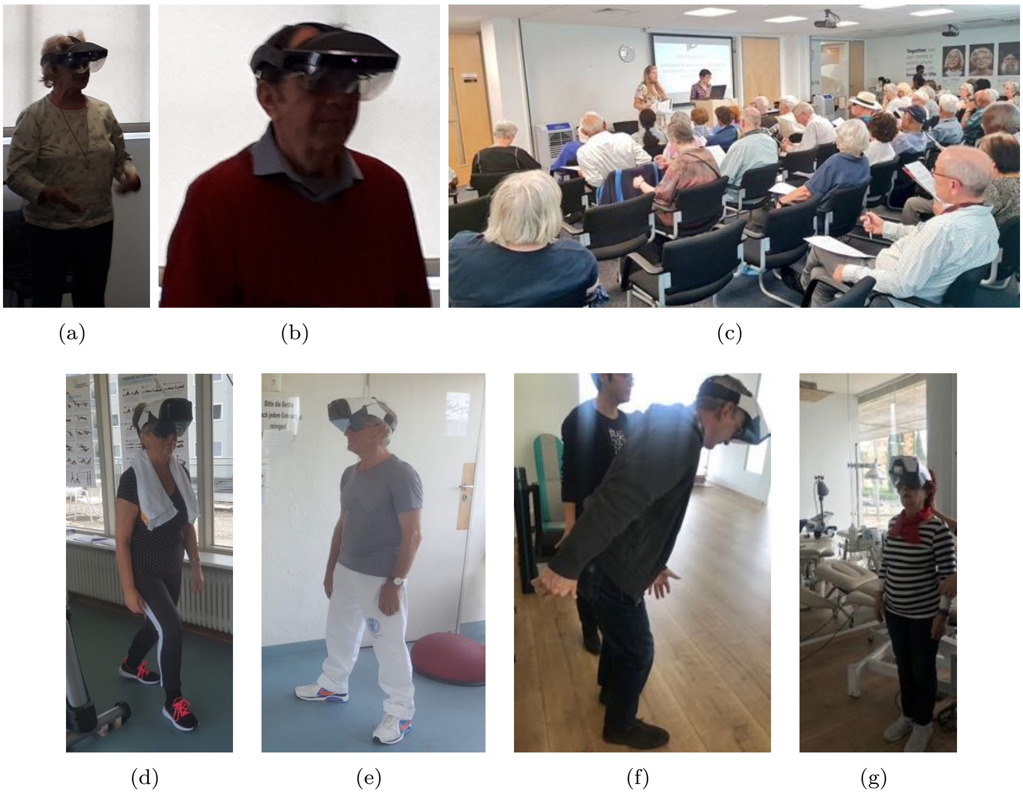


Figure 8.4: Images from the focus groups: (a and b) two older adults wearing the Meta 2 HMD during the focus group I in London (England), (c) focus group II with 47 older adults in London (England), (d and e) older adults wearing the Docooler AR HMD holding a Pixel 3 XL phone during the focus group III in Freiburg (Germany) and (f and h) during focus group IV in Athens (Greece).

Figure 8.3(e)). Each virtual coach demonstrated the correct performance of an exercise (in sitting, standing, and walking postures) to the participants. The participants then performed one standing exercise (i. e., standing and bending over as if to pick up an object from the floor) and received feedback from the virtual coach about their performance. Following that, they fill-out questionnaires and answered our questions in a semi-structure interview.

The first questionnaire was SUS [60] questionnaire. The average SUS score was calculated as 72.5, which is considered to be above average.

The second questionnaire was a modified version of the Co-Presence and Social Presence in Virtual Environments questionnaire [273], which has 15 statements to measure four sub-dimensions of social presence using five-point Likert scales (1=Strongly Disagree - 5=Strongly Agree). This questionnaire was originally developed to measure the impression of the virtual audience during a public speaking task in VR. We modified the questionnaire to evaluate the impression of the virtual coach in our scenario. All four factors of this questionnaires were rated above medium (Figure 8.5): 1) Participant's Reaction to the virtual coach (Mdn=4, M=3.55, SD=1.57), 2) Perceived virtual coach's Reaction (Mdn=4, M=3.6, SD=1.57), 3) Impression of Interaction Possibilities (Mdn=3, M=2.8, SD=1.51), and 4) (Co-)Presence of the virtual coach (Mdn=3, M=3, SD=1.77).

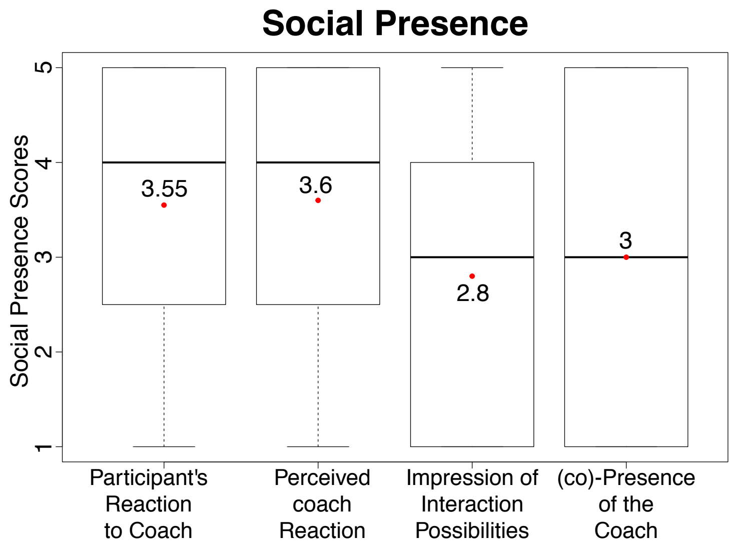


Figure 8.5: Co-Presence and Social Presence Scores

Moreover, the participants were asked about their virtual coach preference. 60% of the participants (two females and one male) preferred the realistic male virtual coach. The mean score for the naturalness of the virtual coach's movement was 3.4 (on a scale of 1 (Very Bad) to 5 (Very Good)) and the mean score for the naturalness of the virtual coach speech was 3.8 (on a scale of 1 to 5 ranging from very bad to very good).

We also received several comments resulting in the following user preferences: Addition of volume control for the virtual coach voice, a lighter and smaller HMD, "make the virtual coach more responsive", provide more feedback and more interaction, "less instruction from the coach with more dialogue and conversation", "UK English accent would be preferred", "should not be a generic experience, should be more adaptive on a person basis", "virtual coach mood should look more optimistic and happy", a progress feedback on a daily basis, and detection of safety by the system.

Focus Group with Older Adults II

Our second focus group in London (see Figure 8.4(c)) was intended to verify if the previous findings of the first focus group were representative. Here, the participants did not use the system and saw only a live demonstration as well as pictures and videos of the system (i. e. all main modules: BPH, MCWS, CTEG, ATT, and APA). 47 older adults (29 females) aged between 60 to 84 years (M=73.61, SD=6.12, Mdn=72.5) took part in this focus group. The participants could ask questions, discuss the system with the technical and clinical experts, which were present during the focus group, and finally express their opinions through questionnaires as well as direct conversations with technical, clinical and usability experts.

Before presenting any module, participants answered a demographic questionnaire. The results of this questionnaire show that 13 individuals (28%) had balance disorder, 14 (30%) persons were not sure and answered the question with *maybe*, and the rest with *no* (42%, 20 people). After that, participants were asked about the frequency of experiencing a fall (min=0, max=14, M=1.07, SD=2.27) or near fall (min=0, max=24, M=3.26, SD=4.83) during the past 6 months, given the definition of each type of falls to them: a fall is defined as an event that results in a person coming to rest inadvertently on the ground or floor or other lower level; whereas a near fall is an event in which a person feels a fall is imminent, but avoids it by compensatory action, such as grabbing a nearby object or controlling the fall. In addition, they stated that they are physically active on a daily (81%) or weekly (15%) basis and only 4% indicated that they never do physical exercises. Besides, half of the participants had $\geq 3m^2$, one third $2m^2$, and the rest (17%) had $\leq 1m^2$ free space at home for performing exercises.

The number of participants, which use computer systems (including smart phones, tablet PCs, etc.) were quite high: 94% daily, 4% weekly, and 2% never use any type of computers. Besides, on a scale of 1 (very bad) to 5 (very good) they rated their confidence in using technologies on average 3.32 (SD = .91, Mdn=3). Around 94% of the participants claimed that they do have broadband internet at home. We solicited information about their means of communication with their medical doctors. While telephone was still the most frequent tool for communication (89%), newer technologies are used as well such as Email (38%), mobile phone (37%), SMS/text message (11%), instant messengers (e. g., WhatsApp) (8%), and social media (6%). Furthermore, we received comments such as “through online appointment system”, “doctor does not have mobile phone (I do have an iPhone)”, “almost impossible to contact”, and “increasingly difficult to make an appointment”. Hence, we can assume that the social media panel of the APA could be of use for the older adults as our participants were social media users (e. g. 64% use Facebook, 36% Twitter, and 21% WhatsApp).

Moreover, 62% of the participants stated that they did not have any previous experience with 3D holograms. Two persons also commented that “[No but I] know what it is though”. Also, 63% had no previous experience with HMDs. They gave us two comments as “[No but I have] read about them for films” and “yes and no (I am aware of them)”. Thus, 37-38% of the participants did have prior experience with 3D holograms and HMDs.

After that we asked whether they have ever used a fitness wrist band (e. g. Fitbit). 64% of our participants answered this question with No. Two people added their comments as “[No] but would [like] a free trial” and “[Yes] my friends have them”.

Next, they were asked about their experience in using exergames (they were provided with an explanation about this term). 85% had no previous experience with such technologies. Three people gave us their comments as “[Yes] but [I] got bit confused using it (pre use: no instruction seen) I guess if I had it explained I would have been [doing] as I thought it was very versatile and easier to body”, “[No] but I have used DVD for exercises”, and “[No just] briefly with grandchildren”. Their previous experience using video games for cognitive training were also limited as 76% of them had no previous experience with such games.

In order to estimate the usability aspects of each module, a modified version of the SUS

Table 8.2: Means (standard deviations), medians, and percentage of the responses to a modified version of the SUS questionnaire during the Focus Group II. Its items are: *Frequency of use:* "I would like to use this system frequently if it meant that it would reduce my risk of falling", *Ease of use:* "I think this system would be easy to use", *Learn to use:* "I would imagine that most people would learn to use this system very quickly", *Confidence to use:* "I would feel very confident using this system", *Need for Assistance:* "I think that I would need assistance to be able to use this system. If yes, what type?". All items (with an exception of the *Need for Assistance*) were answered on a scale of 1 (strongly disagree) to 5 (strongly agree). The *Need for Assistance* item was a multiple choice question.

	BPH		MCWS		CTEG		ATT		APA	
Item	*M (SD)*	*Mdn*	*M (SD)*	*Mdn*	*M (SD)*	*Mdn*	*M (SD)*	*Mdn*	*M (SD)*	*Mdn*
Frequency of use	4.15 (1.02)	4	3.81 (1.31)	4	3.57 (1.33)	4	3.74 (1.15)	4	3.75 (1.28)	4
Ease of use	3.69 (.99)	4	3.6 (1.07)	4	3.64 (1.19)	4	3.64 (1.16)	4	3.46 (1.32)	3
Learn to use	3.04 (1.03)	3	3.05 (1.11)	3	3.32 (1.22)	3	3.38 (1.08)	3	3.53 (.98)	3
Confidence to use	3.67 (.87)	4	3.57 (1.21)	4	3.77 (1.17)	4	3.6 (1.17)	4	3.73 (1.12)	4
Need for Assistance? What type?		%		%		%		%		%
No		41.3		34.9		48.8		43.9		30.6
Physical		10.9		9.3		14		9.8		5.6
Technical		39.1		44.2		25.6		34.1		50
Educational		17.4		18.6		11.6		7.3		16.7
Medical		6.7		7		4.7		7.3		2.8

questionnaire with five statements was repeated after each module's demonstration. Since participants had to fill out 99 questions in 31 pages we reduced the SUS questionnaire to the five most important items of the SUS for our case and reported their mean score. The Assistance question consisted of multiple choice questions. Thus, for each part of the system we could calculate how many percent of the participants needed what type of assistance. As it can be seen in Table 8.2, our participants were positive about using all modules frequently if it would reduce their risk of falling. All modules were also perceived as easy to use, with an exception of APA, which was rated towards a neutral opinion. Indeed, all modules were more advanced in the development than the APA, which existed at that time only as a paper mock-up. The responses to the next statement: "I would imagine that most people would learn to use this system very quickly" were also rather neutral for all modules. On the other hand the participants stated that they would feel confident using any of the modules presented to them. An assistance to use the system was not seen as requirement, unless it is a technical assistance, and as some comments clarify that such technical assistance would be beneficial in particular in the beginning: "Technical assistance to start with" and "to confirm I am using it correctly".

Furthermore, the participants gave feedback and comments about the virtual coach. The most frequent comment was about the computer generated voice: e. g. "speech needs to be quite clear syllables, well-articulated, as older people 'decode' fast speech rather badly quite often". As a result, 61% of the participants wished for a change in the voice of the virtual coach towards a more "natural voice" or a "computerised speech, but it must follow natural English rhythm". Some other characteristics of the virtual coach that participants would like to personalize were gender (48%), age (41%, e. g. "age 35+, looking 'mature' / knowledgeable"), cloths (28%), and size (26%, e. g. "not stick thin"). Some participants in addition asked for a change of the ethnicity of the virtual coach. The repetition of the instruction and the exercises was as well the most desired interaction function.

In order to receive older adults' preferences in respect to the CTEG prototype, a self-made questionnaire was developed. Since this prototype featured realistic 3D models of different types of animals to be augmented in the real environment, i. e., older adults' homes, questions about animals were the main theme of this questionnaire. Based on its results, 68% of the participants liked animals in general and 60% of them have had at least once a pet in their life. Household as well as forest animals were their most favourite types. Mouses/rats, snakes, and desert animals were on the other hand the least favourite ones.

The ATT prototype consisted of two tasks: 1) *speech in noise detection task*, which contained pictures of animals on its user interface and 2) *auditory memory task*, which played a short audio story to be listened carefully (to be able to answer the questions about the story afterwards). Again, to better understand the users' preferences a self-made questionnaire was employed. The pictures of the first task's interface were preferred to be means of transportation (55%), animals (36%), musical instruments (38%), fruits, and "sport, anti-capitalist or anti-racist leader/heroes". The genre of the audio stories to be played in the second task was chosen to be comedy (60%), suspenseful (43%), crime, and "factual information". The last two question of the ATT questionnaire focused on the length of the training session. Based on the responses, most of the participants stated that a 10-15 minutes auditory training per session for 3-4 days a week could be easily followed by them on a regular basis. For instance, a person commented "daily duties vary, most people fit it to their daily routines". They (58%) disliked longer sessions on fewer days.

In the end, we received quite a number of general comments and feedback summarized in the following paragraphs:

- "From what is shown very early days. I think the idea is good and could reach some people. The motivation issue is probably the most difficult thing to solve. This could be very interesting."

- "I would like this to be extended to a group use eventually. I think this is a great preventative idea."

- "I would be keen to have headset + sensors as compact + neat as possible, to take up least space. I would like the system to be easy to clean + maintain ([pressure sensing] socks do not appeal unless washable)."

- "This should not be a substitute for face-to-face contact and group advises, it might help those people who are isolated, but we are social beings and this should be encouraged."

- "Inspirational, brilliant particularly as falls clinics are rare in our regions. You may fall, but not be referred even if you experience multiple falls."

- "Interesting, early stages of develop, but suggest testing out in reality in each stage to ensure connectivity from one stage to the next."

8.4.2 Iteration II

In the second iteration, the requirements and user preferences gathered from the first two focus groups were used to further iterate the prototypes. The most emphasized requirement was a lighter, more affordable, and cableless AR-based HMD. Thus, in the second iteration, the BPH and CTEG were optimized for Pixel 3 XL phone, which can be integrated in the lightweight Docooler AR Headset.

Moreover, in order to address the accessibility of our concept for the older adults in different locations, the third focus group was held in Freiburg (Germany) (see Figure 3 (d and e)) and the fourth in Athens (Greece) (see Figure 3 (f and h)). For the purpose of these focus groups, all prototypes and questionnaires were localized (i. e. prepared in both German and Greek languages). Since the participants of the focus groups in London, wished for a natural human voice in the prototypes, in the second iteration, the voices of a native German and a native Greek speaker were recorded and integrated in the prototypes. All participants used the system for about 15 min supervised by at least a clinical and a technical expert before the interviews started.

Focus Group with Older Adults III & IV

The participants of both focus groups wore Docooler AR HMD and used BPH and CTEG prototypes, which were running on the Pixel 3 XL smartphone. ATT and APA do not require the headset, and therefore were displayed on the smartphone alone. The participants evaluated these prototypes using several standard questionnaires and a few open-ended questions in the end. We also collected their feedback during the focus group.

We interviewed in total 24 older adults (21 female) between 60 and 84 years old (M=71.42, Mdn=72, SD=7.99). Eight of them were interviewed in Freiburg and the rest in Athens. Half of the participants had balance disorder and reported an average of .56 (SD=.82) falls and .85 (SD=1.32) near falls during the past 6 months. Seven responses to the near fall question were not specific, and therefore were not included in the analysis. These responses were: "Daily" (n=2), "2-3 per week"(n=1), "several" (n=1), "only dizziness"(n=1), "a few" (n=1), and 1 empty response.

Our participants were also physically active since more than 70% of them reported daily (37.5%) or weekly (33.33%) performance of exercises. The rest (25%) stated that they never do exercises. To perform exercises, about 70% of the participants indicated that they have $3m^2$ or more free space at home. 13.04% of them had $2m^2$ and the rest (17.39%) had $1m^2$ available space at home. Moreover, about 2/3 of these seniors had broadband internet at home and used computer (including smart phones, tablet PCs, etc.) on a daily (58.33%) or weekly (4.16%) basis. On a scale of 1 (very bad) to 5 (very good) they rated their confidence in using technology on average 2.33 (SD=1.34, Mdn=2.5). Furthermore, the majority of the participants (91.66%) did not have prior experience with AR or VR.

In order to calculate the SUS score we had to remove four responses, which were not fully available. The mean SUS score for the remaining 20 responses was calculated as 57 (min=10,

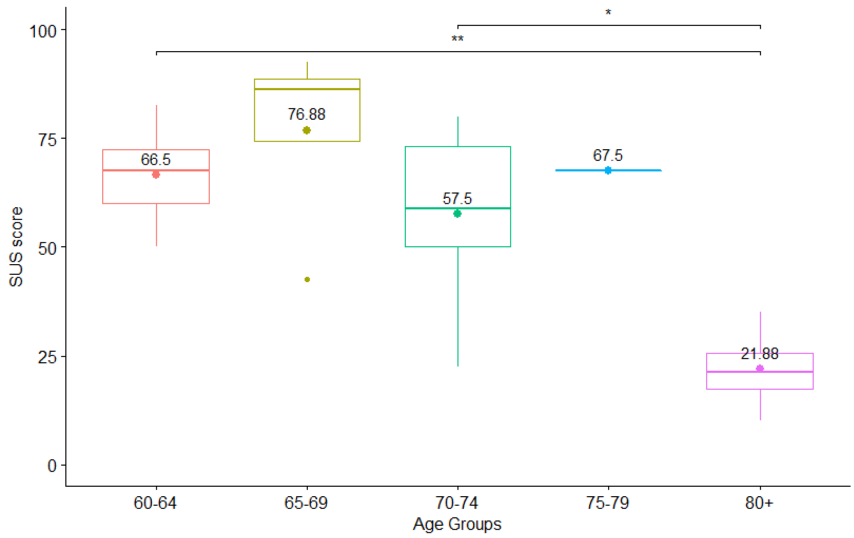


Figure 8.6: System Usability Scale.

max=92, Mdn=60, SD=24.97). Based on the Shapiro-Wilk test, the SUS scores were normally distributed. Therefore, we ran Pearson correlation, which suggested a significant negative correlation between the subject's age and their SUS score (cor=-.628, df = 18, p<.01). In other words, as the age of the participants increased, their evaluation of the usability of the system decreased. Figure 8.6 depicts the SUS scores per age groups. The results of the ANOVA suggested a significant difference between the age groups (F(4,15)=5.57, p<.01, $\eta^2_G = .6$). The pairwise comparisons with Bonferroni corrections showed that the SUS score of the participants, which were over 80 years old, was significantly lower than the participants aged 60-64 (p<.01) and 70-74 (p=.05) years.

As it can be seen in Figure 8.7, the results of the NASA-TLX indicates that despite wearing an HMD, the system was not rated as physically demanding. The effort and frustration levels were also rated low. Although the mental and temporal demand was rated higher than other factors, the values are still around the medium score. Therefore, the overall workload during the testing can be considered as low to medium.

The results of the User Experience Questionnaire (UEQ) are shown in Figure 8.9. All scales have a mean value above .8, which corresponds to a positive evaluation. In particular, the system was perceived as novel, stimulating, and attractive.

The Godspeed questionnaire [28] was employed to evaluate the representation of the virtual coach in the BPH module. As it can be seen in Figure 8.8, different aspects of the virtual coach are rated quite high. Therefore, the participants have perceived the virtual coach like a human (Anthropomorphism), who is alive (Animacy) and calm (Perceived Safety), has intelligence (Perceived Intelligence), and has left a positive impression on them (Likeability).

In addition, the mean Social Presence [23] score (on a scale of -3 to +3) was .26 (SD=1.19). The positive value suggests that the participants perceived the virtual coach as conscious and aware.

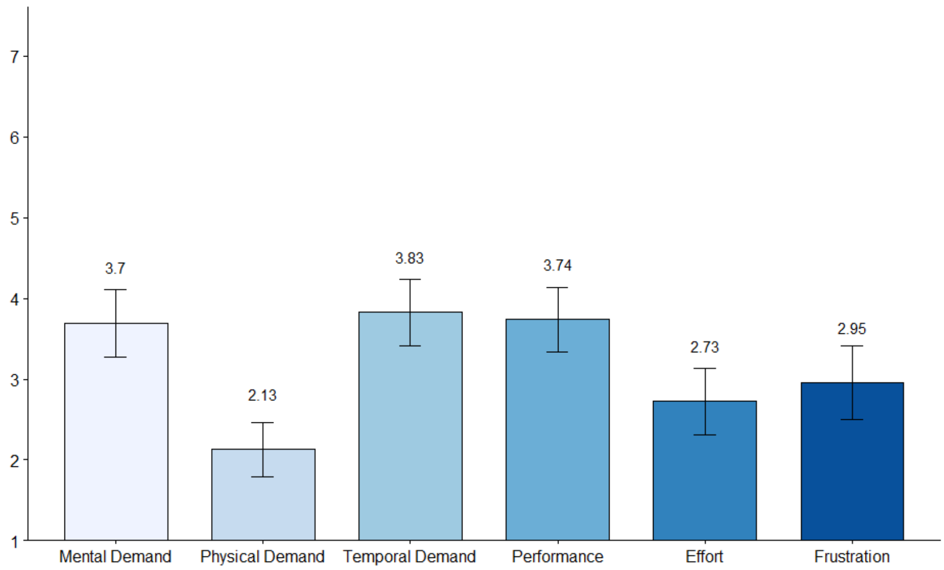


Figure 8.7: NASA Task Load Index.

8.5 Discussion

In order to have a successful technology-based intervention for fall prevention, it is essential to understand the older adults' view on such programs. Through several focus groups across three countries, a large number of requirements were defined, which helped us to design and develop working prototypes for multiple technical modules of our concept.

First, it is essential to be aware of the individual differences between the older adults. The 20 years difference between the age of 60 and 80 makes a significant effect on older adults' use of technology. Our results confirms that the use of our system for participants in their 60s was significantly easier than those over 80 years of age. Indeed, we observed that some 80+ participants did neither have internet at home, nor mobile phones or computers, and as a result had difficulty interacting via the smartphone's touch screen.

Next, one needs to acknowledge the older adults' interest in technology and its application in their everyday life. For instance, the older adults living in London we had the chance to introduce our concept to, were not completely unfamiliar with 3D holographic displays, AR/VR HMDs, or social media. And as a result they were very positive and supportive about the project.

Our findings from the focus groups in Freiburg and Athens suggest that the overall impression of the product was positively evaluated by the older adults. They also showed that the given tasks of the prototypes could be solved without unnecessary effort and they felt in control of the interaction. The system was as well very exciting, creative, and motivating for them.

Furthermore, the results of the social presence and Godspeed questionnaire showed that the older adults perceived the virtual coach not as a computer-generated image, but rather

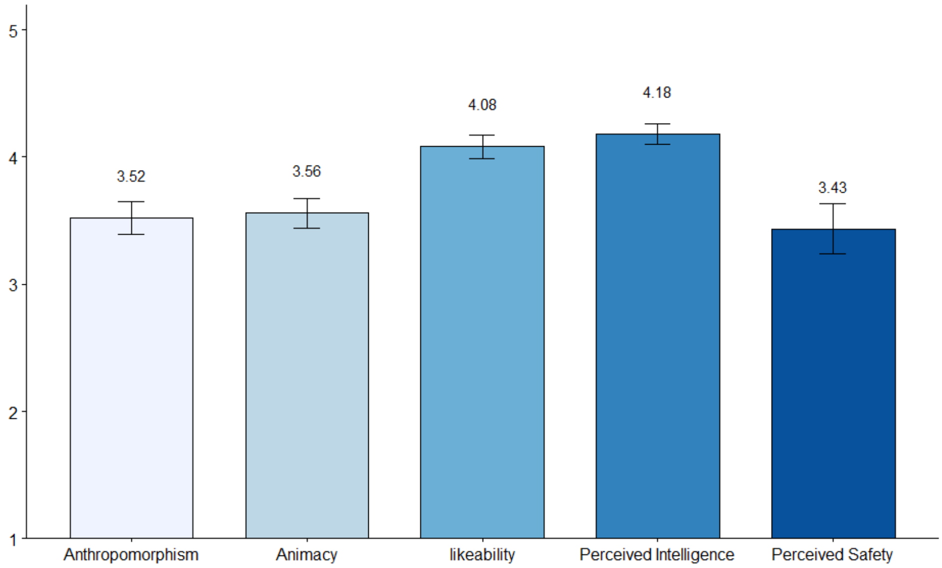


Figure 8.8: Godspeed Questionnaire.

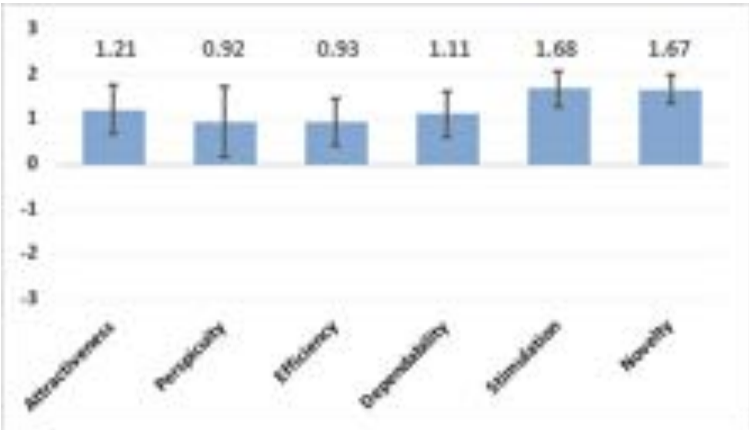


Figure 8.9: User Experience Questionnaire.

a conscious and alive person, who is present in the room and observes them. Moreover, they found the virtual coach intelligent, friendly, and safe to interact with.

8.6 Conclusion

In this chapter, we presented a first investigation of the acceptance of an individualized virtual coaching system based on AR for balance training.

The proposed system is intended to support balance training at home without the requirement of the physical presence of human physiotherapeutic or medical experts. In a human-centered design approach we developed and evaluated several mock-ups and prototypes with more than 70 older adults across three countries.

In general, the results suggest that the participants found the system encouraging and stimulating. Furthermore, the virtual coach was perceived like an alive, calm, intelligent, and friendly human and has therefore enormous potential to enhance social facilitation, which is important for motivation. However, the usability of the AR system showed a significant negative correlation with the participants' age. To summarize, the results give important implications for the use of AR in the overall aging population.

The overall positive feedback of the participants could be caused by a novelty effect, which may have led to the tendency for performance to initially improve when new technology is instituted. Hence, feedback of the participants should be collected during long-term usage of our system. In this context, we plan to validate the effectiveness of this novel rehabilitation and training concept through a randomized control trial on each three locations, which hosted the above described focus groups (i. e. London, Freiburg, and Athens).

Chapter 9

9 Conclusion

In this thesis, we performed a number of experiments to examine the psycho- and physiological effects of exposure to and interaction with MR therapeutic applications.

We started with FloVR (Chapter 2 and Chapter 3), which is a game designed to improve mood and reduce deppressivity. First, we conducted an fMRI study to investigate differences in functional brain connectivity between objectively different levels of immersion, that the gaming display offers. In particular, we considered an MRI compatible HMD, which has the ability to display 3D content, objectively more immersive than a back-projection of a screen via a mirror system [338]. Therefore, we optimized our FloVR to be played inside MRI scanner. The game was displayed using (i) mirror projection and (ii) an MRI compatible HMD with 2D and (iii) 3D content.

Our motivating assumption was that the display that elicits higher brain connectivity should be most suited for long-term brain training interventions, in this case, for therapeutic purposes. The reason is that extended training with the help of such a display could permanently improve brain connectivity on a functional as well as on a structural level.

Our results revealed higher brain connectivity between brain regions in the HMD condition compared to the screen and, in particular, in the stereoscopic condition compared to the monoscopic condition. We interpret these results as a hint that training in MR environments, in contrast to environments displayed on a screen, may be superior in eliciting and therewith facilitating brain connectivity in intervention studies.

In order to determine the psychological effects of FloVR, we conducted another study, which was presented in Chapter 3. Here, we examined the effects of playing FloVR on depressivity, anxiety, affect, flow, and sense of presence, when it was played using a VR HMD or a tablet device. The results suggested that in comparison to a control and the tablet condition, playing FloVR using an HMD could significantly reduce depressivity and negative affect. Moreover, VR could increase the positive affect, yet not significantly different from the tablet and control conditions. The sense of flow and presence were, additionally, higher in the VR condition.

We conclude that playing FloVR in VR brings major objective and subjective advantages over playing it on non-immersive displays. We showed that it can, not only improve mood and reduce depressivity, but also elicit higher functional brain connectivity. Meaning that long term interventions of FloVR in VR can potentially offer an alternative solution for treating mood and depressive disorders.

In the psychological study of FloVR, we used a commercial HMD as an immersive and a tablet device as a non-immersive display. Such a comparison was relevant for the collaborating clinic from a pragmatic point of view, but it meant that the game controls had to be programmed differently. That is, participants in the VR condition navigated in the VE using their head movement and in the Tablet condition by rotating the tablet. The tablet device was also used in the control condition for reading a text. Thus, the level of immersion of the device was not different from the Tablet condition, although the immersion level of the task was different (gaming vs. reading). Finally, the VE of FloVR was a virtual nature, yet not as realistic as an immersive video of a real nature.

In order to control for both the type of display and the realism of the stimuli, we conducted another study (Chapter 4), in which the content of the conditions were different in the level of immersion but were all displayed on an HMD. Thereby, we examined the effects of exposure to immersive videos or photo slideshow of a forest and an urban environment in terms of their respective impact on cognition, physiological reactions, mood, and stress. The results showed that the forest environment had a positive effect on cognition as well as mood and the immersive videos induced a higher sense of presence.

Based on the results of both experiments (Chapter 3 and Chapter 4), we could show the positive effects of exposure to and interaction with immersive natural environments on mood and cognition. The cognitive test, which was used in the latter study, was taken from the TSST [189], which was originally designed to induce moderate psycho-social stress in the laboratory. In our study, we used its mental arithmetic task and showed that a six-min exposure to a virtual natural environment can reduce the induced stress from the previous cognitive task and increase cognitive performance in the follow up test.

The question that arose next was whether a complete TSST in VR has the same psycho-social effects as in real life. The TSST consists of giving a speech and performing mental arithmetic calculations each for five minutes in front of three persons. Therefore, we developed an adaptation of TSST in VR and compared it with in vivo TSST (Chapter 5). Meanwhile, we assumed that the larger the size of virtual audience, the stronger the effects of the test. However, to our surprise, the question of whether the size of virtual audience plays a role in induced stress using virtual TSST had not been answered before. Therefore, we aimed at investigating the effects of different size of the virtual audience (i.e., three, six, and fifteen) on perceived social anxiety as well.

The analysis of different physiological and subjective measures revealed different aspects of the exposure to this psycho-socially stressful situation. Physiological arousal could be observed with VR inducing anxiety, yet less than in vivo TSST. Furthermore, some of the subjective measures showed a state of anxiety experienced during the experiment. An effect of the virtual audience size could be observed only in the heart rate measures as a virtual audience size of three induced the highest values, which was significantly different from an audience of size six and fifteen.

The results imply that MR can be used for psycho-therapeutic interventions to both decrease and increase anxiety depending on the disorder and needs of the therapy receivers. For instance, for treating the social anxiety disorder, one can increase social anxiety in training situations, whereas for treating depressive and mood disorders, one can elevate mood and decrease anxiety during the training sessions. For this reason, MR is an optimal tool to simulate real life situations, such as confrontation to stressful situations or visit to natural environments, and make use of them for therapeutic interventions.

However, one might need to get away from the ordinary, everyday situations and gain an extra-ordinary experience to have a bigger picture of the psychological distresses. Similar to an experience described by some astronauts, who had the privilege to travel far away from the earth and gaze at our planet from the outer space. The experience has been termed the *Overview Effect* [390] and has been reported to have positive, life-lasting effects on space travelers.

In this context, VR is an ideal tool for making this positive experience accessible to more individuals than sole space travelers. Therefore, we designed and developed a VR application, in which participants could experience earth gazing from the surface of the moon in VR. In addition, we combined this experience with Compassion-Focused Therapy and created our Space-Compassion VR application (Chapter 6).

We compared the effects of our application with a controlled VR condition, in a study with individuals with occasional paranoid thoughts. The results showed that our intervention could reduce state paranoid thoughts as well as worry and increase self-compassion feelings. Although, these results show positive effects of our intervention and have practical implications, in future, we will disentangle the Compassion-Focused Therapy from earth gazing, in order to examine the pure experience of the overview effect in VR.

Back to the earth, we learned through our collaborating clinical experts that compulsive consumption of alcohol is a major health problem and there is need for innovative therapeutic interventions, which are more attractive and effective than the current ones. For this purpose, we designed and developed a VR application for AUD therapy, which was evaluated by clinical experts and AUD patients (Chapter 7). The usability of the game was tested as well, which showed an excellent usability for all parts of the application.

Based on the feedback of the participants, we iterated the development of our VR AUD therapeutic application and prepared a psycho-physiological experiment. Thereby, we aimed at comparing our implementation with its traditional equivalent therapy method at the rehabilitation clinics, namely the AAT with joystick and on a computer screen. In addition, it is known that AUD patients have favorite types of alcoholic beverages, for which they feel craving. Therefore, we set out to investigate whether such craving can be induced in VR for virtual representations of alcoholic beverages. For this reason, we customized our experiment for each individual to play with their favorite type of alcoholic beverage with different frequency levels. Unfortunately, due to the COVID-19 pandemic, experiments with AUD patients had be ceased. However, we hope that when the experiment can be resumed again, its results will reveal new insights into the personalization and practical use of VR for AUD therapy.

Personalization of the program is also a key for successful physical therapy, in particular, to prevent falls [328]. On the other hand, the prevalence of individuals seeking preventive or managing individualized programs overwhelms the availability of exerts to provide this treatment. To address this challenge, we designed an AR coaching system to support clinical experts in prescribing and monitoring individualized fall preventive exercises (Chapter 8). In addition, we investigated its acceptance through an evaluation with 76 older adults in three countries across Europe. The results suggested that older adults find the system encouraging and stimulating. The virtual coach was perceived alive, calm, intelligent, and friendly. However, usability of the entire AR system showed a significant negative correlation with participants' age.

It is worth noting that depending on the context and the target group, the perception of virtual humans, as part of the MR intervention, may vary in different MR applications. We argue this as one particular model of virtual human was used in both TSST study and the AR coaching system. Therefore, while it facilitated evoking social anxiety by playing the role of the main jury in all variations of our virtual TSSTs, it was perceived friendly and calm while playing the role of a virtual coach in our AR coaching system.

Considering all reported experiments so far, one can argue that the type of virtual stimuli in MR was limited to only visual and auditory, although further types of stimuli could enrich the illusion and result in more immersive experiences. For instance, a visit to a virtual forest could be more realistic for the participants when, in addition to the sight and sound of the forest, they could smell the wood (e.g., using an olfactory system), feel the wind breeze on their skin (e.g., using a gentle fan), receive proper haptic feedback when they touch the trees (e.g., with haptic gloves) or walk on the leaves (e.g., using haptic shoes), and feel the warming effect of the sunlight (e.g., using a thermal display). Thereby, the effects of exposure to such an immersive virtual nature could be stronger.

Furthermore, our measures in all our experiments were mostly cognitive tests, clinical and HCI-related questionnaires as well as heart rate and skin conductance recordings. In one of our experiments, salivary cortisol hormone level was measured as well. However, understanding the mechanisms of change in MR-based interventions require brain imaging to acquire data resulting from behavior as well as brain processes. For this reason, in Chapter 2, we used the fMRI technique to reveal the effects of VR-based gaming on brain connectivity.

fMRI is one of the most commonly used brain imaging techniques [88], but the use of this technique with MR paradigms is associated with some limiting factors [321], such as electromagnetic interferences with other instruments, disturbing noise, huge machinery dimensions, high costs as well as the horizontal and unnatural position of the participants during scans' acquisition which restricts their head movement. Thereby, the studies that report using VR inside MRI scanner, including ours, use either a conventional screen back-projection via a mirror or compatible goggles with limited resolution and FOV [392]. Thus, participants do not benefit from a full immersion into the VE.

To overcome this limitation, a comparable brain imaging technique to fMRI, namely functional Near-infrared Spectroscopy (fNIRS), may be employed and combined with a common immersive VR system. This combination could deliver a more immersive VR experience and consequently a more ecologically valid experiment. The reason is that both fNIRS and fMRI rely on the BOLD responses and thus are very comparable. Yet, fNIRS has some advantages compared to fMRI, especially in respect to simultaneous use of immersive VR. For instance, it is easily portable and enables participants to not only move their head freely, but also use their hands, and even walk within the restricted physical environments, that grant a higher immersion into the VE. Thus, the combination of immersive VR and fNIRS to study the neuro-physiological effects of VR-based therapeutic interventions remains a topic for future work. On the other hand, fNRIS is also associated with some limitations, e.g., only cortical activity can be measured and the anatomical information cannot be directly inferred. To address these limitations, a replication of the experiments using fMRI should performed as well.

Finally, a limitation of most of our experiments is their relatively small sample size. Also, the sample was not always diagnosed with the targeted disorder, for which we designed and developed our MR interventions. The reason is the research tradition of exploring clinical phenomena on a sub-threshold level, which investigates symptoms on a dimensional level. This tradition has been recently promoted by the Research Domain Criteria (RDoC) framework led by the National Institute of Mental Health (NIMH) [168]. Thus, from the perspective of clinical research, the scale of the research presented in this thesis falls in the category of preliminary to pilot research, whose outcomes form the foundations for planning future full-size randomized clinical trials.

All in all, ***MR Pharmacy*** aimed at providing a collection of MR-based therapeutic applications. Such applications may become available as ***MR Medication (MRM)s*** in the future. Thus, similar to a real pharmacy, an MR pharmacy should ideally offer a variety of MRMs as alternatives to pharmacological drugs. Likewise, MRMs should be clinically approved, whose side-effects, recommended dosage, etc. are known to the consumer. This way, similar to taking, for example, a painkiller for a headache, one could receive prescribed MRMs for their needs. Equally compared with availability of the standardized pharmacological medications all around the world, MRMs shall become standard and available all around the world as well. Ideally, the process of prescribing and monitoring such medications may also become digital. For this purpose, ***MR Therapists*** may assist the health system by taking the role of real therapists, when it is needed. I believe that both therapists and therapy receivers will benefit from the enormous potentials of innovative forms of therapy and rehabilitation, in particular, MR interventions. My confidence is based on our numerous interviews with health professionals as well as patients in different age groups and with different disorders together with the results of the psycho-physiological experiments as they were presented in this thesis.

Bibliography

[1] Group ica toolbox. `http://icatb.sourceforge.net/groupica.htm`. (Accessed on 02/23/2021).

[2] Holobalance – holograms for balance rehabilitation. `https://holobalance.eu/`. (Accessed on 03/26/2021).

[3] Indoor air quality — european environment agency. `https://www.eea.europa.eu/signals/signals-2013/articles/indoor-air-quality`. (Accessed on 04/14/2020).

[4] Mathworks – entwickler von matlab und simulink - matlab & simulink. `https://de.mathworks.com/`. (Accessed on 02/23/2021).

[5] Stefano Abbate, Marco Avvenuti, Francesco Bonatesta, Guglielmo Cola, Paolo Corsini, and Alessio Vecchio. A smartphone-based fall detection system. *Pervasive and Mobile Computing*, 8(6):883–899, 2012.

[6] Gregory D Abowd, Anind K Dey, Peter J Brown, Nigel Davies, Mark Smith, and Pete Steggles. Towards a better understanding of context and context-awareness. In *International Symposium on Handheld and Ubiquitous Computing*, pages 304–307. Springer, 1999.

[7] Substance Abuse, Mental Health Services Administration, et al. 2015 national survey on drug use and health. 2016.

[8] Assistant Secretary for Public Affairs. System usability scale (sus). =https://www.usability.gov/how-to-and-tools/methods/system-usability-scale.html, 2013. Last visited 2.8.2018.

[9] Inaki Merino Albaina, Thomas Visser, Charles APG Van Der Mast, and Martijn H Vastenburg. Flowie: A persuasive virtual coach to motivate elderly individuals to walk. In *2009 3rd International Conference on Pervasive Computing Technologies for Healthcare*, pages 1–7. IEEE, 2009.

[10] Mihyang An, Stephen M Colarelli, Kimberly O'Brien, and Melanie E Boyajian. Why we need more nature at work: Effects of natural elements and sunlight on employee mental health and work attitudes. *PloS One*, 11(5):e0155614, 2016.

[11] Allison P Anderson, Michael D Mayer, Abigail M Fellows, Devin R Cowan, Mark T Hegel, and Jay C Buckey. Relaxation with immersive natural scenes presented using virtual reality. *Aerospace Medicine and Human Performance*, 88(6):520–526, 2017.

[12] Page L Anderson, Matthew Price, Shannan M Edwards, Mayowa A Obasaju, Stefan K Schmertz, Elana Zimand, and Martha R Calamaras. Virtual reality exposure therapy for social anxiety disorder: A randomized controlled trial. *Journal of Consulting and Clinical Psychology*, 81(5):751, 2013.

[13] G Andrews, A Basu, P Cuijpers, MG Craske, Peter McEvoy, CL English, and JM Newby. Computer therapy for the anxiety and depression disorders is effective, acceptable and practical health care: an updated meta-analysis. *Journal of Anxiety Disorders*, 55:70–78, 2018.

[14] Thelma Androutsou, Ioannis Kouris, Athanasios Anastasiou, Sotiris Pavlopoulos, Fariba Mostajeran, Doris-Eva Bamiou, Greg Genna, Sergi G Costafreda, and Dimitrios Koutsouris. A smartphone application designed to engage the elderly in home-based rehabilitation. *Frontiers in Digital Health*, 2:15, 2020.

[15] Matilda Annerstedt, Peter Jönsson, Mattias Wallergård, Gerd Johansson, Björn Karlson, Patrik Grahn, Åse Marie Hansen, and Peter Währborg. Inducing physiological stress recovery with sounds of nature in a virtual reality forest—results from a pilot study. *Physiology & Behavior*, 118:240–250, 2013.

[16] Matilda Annerstedt, Johan Norman, Mattias Boman, Leif Mattsson, Patrik Grahn, and Peter Währborg. Finding stress relief in a forest. *Ecological Bulletins*, pages 33–42, 2010.

[17] Jay Appleton. Prospects and refuges re-visited. *Landscape Journal*, 3(2):91–103, 1984.

[18] Solomon E Asch. Effects of group pressure upon the modification and distortion of judgments. *Organizational Influence Processes*, pages 295–303, 1951.

[19] Solomon E Asch. Opinions and social pressure. *Scientific American*, 193(5):31–35, 1955.

[20] Leonie Ascone, Johanna Sundag, Björn Schlier, and Tania M Lincoln. Feasibility and effects of a brief compassion-focused imagery intervention in psychotic patients with paranoid ideation: A randomized experimental pilot study. *Clinical Psychology & Psychotherapy*, 24(2):348–358, 2017.

[21] Entertainment Software Association. Essential facts about the computer and video game industry. `https://www.theesa.com/wp-content/uploads/2019/05/ESA_Essential_facts_2019_final.pdf`, 2019. (Accessed on 09/07/2020).

[22] Ronald Azuma, Yohan Baillot, Reinhold Behringer, Steven Feiner, Simon Julier, and Blair MacIntyre. Recent advances in augmented reality. *IEEE Computer Graphics and Applications*, 21(6):34–47, 2001.

[23] Jeremy N Bailenson, Jim Blascovich, Andrew C Beall, and Jack M Loomis. Interpersonal distance in immersive virtual environments. *Personality and Social Psychology Bulletin*, 29(7):819–833, 2003.

[24] Jeremy N Bailenson, Nick Yee, Dan Merget, and Ralph Schroeder. The effect of behavioral realism and form realism of real-time avatar faces on verbal disclosure, nonverbal disclosure, emotion recognition, and copresence in dyadic interaction. *Presence: Teleoperators and Virtual Environments*, 15(4):359–372, 2006.

[25] Yoshihiro Banchi, Keisuke Yoshikawa, and Takashi Kawai. Evaluating user experience of 180 and 360 degree images. *Electronic Imaging*, 2020(2):244–1, 2020.

[26] Aaron Bangor, Philip Kortum, and James Miller. Determining what individual sus scores mean: Adding an adjective rating scale. *Journal of Usability Studies*, 4(3):114–123, 2009.

[27] Rosa María Baños, Veronica Guillen, Soledad Quero, Azucena Garcia-Palacios, M Alcaniz, and Cristina Botella. A virtual reality system for the treatment of stress-related disorders: A preliminary analysis of efficacy compared to a standard cognitive behavioral program. *International Journal of Human-Computer Studies*, 69(9):602–613, 2011.

[28] Christoph Bartneck, Dana Kulić, Elizabeth Croft, and Susana Zoghbi. Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots. *International Journal of Social Robotics*, 1(1):71–81, 2009.

[29] Jo Barton and Jules Pretty. What is the best dose of nature and green exercise for improving mental health? a multi-study analysis. *Environmental Science & Technology*, 44(10):3947–3955, 2010.

[30] Thomas Baumgartner, Dominique Speck, Denise Wettstein, Ornella Masnari, Gian Beeli, and Lutz Jäncke. Feeling present in arousing virtual reality worlds: prefrontal brain regions differentially orchestrate presence experience in adults and children. *Frontiers in Human Neuroscience*, 2:8, 2008.

[31] Thomas Baumgartner, Lilian Valko, Michaela Esslen, and Lutz Jäncke. Neural correlate of spatial presence in an arousing and noninteractive virtual reality: an eeg and psychophysiology study. *CyberPsychology & Behavior*, 9(1):30–45, 2006.

[32] Olivier Beauchet, Véronique Dubost, François Herrmann, Muriel Rabilloud, Régis Gonthier, and Reto W Kressig. Relationship between dual-task related gait changes and intrinsic risk factors for falls among transitional frail older adults. *Aging Clinical and Experimental Research*, 17(4):270–275, 2005.

[33] Antoine Bechara. Decision making, impulse control and loss of willpower to resist drugs: a neurocognitive perspective. *Nature Neuroscience*, 8(11):1458, 2005.

[34] Aaron T Beck. *Cognitive therapy of depression.* Guilford press, 1979.

[35] Gian Beeli, Gianclaudio Casutt, Thomas Baumgartner, and Lutz Jäncke. Modulating presence and impulsiveness by external stimulation of the brain. *Behavioral and Brain Functions*, 4(1):1–7, 2008.

[36] David M Benedek, Matthew J Friedman, Douglas Zatzick, and Robert J Ursano. Guideline watch (march 2009): Practice guideline for the treatment of patients with acute stress disorder and posttraumatic stress disorder. *Focus*, 7(2):204–213, 2009.

[37] David Benyon. *Designing interactive systems: A comprehensive guide to HCI, UX and interaction design*. Pearson Edinburgh, 2014.

[38] Ralf Bergner-Köther. *Zur Differenzierung von Angst und Depression: ein Beitrag zur Konstruktvalidierung des State-Trait-Angst-Depressions-Inventars*, volume 18. University of Bamberg Press, 2014.

[39] Marc G Berman, John Jonides, and Stephen Kaplan. The cognitive benefits of interacting with nature. *Psychological Science*, 19(12):1207–1212, 2008.

[40] Michael Lo Bianco, Sonja Pedell, and Gianni Renda. Augmented reality and home modifications: a tool to empower older adults in fall prevention. In *Proceedings of the 28th Australian Conference on Computer-Human Interaction*, pages 499–507. ACM, 2016.

[41] Timothy W Bickmore, Rebecca A Silliman, Kerrie Nelson, Debbie M Cheng, Michael Winter, Lori Henault, and Michael K Paasche-Orlow. A randomized controlled trial of an automated exercise coach for older adults. *Journal of the American Geriatrics Society*, 61(10):1676–1683, 2013.

[42] Frank Biocca. The cyborg's dilemma: Progressive embodiment in virtual environments. *Journal of Computer-Mediated Communication*, 3(2):JCMC324, 1997.

[43] Frank Biocca, Chad Harms, and Jenn Gregg. The networked minds measure of social presence: Pilot test of the factor structure and concurrent validity. In *4th Annual International Workshop on Presence, Philadelphia, PA*, pages 1–9, 2001.

[44] Josiane Bissonnette, Francis Dubé, Martin D Provencher, and MT Moreno Sala. Virtual reality exposure training for musicians: Its effect on performance anxiety and quality. *Medical Problems of Performing Artists*, 30(3):169–177, 2015.

[45] Lukas Bittner, Fariba Mostajeran, Frank Steinicke, Jürgen Gallinat, and Simone Kühn. Evaluation of flowvr: a virtual reality game for improvement of depressive mood. *Biorxiv*, page 451245, 2018.

[46] Véronique D Bohbot, Milagros S Copara, Jean Gotman, and Arne D Ekstrom. Low-frequency theta oscillations in the human hippocampus during real-world and virtual navigation. *Nature communications*, 8(1):1–7, 2017.

[47] Rod Bond. Group size and conformity. *Group processes & intergroup relations*, 8(4):331–354, 2005.

[48] Patrick S Bordnick, Amy Traylor, Hilary L Copp, Ken M Graap, Brian Carter, Mirtha Ferrer, and Alicia P Walton. Assessing reactivity to virtual reality alcohol based cues. *Addictive Behaviors*, 33(6):743–756, 2008.

[49] Cristina M Botella, M Carmen Juan, Rosa María Baños, M Alcañiz, Veronica Guillén, and Beatriz Rey. Mixing realities? an application of augmented reality for the treatment of cockroach phobia. *CyberPsychology & Behavior*, 8(2):162–171, 2005.

[50] Stephane Bouchard, Stephanie Dumoulin, Genevieve Robillard, Tanya Guitard, Evelyne Klinger, Helene Forget, Claudie Loranger, and Francois Xavier Roucaut. Virtual reality compared with in vivo exposure in the treatment of social anxiety disorder: a three-arm randomised controlled trial. *The British Journal of Psychiatry*, 210(4):276–283, 2017.

[51] Stéphane Bouchard and AA Rizzo. *Virtual reality for psychological and neurocognitive interventions*. Springer, 2019.

[52] Wolfram Boucsein. *Electrodermal activity*. Springer Science & Business Media, 2012.

[53] Diana E Bowler, Lisette M Buyung-Ali, Teri M Knight, and Andrew S Pullin. A systematic review of evidence for the added benefits to health of exposure to natural environments. *BMC Public Health*, 10(1):456, 2010.

[54] Nicholas D Bowman and Ron Tamborini. Task demand and mood repair: The intervention potential of computer games. *New Media & Society*, 14(8):1339–1357, 2012.

[55] Nicholas D Bowman and Ron Tamborini. “in the mood to game”: Selective exposure and mood management processes in computer game play. *New Media & Society*, 17(3):375–393, 2015.

[56] Margaret M Bradley and Peter J Lang. Measuring emotion: the self-assessment manikin and the semantic differential. *Journal of Behavior Therapy and Experimental Psychiatry*, 25(1):49–59, 1994.

[57] Elmar Brähler, Jörg Schumacher, and Bernhard Strauß. *Diagnostische Verfahren in der Psychotherapie*. Hogrefe, Verlag für Psychologie, 2002.

[58] Gregory N Bratman, J Paul Hamilton, and Gretchen C Daily. The impacts of nature experience on human cognitive function and mental health. *Annals of the New York Academy of Sciences*, 1249(1):118–136, 2012.

[59] Juani Bretón-López, Soledad Quero, Cristina Botella, Azucena García-Palacios, Rosa Maria Baños, and Mariano Alcañiz. An augmented reality system validation for the treatment of cockroach phobia. *Cyberpsychology, Behavior, and Social Networking*, 13(6):705–710, 2010.

[60] John Brooke et al. Sus-a quick and dirty usability scale. *Usability Evaluation in Industry*, 189(194):4–7, 1996.

[61] Matthew HEM Browning, Katherine J Mimnaugh, Carena J van Riper, Heidemarie K Laurent, and Steven M LaValle. Can simulated nature support mental health? comparing short, single-doses of 360-degree nature videos in virtual reality with the outdoors. *Frontiers in Psychology*, 10, 2019.

[62] Matthew HEM Browning, Fatemeh Saeidi-Rizi, Olivia McAnirlin, Hyunseo Yoon, and Yue Pei. The role of methodological choices in the effects of experimental exposure to simulated natural landscapes on human health and cognitive performance: A systematic review. *Environment and Behavior*, page 0013916520906481, 2020.

[63] Ed Bullmore and Olaf Sporns. Complex brain networks: graph theoretical analysis of structural and functional systems. *Nature Reviews Neuroscience*, 10(3):186–198, 2009.

[64] Elizabeth Burns and Ramakrishna Kakara. Deaths from falls among persons aged$\geq$ 65 years-united states, 2007–2016. *Morbidity and Mortality Weekly Report*, 67(18):509, 2018.

[65] Alison L Calear and Helen Christensen. Review of internet-based prevention and treatment programs for anxiety and depression in children and adolescents. *Medical Journal of Australia*, 192:S12–S14, 2010.

[66] Vince D Calhoun, Tulay Adali, Godfrey D Pearlson, and James J Pekar. A method for making group inferences from functional mri data using independent component analysis. *Human Brain Mapping*, 14(3):140–151, 2001.

[67] Giovanna Calogiuri, Sigbjørn Litleskare, Kaia A Fagerheim, Tore L Rydgren, Elena Brambilla, and Miranda Thurston. Experiencing nature through immersive virtual environments: environmental perceptions, physical engagement, and affective responses during a simulated nature walk. *Frontiers in Psychology*, 8:2321, 2018.

[68] Mónica S Cameirão, Sergi Bermúdez i Badia, Esther Duarte Oller, and Paul FMJ Verschure. Neurorehabilitation using the virtual reality based rehabilitation gaming system: methodology, design, psychometrics, usability and validation. *Journal of Neuroengineering and Rehabilitation*, 7(1):1–14, 2010.

[69] Jana Campbell and Ulrike Ehlert. Acute psychosocial stress: does the emotional stress response correspond with physiological responses? *Psychoneuroendocrinology*, 37(8):1111–1134, 2012.

[70] Emily Carl, Aliza T Stein, Andrew Levihn-Coon, Jamie R Pogue, Barbara Rothbaum, Paul Emmelkamp, Gordon JG Asmundson, Per Carlbring, and Mark B Powers. Virtual reality exposure therapy for anxiety and related disorders: A meta-analysis of randomized controlled trials. *Journal of Anxiety Disorders*, 61:27–36, 2019.

[71] Alice Chirico, Pietro Cipresso, David B Yaden, Federica Biassoni, Giuseppe Riva, and Andrea Gaggioli. Effectiveness of immersive videos in inducing awe: an experimental study. *Scientific Reports*, 7(1):1–11, 2017.

[72] Alice Chirico and Andrea Gaggioli. When virtual feels real: comparing emotional responses and presence in virtual and natural environments. *Cyberpsychology, Behavior, and Social Networking*, 22(3):220–226, 2019.

[73] Min Ho Chun, Min Cheol Chang, and Sung-Jae Lee. The effects of forest therapy on depression and anxiety in patients with chronic stroke. *International Journal of Neuroscience*, 127(3):199–203, 2017.

[74] Kyungmi Chung, Daeho Lee, and Jin Young Park. Involuntary attention restoration during exposure to mobile-based 360 virtual nature in healthy adults with different levels of restorative experience: Event-related potential study. *Journal of Medical Internet Research*, 20(11):e11152, 2018.

[75] Błażej Cieślik, Justyna Mazurek, Sebastian Rutkowski, Paweł Kiper, Andrea Turolla, and Joanna Szczepańska-Gieracha. Virtual reality in psychiatric disorders: A systematic review of reviews. *Complementary Therapies in Medicine*, page 102480, 2020.

[76] Pietro Cipresso, Irene Alice Chicchi Giglioli, Mariano Alcañiz Raya, and Giuseppe Riva. The past, present, and future of virtual and augmented reality research: a network and cluster analysis of the literature. *Frontiers in Psychology*, 9:2086, 2018.

[77] Jacob Cohen. A power primer. *Psychological Bulletin*, 112(1):155, 1992.

[78] Sheldon Cohen, T Kamarck, R Mermelstein, et al. Perceived stress scale. *Measuring stress: A guide for health and social scientists*, pages 235–283, 1994.

[79] Sandra Cornelisse, Anda H van Stegeren, and Marian Joëls. Implications of psychosocial stress on memory formation in a typical male versus female student sample. *Psychoneuroendocrinology*, 36(4):569–578, 2011.

[80] Ben Cowley, Darryl Charles, Michaela Black, and Ray Hickey. Toward an understanding of flow in video games. *Computers in Entertainment (CIE)*, 6(2):1–27, 2008.

[81] Fulton Timm Crews and Charlotte Ann Boettiger. Impulsivity, frontal lobes and risk for addiction. *Pharmacology Biochemistry and Behavior*, 93(3):237–247, 2009.

[82] Carolina Cruz-Neira. Virtual reality overview. In *SIGGRAPH*, volume 93, pages 1–1, 1993.

[83] Mihaly Csikszentmihalyi. *Beyond boredom and anxiety.* Jossey-Bass, 2000.

[84] Mihaly Csikszentmihalyi. Flow and the foundations of positive psychology: The collected works of mihaly csikszentmihalyi, 2014.

[85] Cubicle Ninjas. *Guided Meditation VR.* Game [Steam, Oculus Store], January 2016. Glen Ellyn, USA.

[86] Alex Dan and Miriam Reiner. Eeg-based cognitive load of processing events in 3d virtual worlds is lower than processing events in 2d displays. *International Journal of Psychophysiology*, 122:75–84, 2017.

[87] Eling D de Bruin, Annina Reith, Manuela Dörflinger, and Kurt Murer. Feasibility of strength-balance training extended with computer game dancing in older people; does it affect dual task costs of walking. *Journal of Novel Physiotherapies*, 1(1):104, 2011.

[88] R Christopher DeCharms. Applications of real-time fmri. *Nature Reviews Neuroscience*, 9(9):720–729, 2008.

[89] Michael H Depledge, Robert J Stone, and WJ Bird. Can natural and virtual environments be used to promote improved human health and wellbeing?, 2011.

[90] Sally S Dickerson and Margaret E Kemeny. Acute stressors and cortisol responses: a theoretical integration and synthesis of laboratory research. *Psychological Bulletin*, 130(3):355, 2004.

[91] Artemisa Rocha Dores, I Almeida, F Barbosa, M Castelo-Branco, L Monteiro, M Reis, L de Sousa, and A Castro Caldas. Effects of emotional valence and three-dimensionality of visual stimuli on brain activation: an fmri study. *NeuroRehabilitation*, 33(4):505–512, 2013.

[92] Julie Doyle, Cathy Bailey, Ben Dromey, and Cliodhna Ni Scanaill. Base-an interactive technology solution to deliver balance and strength exercises to older adults. In *2010 4th International Conference on Pervasive Computing Technologies for Healthcare*, pages 1–5. IEEE, 2010.

[93] Karen Drexler. *Current Definitions of Alcohol Use Disorders and the Use of Validated Questionnaires in Clinical Practice and Research*, pages 11–17. Springer New York, New York, NY, 2014.

[94] Ruixiang Du, Vinayak Jagtap, Yanren Long, Oke Onwuka, and Taskin Padir. Robotics enabled in-home environment screening for fall risks. In *Proceedings of the 2014 Workshop on Mobile Augmented Reality and Robotic Technology-Based Systems*, pages 9–12. ACM, 2014.

[95] David Daniel Ebert, Anna-Carlotta Zarski, Helen Christensen, Yvonne Stikkelbroek, Pim Cuijpers, Matthias Berking, and Heleen Riper. Internet and computer-based cognitive behavioral therapy for anxiety and depression in youth: a meta-analysis of randomized controlled outcome trials. *PloS One*, 10(3):e0119895, 2015.

[96] Paul MG Emmelkamp, M Krijn, AM Hulsbosch, S De Vries, Martijn J Schuemie, and Charles APG van der Mast. Virtual reality treatment versus exposure in vivo: a comparative evaluation in acrophobia. *Behaviour Research and Therapy*, 40(5):509–516, 2002.

[97] Stefan Engeser and Falko Rheinberg. Flow, performance and moderators of challenge-skill balance. *Motivation and Emotion*, 32(3):158–172, 2008.

[98] Stefan Hermann Engeser. Lernmotivation und volitionale handlungssteuerung: Eine längsschnittsuntersuchung beim statistik lernen im psychologiestudium. 2005.

[99] Caroline J Falconer, Aitor Rovira, John A King, Paul Gilbert, Angus Antley, Pasco Fearon, Neil Ralph, Mel Slater, and Chris R Brewin. Embodying self-compassion within virtual reality and its effects on patients with depression. *BJPsych Open*, 2(1):74–80, 2016.

[100] Ingrid C Fedoroff and Steven Taylor. Psychological and pharmacological treatments of social phobia: a meta-analysis. *Journal of Clinical Psychopharmacology*, 21(3):311–324, 2001.

[101] Yasmin Felberbaum, Joel Lanir, and Patrice L Tamar Weiss. Challenges and requirements for technology to support mobility of older adults. In *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems*, page LBW545. ACM, 2018.

[102] Alicia Fernández-Sotos, Antonio Fernández-Caballero, and José M Latorre. Influence of tempo and rhythmic unit in musical emotion regulation. *Frontiers in Computational Neuroscience*, 10, 2016.

[103] Marisa Ferrari, Barbara Harrison, Osamah Rawashdeh, Muawea Rawashdeh, Robert Hammond, and Michael Maddens. A pilot study testing a fall prevention intervention for older adults: determining the feasibility of a five-sensor motion detection system. *Journal of Gerontological Nursing*, 38(1):13–16, 2012.

[104] Bruno Neves Ferreira, Vânia Guimarães, and Hugo Sereno Ferreira. Smartphone based fall prevention exercises. In *2013 IEEE 15th International Conference on e-Health Networking, Applications and Services (Healthcom 2013)*, pages 643–647. IEEE, 2013.

[105] Lars Brorson Fich, Peter Jönsson, Poul Henning Kirkegaard, Mattias Wallergård, Anne Helene Garde, and Åse Hansen. Can architectural design alter the physiological reaction to psychosocial stress? a virtual tsst experiment. *Physiology & Behavior*, 135:91–97, 2014.

[106] Matthew T Fish, Carmen V Russoniello, and Kevin O'Brien. The efficacy of prescribed casual videogame play in reducing symptoms of anxiety: a randomized controlled study. *Games for Health*, 3(5):291–295, 2014.

[107] Matthew T Fish, Carmen V Russoniello, and Kevin O'Brien. Zombies vs. anxiety: An augmentation study of prescribed video game play compared to medication in reducing anxiety symptoms. *Simulation & Gaming*, 49(5):553–566, 2018.

[108] Theresa M Fleming, Lynda Bavin, Karolina Stasiak, Eve Hermansson-Webb, Sally N Merry, Colleen Cheek, Mathijs Lucassen, Ho Ming Lau, Britta Pollmuller, and Sarah Hetrick. Serious games and gamification for mental health: current status and promising directions. *Frontiers in Psychiatry*, 7:215, 2017.

[109] Paul Foley and Clemens Kirschbaum. Human hypothalamus–pituitary–adrenal axis responses to acute psychosocial stress in laboratory settings. *Neuroscience & Biobehavioral Reviews*, 35(1):91–96, 2010.

[110] National Institute for Health and Care Excellence. *Falls in older people: assessing risk and prevention.* NICE, 2013.

[111] Caroline Garcia Forlim, Lukas Bittner, Fariba Mostajeran, Frank Steinicke, Jürgen Gallinat, and Simone Kühn. Stereoscopic rendering via goggles elicits higher functional connectivity during virtual reality gaming. *Frontiers in Human Neuroscience*, 13:365, 2019.

[112] Chloe Foster, Helen Startup, Laura Potts, and Daniel Freeman. A randomised controlled trial of a worry intervention for individuals with persistent persecutory delusions. *Journal of Behavior Therapy and Experimental Psychiatry*, 41(1):45–51, 2010.

[113] Daniel Freeman, Jonathan Bradley, Angus Antley, Emilie Bourke, Natalie DeWeever, Nicole Evans, Emma Černis, Bryony Sheaves, Felicity Waite, Graham Dunn, et al. Virtual reality in the treatment of persecutory delusions: randomised controlled experimental study testing how to reduce delusional conviction. *British Journal of Psychiatry*, 209(1):62–67, 2016.

[114] Daniel Freeman, Graham Dunn, David Fowler, Paul Bebbington, Elizabeth Kuipers, Richard Emsley, Suzanne Jolley, and Philippa Garety. Current paranoid thinking in patients with delusions: the presence of cognitive-affective biases. *Schizophrenia Bulletin*, 39(6):1281–1287, 2012.

[115] Daniel Freeman, Graham Dunn, Helen Startup, Katherine Pugh, Jacinta Cordwell, Helen Mander, Emma Černis, Gail Wingham, Katherine Shirvell, and David Kingdon. Effects of cognitive behaviour therapy for worry on persecutory delusions in patients with psychosis (wit): a parallel, single-blind, randomised controlled trial with a mediation analysis. *The Lancet Psychiatry*, 2(4):305–313, 2015.

[116] Daniel Freeman, Philippa A Garety, Paul E Bebbington, Benjamin Smith, Rebecca Rollinson, David Fowler, Elizabeth Kuipers, Katarzyna Ray, and Graham Dunn. Psychological investigation of the structure of paranoia in a non-clinical population. *The British Journal of Psychiatry*, 186(5):427–435, 2005.

[117] Daniel Freeman, Sarah Reeve, A Robinson, Anke Ehlers, David Clark, Bernhard Spanlang, and Mel Slater. Virtual reality in the assessment, understanding, and treatment of mental health disorders. *Psychological Medicine*, 47(14):2393–2400, 2017.

[118] John Frewen, Bellinda King-Kallimanis, Gerard Boyle, and Rose Anne Kenny. Recent syncope and unexplained falls are associated with poor cognitive performance. *Age and Ageing*, 44(2):282–286, 2014.

[119] Howard Frumkin. Healthy places: exploring the evidence. *American Journal of Public Health*, 93(9):1451–1456, 2003.

[120] Henry Fuchs and Gary Bishop. Research directions in virtual environments, 1992.

[121] Clive J Fullagar and E Kevin Kelloway. Flow at work: An experience sampling approach. *Journal of Occupational and Organizational Psychology*, 82(3):595–615, 2009.

[122] Patricia M Furlan, Nicholas DeMartinis, Edward Schweizer, Karl Rickels, and Irwin Lucki. Abnormal salivary cortisol levels in social phobic patients in response to acute psychological but not physical stress. *Biological Psychiatry*, 50(4):254–259, 2001.

[123] Michael Gaebler, Felix Biessmann, Jan-Peter Lamke, Klaus-Robert Müller, Henrik Walter, and Stefan Hetzer. Stereoscopic depth increases intersubject correlations of brain networks. *NeuroImage*, 100:427–434, 2014.

[124] Katherine R Gamble, James H Howard Jr, and Darlene V Howard. Not just scenery: viewing nature pictures improves executive attention in older adults. *Experimental Aging Research*, 40(5):513–530, 2014.

[125] Azucena Garcia-Palacios, Cristina Botella, H Hoffman, and Sonia Fabregat. Comparing acceptance and refusal rates of virtual reality exposure vs. in vivo exposure by patients with specific phobias. *CyberPsychology & Behavior*, 10(5):722–724, 2007.

[126] Azucena Garcia-Palacios, Hunter G Hoffman, Sheree Kwong See, AMY Tsai, and Cristina Botella. Redefining therapeutic success with virtual reality exposure therapy. *CyberPsychology & Behavior*, 4(3):341–348, 2001.

[127] Hilde AE Geraedts, Wiebren Zijlstra, Wei Zhang, Sjoerd Bulstra, and Martin Stevens. Adherence to and effectiveness of an individually tailored home-based exercise program for frail older adults, driven by mobility monitoring: design of a prospective cohort study. *BMC Public Health*, 14(1):570, 2014.

[128] Kathrin M Gerling, Regan L Mandryk, and Conor Linehan. Long-term use of motion-based video games in care home settings. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, pages 1573–1582. ACM, 2015.

[129] Kathrin Maria Gerling, Jonas Schild, and Maic Masuch. Exergame design for elderly users: the case study of silverbalance. In *Proceedings of the 7th International Conference on Advances in Computer Entertainment Technology*, pages 66–69. ACM, 2010.

[130] Michael A Gigante. Virtual reality: definitions, history and applications. In *Virtual Reality Systems*, pages 3–14. Elsevier, 1993.

[131] Paul Gilbert. *Compassion focused therapy: Distinctive features.* Routledge, 2010.

[132] Paul Gilbert. The origins and nature of compassion focused therapy. *British Journal of Clinical Psychology*, 53(1):6–41, 2014.

[133] Tobias Gleich, Robert C Lorenz, Jürgen Gallinat, and Simone Kühn. NeuroImage Functional changes in the reward circuit in response to gaming-related cues after training with a commercial video game. *NeuroImage*, 152(August 2016):467–475, 2017.

[134] William K Goodman, Johanna Janson, and Jutta M Wolf. Meta-analytical assessment of the effects of protocol variations on cortisol responses to the trier social stress test. *Psychoneuroendocrinology*, 80:26–35, 2017.

[135] James J Gross and Ricardo F Muñoz. Emotion regulation and mental health. *Clinical Psychology: Science and Practice*, 2(2):151–164, 1995.

[136] Edward T Hall, Ray L Birdwhistell, Bernhard Bock, Paul Bohannan, A Richard Diebold Jr, Marshall Durbin, Munro S Edmonson, JL Fischer, Dell Hymes, Solon T Kimball, et al. Proxemics [and comments and replies]. *Current Anthropology*, 9(2/3):83–108, 1968.

[137] Julian Hamm, Arthur G Money, Anita Atwal, and Ioannis Paraskevopoulos. Fall prevention intervention technologies: A conceptual framework and survey of the state of the art. *Journal of Biomedical Informatics*, 59:319–345, 2016.

[138] Patricia Hansen-Ketchum, Patricia Marck, and Linda Reutter. Engaging with nature to promote health: new directions for nursing research. *Journal of Advanced Nursing*, 65(7):1527–1538, 2009.

[139] Sandro Hardy, Stefan Göbel, and Ralf Steinmetz. Adaptable and personalized game-based training system for fall prevention. In *Proceedings of the 21st ACM International Conference on Multimedia*, pages 431–432, 2013.

[140] Sandra R Harris, Robert L Kemmerling, and Max M North. Brief virtual reality therapy for public speaking anxiety. *CyberPsychology & Behavior*, 5(6):543–550, 2002.

[141] Sandra G Hart. Nasa-task load index (nasa-tlx); 20 years later. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, volume 50, pages 904–908. Sage Publications Sage CA: Los Angeles, CA, 2006.

[142] Richard Harte, Liam Glynn, Alejandro Rodríguez-Molinero, Paul MA Baker, Thomas Scharf, Leo R Quinlan, and Gearóid ÓLaighin. A human-centered design methodology to enhance the usability, human factors, and user experience of connected health systems: a three-phase methodology. *JMIR Human Factors*, 4(1), 2017.

[143] Terry Hartig, Florian G Kaiser, and Einar Strumse. Psychological restoration in nature as a source of motivation for ecological behaviour. *Environmental Conservation*, 34(4):291–299, 2007.

[144] Terry Hartig, Marlis Mang, and Gary W Evans. Restorative effects of natural environment experiences. *Environment and Behavior*, 23(1):3–26, 1991.

[145] Marc Hassenzahl, Michael Burmester, and Franz Koller. Attrakdiff: Ein fragebogen zur messung wahrgenommener hedonischer und pragmatischer qualität. In *Mensch & Computer 2003*, pages 187–196. Springer, 2003.

[146] M Hautzinger, F Keller, and Ch Kühner. Beck depressions-inventar - revision (bdi-ii) (2. auflage), 2009.

[147] Martin Hautzinger, Ferdinand Keller, and Christine Kühner. *Beck Depressions-Inventar (BDI-II)*. Harcourt Test Services Frankfurt, 2006.

[148] Michael Havranek, Nicolas Langer, Marcus Cheetham, and Lutz Jäncke. Perspective and agency during video gaming influences spatial presence experience and brain activation patterns. *Behavioral and Brain Functions*, 8(1):1–13, 2012.

[149] Jennifer Anne Healey. *Wearable and automotive systems for affect recognition from physiology*. PhD thesis, Massachusetts Institute of Technology, 2000.

[150] Marcus Hedblom, Bengt Gunnarsson, Behzad Iravani, Igor Knez, Martin Schaefer, Pontus Thorsson, and Johan N Lundström. Reduction of physiological stress by urban green space in a multisensory virtual experiment. *Scientific Reports*, 9(1):1–11, 2019.

[151] Carrie Heeter. Interactivity in the context of designed experiences. *Journal of Interactive Advertising*, 1(1):3–14, 2000.

[152] William S Helton and Katharina Näswall. Short stress state questionnaire. *European Journal of Psychological Assessment*, 2015.

[153] Claudia Hendrix and Woodrow Barfield. The sense of presence within auditory virtual environments. *Presence: Teleoperators & Virtual Environments*, 5(3):290–301, 1996.

[154] Bruno Herbelin. *Virtual reality exposure therapy for social phobia*. PhD thesis, EPFL, 2005.

[155] Thierry Hergueta, Roxy Baker, and Geoffrey C Dunbar. The mini-international neuropsychiatric interview (mini): the development and validation of a structured diagnostic psychiatric interview for dsm-ivand icd-10. *Journal of Clinical Psychiatry*, 59(Suppl 20):2233, 1998.

[156] S Het, N Rohleder, D Schoofs, C Kirschbaum, and OT Wolf. Neuroendocrine and psychometric evaluation of a placebo version of the 'trier social stress test'. *Psychoneuroendocrinology*, 34(7):1075–1086, 2009.

[157] Juan Luis Higuera-Trujillo, Juan López-Tarruella Maldonado, and Carmen Llinares Millán. Psychological and physiological human responses to simulated and real environments: A comparison between photographs, 360 panoramas, and virtual reality. *Applied Ergonomics*, 65:398–409, 2017.

[158] Hunter G Hoffman, David R Patterson, Jeff Magula, Gretchen J Carrougher, Karen Zeltzer, Stephen Dagadakis, and Sam R Sharar. Water-friendly virtual reality pain control during wound care. *Journal of Clinical Psychology*, 60(2):189–195, 2004.

[159] Antoine Hone-Blanchet, Tobias Wensing, and Shirley Fecteau. The use of virtual reality in craving assessment and cue-exposure therapy in substance use disorders. *Frontiers in Human Neuroscience*, 8, 2014.

[160] Omer Horovitz, Brachali Roitburd, Rany Abend, Daniel Ziskind, and Tomer Shechner. Distraction versus training attention away from threat: How to best wait for the dentist? *Australian Journal of Psychology*, 68(3):191–199, 2016.

[161] Gabriela Husain, William Forde Thompson, and E Glenn Schellenberg. Effects of musical tempo and mode on arousal, mood, and spatial abilities. *Music Perception: An Interdisciplinary Journal*, 20(2):151–171, 2002.

[162] Zaheer Hussain and Mark D Griffiths. The attitudes, feelings, and experiences of online gamers: A qualitative analysis. *CyberPsychology & Behavior*, 12(6):747–753, 2009.

[163] Paul Hutton, James Kelly, Ian Lowens, Peter J Taylor, and Sara Tai. Self-attacking and self-reassurance in persecutory delusions: A comparison of healthy, depressed and paranoid individuals. *Psychiatry Research*, 205(1-2):127–136, 2013.

[164] Ioanna Iacovides and Elisa D Mekler. The role of gaming during difficult life experiences. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, pages 1–12, 2019.

[165] Frank F Ibarra, Omid Kardan, MaryCarol R Hunter, Hiroki P Kotabe, Francisco AC Meyer, and Marc G Berman. Image feature types and their predictions of aesthetic preference and naturalness. *Frontiers in Psychology*, 8:632, 2017.

[166] Resonance Technology Inc. Mri video systems — northridge, ca — visuastimdigital. `http://www.mrivideo.com/visuastimdigital.php`. (Accessed on 02/23/2021).

[167] Rick E Ingram. Toward an information-processing analysis of depression. *Cognitive Therapy and Research*, 8(5):443–477, 1984.

[168] T Insel, B Cuthbert, M Garvey, R Heinssen, DS Pine, K Quinn, and PW Wang. Research domain criteria (rdoc): Developing a valid diagnostic framework for research on mental disorders. *American Journal of Psychiatry*, 167(7):748–751, 2010.

[169] Katherine N Irvine and Sara L Warber. Greening healthcare: Practicing as if the natural environment really mattered. *Alternative Therapies in Health and Medicine*, 8:76–83, 2002.

[170] Lutz Jäncke, Marcus Cheetham, and Thomas Baumgartner. Virtual reality and the role of the prefrontal cortex in adults and children. *Frontiers in Neuroscience*, 3:6, 2009.

[171] Hyunju Jo, Chorong Song, and Yoshifumi Miyazaki. Physiological benefits of viewing nature: A systematic review of indoor experiments. *International Journal of Environmental Research and Public Health*, 16(23):4739, 2019.

[172] Marian Joëls, Zhenwei Pu, Olof Wiegert, Melly S Oitzl, and Harm J Krugers. Learning under stress: how does it work? *Trends in Cognitive Sciences*, 10(4):152–158, 2006.

[173] Peter Jönsson, Mattias Wallergård, Kai Österberg, Åse Marie Hansen, Gerd Johansson, and Björn Karlson. Cardiovascular and cortisol reactivity and habituation to a virtual reality version of the trier social stress test: a pilot study. *Psychoneuroendocrinology*, 35(9):1397–1403, 2010.

[174] Martin G Jorgensen, Uffe Laessoe, Carsten Hendriksen, Ole Bruno Faurholt Nielsen, and Per Aagaard. Efficacy of nintendo wii training on mechanical leg muscle function and postural balance in community-dwelling older adults: a randomized controlled trial. *Journals of Gerontology Series A: Biomedical Sciences and Medical Sciences*, 68(7):845–852, 2012.

[175] M Carmen Juan, Mariano Alcaniz, Carlos Monserrat, Cristina Botella, Rosa María Baños, and Belen Guerrero. Using augmented reality to treat phobias. *IEEE Computer Graphics and Applications*, 25(6):31–37, 2005.

[176] Antonia N Kaczkurkin and Edna B Foa. Cognitive-behavioral therapy for anxiety disorders: an update on the empirical evidence. *Dialogues in Clinical Neuroscience*, 17(3):337, 2015.

[177] Isabel L Kampmann, Paul MG Emmelkamp, Dwi Hartanto, Willem-Paul Brinkman, Bonne JH Zijlstra, and Nexhmedin Morina. Exposure to virtual social interactions in the treatment of social anxiety disorder: A randomized controlled trial. *Behaviour Research and Therapy*, 77:147–156, 2016.

[178] Rachel Kaplan. The psychological benefits of nearby nature. In *Role of horticulture in human well-being and social development: A national symposium*, pages 125–133. Arlington, Va.: Timber Press, 1992.

[179] Rachel Kaplan. The nature of the view from home: Psychological benefits. *Environment and Behavior*, 33(4):507–542, 2001.

[180] Rachel Kaplan and Stephen Kaplan. *The experience of nature: A psychological perspective*. CUP Archive, 1989.

[181] Omid Kardan, Emre Demiralp, Michael C Hout, MaryCarol R Hunter, Hossein Karimi, Taylor Hanayik, Grigori Yourganov, John Jonides, and Marc G Berman. Is the preference of natural versus man-made scenes driven by bottom–up processing of the visual features of nature? *Frontiers in Psychology*, 6:471, 2015.

[182] Aaron H Katcher and Alan M Beck. Health and caring for living things. *Anthrozoös*, 1(3):175–183, 1987.

[183] Hiroki Kayama, Kazuya Okamoto, Shu Nishiguchi, Minoru Yamada, Tomohiro Kuroda, and Tomoki Aoyama. Effect of a kinect-based exercise game on improving executive cognitive performance in community-dwelling elderly: case control study. *Journal of Medical Internet Research*, 16(2):e61, 2014.

[184] Margaret E Kemeny. The psychobiology of stress. *Current Directions in Psychological Science*, 12(4):124–129, 2003.

[185] Lucy E Keniger, Kevin J Gaston, Katherine N Irvine, and Richard A Fuller. What are the benefits of interacting with nature? *International Journal of Environmental Research and Public Health*, 10(3):913–935, 2013.

[186] Robert S Kennedy, Norman E Lane, Kevin S Berbaum, and Michael G Lilienthal. Simulator sickness questionnaire: An enhanced method for quantifying simulator sickness. *The International Journal of Aviation Psychology*, 3(3):203–220, 1993.

[187] Hyeyun Kim, Yong Won Lee, Hyo Jin Ju, Bong Jin Jang, and Yeong In Kim. An exploratory study on the effects of forest therapy on sleep quality in patients with gastrointestinal tract cancers. *International Journal of Environmental Research and Public Health*, 16(14):2449, 2019.

[188] Tae-Hoon Kim, Gwang-Woo Jeong, Han-Su Baek, Gwang-Won Kim, Thirunavukkarasu Sundaram, Heoung-Keun Kang, Seung-Won Lee, Hyung-Joong Kim, and Jin-Kyu Song. Human brain activation in response to visual stimulation with rural and urban scenery pictures: a functional magnetic resonance imaging study. *Science of the Total Environment*, 408(12):2600–2607, 2010.

[189] Clemens Kirschbaum, Karl-Martin Pirke, and Dirk H Hellhammer. The 'trier social stress test'–a tool for investigating psychobiological stress responses in a laboratory setting. *Neuropsychobiology*, 28(1-2):76–81, 1993.

[190] Clemens Kirschbaum, Stefan Wüst, and Dirk Hellhammer. Consistent sex differences in cortisol responses to psychological stress. *Psychosomatic Medicine*, 54(6):648–657, 1992.

[191] Evelyne Klinger, Stéphane Bouchard, Patrick Légeron, Stéphane Roy, Françoise Lauer, Isabelle Chemin, and Pierre Nugues. Virtual reality therapy versus cognitive behavior therapy for social phobia: A preliminary controlled study. *CyberPsychology & Behavior*, 8(1):76–88, 2005.

[192] Elisabeth Klumbies, David Braeuer, Juergen Hoyer, and Clemens Kirschbaum. The reaction to social stress in social phobia: discordance between physiological and subjective parameters. *PloS One*, 9(8):e105670, 2014.

[193] Kevin Koban, Johannes Breuer, Diana Rieger, M Rohangis Mohseni, Stephanie Noack, Gary Bente, and Peter Ohler. Playing for the thrill and skill. quiz games as means for mood and competence repair. *Media Psychology*, 22(5):743–768, 2019.

[194] Silvia Erika Kober, Jürgen Kurzmann, and Christa Neuper. Cortical correlate of spatial presence in 2d and 3d interactive virtual reality: an eeg study. *International Journal of Psychophysiology*, 83(3):365–374, 2012.

[195] Oswald D Kothgassner, Anna Felnhofer, Helmut Hlavacs, Leon Beutl, Rupert Palme, Ilse Kryspin-Exner, and Lisa M Glenk. Salivary cortisol and cardiovascular reactivity to a public speaking task in a virtual and real-life environment. *Computers in Human Behavior*, 62:124–135, 2016.

[196] Martina Krämer, Wiebke Lina Seefeldt, Nina Heinrichs, Brunna Tuschen-Caffier, Julian Schmitz, Oliver Tobias Wolf, and Jens Blechert. Subjective, autonomic, and endocrine reactivity during social stress in children with social phobia. *Journal of Abnormal Child Psychology*, 40(1):95–104, 2012.

[197] Deutsches Krebsforschungszentrum. *Alkoholatlas Deutschland 2017*. Pabst Science Publishers, 2017.

[198] Merel Krijn, Paul MG Emmelkamp, Roeline Biemond, Claudius de Wilde de Ligny, Martijn J Schuemie, and Charles APG van der Mast. Treatment of acrophobia in virtual reality: The role of immersion and presence. *Behaviour Research and Therapy*, 42(2):229–239, 2004.

[199] BM Kudielka, A Buske-Kirschbaum, DH Hellhammer, and C Kirschbaum. Hpa axis responses to laboratory psychosocial stress in healthy elderly adults, younger adults, and children: impact of age and gender. *Psychoneuroendocrinology*, 29(1):83–98, 2004.

[200] Brigitte M Kudielka, Nicole C Schommer, Dirk H Hellhammer, and Clemens Kirschbaum. Acute hpa axis responses, heart rate, and mood changes to psychosocial stress (tsst) in humans at different times of day. *Psychoneuroendocrinology*, 29(8):983–992, 2004.

[201] Simone Kühn and Jürgen Gallinat. Amount of lifetime video gaming is positively associated with entorhinal, hippocampal and occipital volume. *Molecular Psychiatry*, 19(7):842, 2014.

[202] Simone Kühn, Tobias Gleich, Robert C Lorenz, Ulman Lindenberger, and Jürgen Gallinat. Playing super mario induces structural brain plasticity: gray matter changes resulting from training with a commercial video game. *Molecular Psychiatry*, 19(2):265, 2014.

[203] Byoung-Suk Kweon, Roger S Ulrich, Verrick D Walker, and Louis G Tassinary. Anger and stress: The role of landscape posters in an office setting. *Environment and Behavior*, 40(3):355–381, 2008.

[204] Daniël Lakens. Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and anovas. *Frontiers in Psychology*, 4:863, 2013.

[205] Bibb Latané. The psychology of social impact. *American Psychologist*, 36(4):343, 1981.

[206] Bibb Latane, Steve Nida, et al. Social impact theory and group influence: A social engineering perspective. *Psychology of Group Influence*, pages 3–34, 1980.

[207] Lothar Laux, Michael Hock, Ralf Bergner-Köther, Volker Hodapp, and Karl-Heinz Renner. *Das State-Trait-Angst-Depressions-Inventar: STADI; Manual*. Hogrefe, Göttingen, 2013.

[208] Lothar Laux, Michael Hock, Bergner-Köther Ralf, Hodapp Volker, and Renner Karl-Heinz. Stadi: State-trait-angst-depressions-inventar. 2013.

[209] Phil Leather, Mike Pyrgas, Di Beale, and Claire Lawrence. Windows in the workplace: Sunlight, view, and occupational stress. *Environment and Behavior*, 30(6):739–762, 1998.

[210] Hyun Jin Lee and Sung Ae Son. Qualitative assessment of experience on urban forest therapy program for preventing dementia of the elderly living alone in low-income class. *Journal of People, Plants, and Environment*, 21:565–574, 2018.

[211] Hyun Jin Lee, Yong-Hoon Son, Songyi Kim, and Dong Kun Lee. Healing experiences of middle-aged women through an urban forest therapy program. *Urban Forestry & Urban Greening*, 38:383–391, 2019.

[212] I-Jui Lee, Chien-Hsu Chen, and Kuo-Ping Chang. Augmented reality technology combined with three-dimensional holography to train the mental rotation ability of older adults. *Computers in Human Behavior*, 65:488–500, 2016.

[213] Jang-Han Lee, Hyoseok Kwon, Joonho Choi, and Byung-Hwan Yang. Cue-Exposure Therapy to Decrease Alcohol Craving in Virtual Environment. *CyberPsychology & Behavior*, 10(5):617–623, 2007.

[214] Jang-Han Lee, Youngsik Lim, Brenda K Wiederhold, and Simon J Graham. A functional magnetic resonance imaging (fmri) study of cue-induced smoking craving in virtual environments. *Applied Psychophysiology and Biofeedback*, 30(3):195–204, 2005.

[215] Juyoung Lee, B-J Park, Yuko Tsunetsugu, Tatsuro Ohira, Takahide Kagawa, and Yoshifumi Miyazaki. Effect of forest bathing on physiological and psychological responses in young japanese male subjects. *Public Health*, 125(2):93–100, 2011.

[216] Juyoung Lee, Bum-Jin Park, Yuko Tsunetsugu, Takahide Kagawa, and Yoshifumi Miyazaki. Restorative effects of viewing real forest landscapes, based on a comparison with urban landscapes. *Scandinavian Journal of Forest Research*, 24(3):227–234, 2009.

[217] Juyoung Lee, Yuko Tsunetsugu, Norimasa Takayama, Bum-Jin Park, Qing Li, Chorong Song, Misako Komatsu, Harumi Ikei, Liisa Tyrväinen, Takahide Kagawa, et al. Influence of forest therapy on cardiovascular relaxation in young adults. *Evidence-Based Complementary and Alternative Medicine*, 2014, 2014.

[218] Joelle LeMoult, K Lira Yoon, and Jutta Joormann. Affective priming in major depressive disorder. *Frontiers in Integrative Neuroscience*, 6:76, 2012.

[219] Charles A Lewis. *Green nature/human nature: The meaning of plants in our lives.* University of Illinois Press, 1996.

[220] Jinhui Li, Yin-Leng Theng, and Schubert Foo. Game-based digital interventions for depression therapy: a systematic review and meta-analysis. *Cyberpsychology, Behavior, and Social Networking*, 17(8):519–527, 2014.

[221] William HC Li, Joyce OK Chung, and Eva KY Ho. The effectiveness of therapeutic play, using virtual reality computer games, in promoting the psychological well-being of children hospitalised with cancer. *Journal of Clinical Nursing*, 20(15-16):2135–2143, 2011.

[222] Sha Liang. Design principles of augmented reality focusing on the ageing population. In *Proceedings of the 30th International BCS Human Computer Interaction Conference*, pages 1–7. BCS Learning & Development Ltd., 2016.

[223] Michael R Liebowitz. Social phobia. In *Anxiety*, volume 22, pages 141–173. Karger Publishers, 1987.

[224] Tania M Lincoln, Felicitas Hohenhaus, and Maike Hartmann. Can paranoid thoughts be reduced by targeting negative emotions and self-esteem? an experimental investigation of a brief compassion-focused intervention. *Cognitive Therapy and Research*, 37(2):390–402, 2013.

[225] Tania M Lincoln and Eva Keller. Delusions and hallucinations in students compared to the general population. *Psychology and Psychotherapy: Theory, Research and Practice*, 81(3):231–235, 2008.

[226] Matthew Lombard and Theresa Ditton. At the heart of it all: The concept of presence. *Journal of Computer-Mediated Communication*, 3(2):JCMC321, 1997.

[227] Jack M Loomis, James J Blascovich, and Andrew C Beall. Immersive virtual environment technology as a basic research tool in psychology. *Behavior Research Methods, Instruments, & Computers*, 31(4):557–564, 1999.

[228] B Lörch. Cue Exposure. *Verhaltenstherapiemanual*, pages 107–111, 2011.

[229] Mathias Luethi, Beat Meier, and Carmen Sandi. Stress effects on working memory, explicit memory, and implicit memory for neutral and emotional stimuli in healthy men. *Frontiers in Behavioral Neuroscience*, 2:5, 2009.

[230] DT Lykken, R Rose, B Luther, and M Maley. Correcting psychophysiological measures for individual differences in range. *Psychological Bulletin*, 66(6):481, 1966.

[231] Cecily Maller, Mardie Townsend, Anita Pryor, Peter Brown, and Lawrence St Leger. Healthy nature healthy people:'contact with nature'as an upstream health promotion intervention for populations. *Health Promotion International*, 21(1):45–54, 2006.

[232] André Marchand and Thorsten Hennig-Thurau. Value creation in the video game industry: Industry economics, consumer benefits, and research opportunities. *Journal of Interactive Marketing*, 27(3):141–157, 2013.

[233] Mark Matthews and Gavin Doherty. In the mood: engaging teenagers in psychotherapy using mobile phones. In *Proceedings of the 2011 CHI Conference on Human Factors in Computing Systems*, pages 2947–2956, 2011.

[234] Sophie L Mayhew and Paul Gilbert. Compassionate mind training with people who hear malevolent voices: A case series report. *Clinical Psychology & Psychotherapy*, 15(2):113–138, 2008.

[235] Elizabeth McAllister, Navjot Bhullar, and Nicola S Schutte. Into the woods or a stroll in the park: How virtual contact with nature impacts positive and negative affect. *International Journal of Environmental Research and Public Health*, 14(7):786, 2017.

[236] Mark E McKinney, Robert J Gatchel, and Paul B Paulus. The effects of audience size on high and low speech-anxious subjects during an actual speaking task. *Basic and Applied Social Psychology*, 4(1):73–87, 1983.

[237] Robert N McLay, Dennis P Wood, Jennifer A Webb-Murphy, James L Spira, Mark D Wiederhold, Jeffrey M Pyne, and Brenda K Wiederhold. A randomized, controlled trial of virtual reality-graded exposure therapy for post-traumatic stress disorder in active duty service members with combat-related post-traumatic stress disorder. *Cyberpsychology, Behavior, and Social Networking*, 14(4):223–229, 2011.

[238] Michael Meehan, Brent Insko, Mary Whitton, and Frederick P Brooks Jr. Physiological measures of presence in stressful virtual environments. *Acm Transactions on Graphics (tog)*, 21(3):645–652, 2002.

[239] Stephanie Mehl, Dirk Werner, and Tania M Lincoln. Does cognitive behavior therapy for psychosis (cbtp) show a sustainable effect on delusions? a meta-analysis. *Frontiers in Psychology*, 6:1450, 2015.

[240] Mary Metcalf, Karen Rossie, Katie Stokes, Christina Tallman, and Bradley Tanner. Virtual reality cue refusal video game for alcohol and cigarette recovery support: Summative study. *JMIR Serious Games*, 6(2), 2018.

[241] Paul Milgram and Fumio Kishino. A taxonomy of mixed reality visual displays. *IEICE TRANSACTIONS on Information and Systems*, 77(12):1321–1329, 1994.

[242] Paul Milgram, Haruo Takemura, Akira Utsumi, and Fumio Kishino. Augmented reality: A class of displays on the reality-virtuality continuum. In *Telemanipulator and Telepresence Technologies*, volume 2351, pages 282–292. International Society for Optics and Photonics, 1995.

[243] Gregory Miller, Nicolas Rohleder, and Steve W Cole. Chronic interpersonal stress predicts activation of pro-and anti-inflammatory signaling pathways six months later. *Psychosomatic Medicine*, 71(1):57, 2009.

[244] Sarah D Miyahira, Raymond A Folen, Hunter G Hoffman, Azucena Garcia-Palacios, James L Spira, and Michelle Kawasaki. The effectiveness of vr exposure therapy for ptsd in returning warfighters. *Annual Review of Cybertherapy and Telemedicine*, 181:128–132, 2012.

[245] Monobanda PLAY. *DEEP*. Game, 2016. Utrecht, Netherlands.

[246] Eva Montero-López, Ana Santos-Ruiz, M Carmen García-Ríos, Raúl Rodríguez-Blázquez, Miguel Pérez-García, and María Isabel Peralta-Ramírez. A virtual reality approach to the trier social stress test: Contrasting two distinct protocols. *Behavior research methods*, 48(1):223–232, 2016.

[247] Emi Morita, S Fukuda, Jun Nagano, N Hamajima, H Yamamoto, Y Iwai, T Nakashima, H Ohira, and TJPH Shirakawa. Psychological effects of forest environments on healthy adults: Shinrin-yoku (forest-air bathing, walking) as a possible method of stress reduction. *Public Health*, 121(1):54–63, 2007.

[248] Fariba Mostajeran, Melik Berk Balci, Frank Steinicke, Simone Kühn, and Jürgen Gallinat. The effects of virtual audience size on social anxiety during public speaking. In *2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pages 303–312. IEEE, 2020.

[249] John Murphy and Bernard Isaacs. The post-fall syndrome. *Gerontology*, 28(4):265–270, 1982.

[250] Michael Naimark. Elements of real-space imaging: a proposed taxonomy. In *Stereoscopic Displays and Applications II*, volume 1457, pages 169–180. International Society for Optics and Photonics, 1991.

[251] Kristin D Neff. The development and validation of a scale to measure self-compassion. *Self and Identity*, 2(3):223–250, 2003.

[252] Kristin D Neff. Self-compassion, self-esteem, and well-being. *Social and Personality Psychology Compass*, 5(1):1–12, 2011.

[253] Michelle G Newman, Kevin E Kachin, Andrea R Zuellig, Micheal J Constantino, and Laurie Cashman-McGrath. The social phobia diagnostic questionnaire: Preliminary validation of a new self-report diagnostic measure of social phobia. *Psychological Medicine*, 33(4):623–635, 2003.

[254] John C Norcross, Rory A Pfund, and James O Prochaska. Psychotherapy in 2022: A delphi poll on its future. *Professional Psychology: Research and Practice*, 44(5):363, 2013.

[255] Hiroko Ochiai, Harumi Ikei, Chorong Song, Maiko Kobayashi, Ako Takamatsu, Takashi Miura, Takahide Kagawa, Qing Li, Shigeyoshi Kumeda, Michiko Imai, et al. Physiological and psychological effects of forest therapy on middle-aged males with high-normal blood pressure. *International Journal of Environmental Research and Public Health*, 12(3):2532–2542, 2015.

[256] Corinna Ogonowski, Konstantin Aal, Daryoush Vaziri, Thomas Von Rekowski, Dave Randall, Dirk Schreiber, Rainer Wieching, and Volker Wulf. Ict-based fall prevention system for older adults: qualitative results from a long-term field study. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 23(5):29, 2016.

[257] Heather Ohly, Mathew P White, Benedict W Wheeler, Alison Bethel, Obioha C Ukoumunne, Vasilis Nikolaou, and Ruth Garside. Attention restoration theory: A systematic review of the attention restoration potential of exposure to natural environments. *Journal of Toxicology and Environmental Health, Part B*, 19(7):305–343, 2016.

[258] APA Presidential Task Force on Evidence-Based Practice et al. Evidence-based practice in psychology. *The American Psychologist*, 61(4):271, 2006.

[259] Adam Palanica, Aleksandra Lyons, Madeline Cooper, Andrew Lee, and Yan Fossat. A comparison of nature and urban environments on creative thinking across different levels of reality. *Journal of Environmental Psychology*, 63:44–51, 2019.

[260] Federica Pallavicini and Alessandro Pepe. Comparing player experience in video games played in virtual reality or on desktop displays: Immersion, flow, and positive emotions. In *Extended Abstracts of the Annual Symposium on Computer-Human Interaction in Play*, pages 195–210, 2019.

[261] American Geriatrics Society Panel on Prevention of Falls in Older Persons and British Geriatrics Society. Summary of the updated american geriatrics society/british geriatrics society clinical practice guideline for prevention of falls in older persons. *Journal of the American Geriatrics Society*, 59(1):148–157, 2011.

[262] Bum-Jin Park, Yuko Tsunetsugu, Tamami Kasetani, Hideki Hirano, Takahide Kagawa, Masahiko Sato, and Yoshifumi Miyazaki. Physiological effects of shinrin-yoku (taking in the atmosphere of the forest)—using salivary cortisol and cerebral activity as indicators—. *Journal of Physiological Anthropology*, 26(2):123–128, 2007.

[263] Eunice Park-Lee, Rachel N Lipari, Sarra L Hedden, Larry A Kroutil, and Jeremy D Porter. Receipt of services for substance use and mental health issues among adults: results from the 2016 national survey on drug use and health. In *CBHSQ Data Review*. Substance Abuse and Mental Health Services Administration (US), 2017.

[264] Thomas D Parsons and Albert A Rizzo. Affective outcomes of virtual reality exposure therapy for anxiety and specific phobias: A meta-analysis. *Journal of Behavior Therapy and Experimental Psychiatry*, 39(3):250–261, 2008.

[265] Peter J. Passmore, Maxine Glancy, Adam Philpot, Amelia Roscoe, Andrew Wood, and Bob Fields. Effects of Viewing Condition on User Experience of Panoramic Video. In Dirk Reiners, Daisuke Iwai, and Frank Steinicke, editors, *ICAT-EGVE 2016 - International Conference on Artificial Reality and Telexistence and Eurographics Symposium on Virtual Environments*. The Eurographics Association, 2016.

[266] Emilee Patrick, Dennis Cosgrove, Aleksandra Slavkovic, Jennifer A Rode, Thom Verratti, and Greg Chiselko. Using a large projection screen as an alternative to head-mounted displays for virtual environments. In *Proceedings of the 2000 CHI conference on Human Factors in Computing Systems*, pages 478–485, 2000.

[267] Marousa Pavlou, Ari Lingeswaran, Rosalyn A Davies, Michael A Gresty, and Adolfo M Bronstein. Simulator based rehabilitation in refractory dizziness. *Journal of Neurology*, 251(8):983–995, 2004.

[268] Mara Pereira Guerreiro, Isa Brito Félix, Afonso Cavaco, Ana Paula Cláudio, Anabela Mendes, João Balsa, Maria Beatriz Carmo, Nuno Pimenta, and Adriana Henriques. Development of a complex intervention to improve adherence to antidiabetic medication in older people using an anthropomorphic virtual assistant software. *Frontiers in Pharmacology*, 10:680, 2019.

[269] Gamini Perhakaran, Azmi Mohd Yusof, Mohd Ezanee Rusli, Mohd Zaliman Mohd Yusoff, Imran Mahalil, and Ahmad Redza Razieff Zainuddin. A study of meditation effectiveness for virtual reality based stress therapy using eeg measurement and questionnaire approaches. In *Innovation in Medicine and Healthcare 2015*, pages 365–373. Springer, 2016.

[270] David-Paul Pertaub, Mel Slater, and Chris Barker. An experiment on public speaking anxiety in response to three different types of virtual audience. *Presence: Teleoperators & Virtual Environments*, 11(1):68–78, 2002.

[271] Russell Pine, Theresa Fleming, Simon McCallum, and Kylie Sutcliffe. The effects of casual videogames on anxiety, depression, stress, and low mood: A systematic review. *Games for Health Journal*, 2020.

[272] Yusuf Pisan, Jaime Garcia Marin, and Karla Felix Navarro. Improving lives: using microsoft kinect to predict the loss of balance for elderly users under cognitive load. In *Proceedings of The 9th Australasian Conference on Interactive Entertainment: Matters of Life and Death*, page 29. ACM, 2013.

[273] Sandra Poeschl and Nicola Doering. Measuring co-presence and social presence in virtual environments–psychometric construction of a german scale for a fear of public

speaking scenario. *Annual Review of Cybertherapy and Telemedicine 2015*, page 58, 2015.

[274] Albert T Poffenberger and BE Barrows. The feeling value of lines. *Journal of Applied Psychology*, 8(2):187, 1924.

[275] Roos MCA Pot-Kolder, Chris NW Geraets, Wim Veling, Marije van Beilen, Anton BP Staring, Harm J Gijsman, Philippe AEG Delespaul, and Mark van der Gaag. Virtual-reality-based cognitive behavioural therapy versus waiting list control for paranoid ideation and social avoidance in patients with psychotic disorders: a single-blind randomised controlled trial. *The Lancet Psychiatry*, 5(3):217–226, 2018.

[276] Jonathan D Power, Kelly A Barnes, Abraham Z Snyder, Bradley L Schlaggar, and Steven E Petersen. Spurious but systematic correlations in functional connectivity mri networks arise from subject motion. *Neuroimage*, 59(3):2142–2154, 2012.

[277] Mark B Powers and Paul MG Emmelkamp. Virtual reality exposure therapy for anxiety disorders: A meta-analysis. *Journal of Anxiety Disorders*, 22(3):561–569, 2008.

[278] Jules Pretty. How nature contributes to mental and physical health. *Spirituality and Health International*, 5(2):68–78, 2004.

[279] Matthew Price, Natasha Mehta, Erin B Tone, and Page L Anderson. Does engagement with exposure yield better outcomes? components of presence as a predictor of treatment response for virtual reality exposure therapy for social phobia. *Journal of Anxiety Disorders*, 25(6):763–770, 2011.

[280] BS Rajaratnam, J Gui Kaien, K Lee Jialin, Kwek SweeSin, S Sim FenRu, Lee Enting, E Ang YiHsia, Ng KeatHwee, Su Yunfeng, W Woo YingHowe, et al. Does the inclusion of virtual reality games within conventional rehabilitation enhance balance retraining after a recent episode of stroke? *Rehabilitation Research and Practice*, 2013, 2013.

[281] Keeren Sundara Rajoo, Daljit Singh Karam, and Nor Akmar Abdul Aziz. Developing an effective forest therapy program to manage academic stress in conservative societies: A multi-disciplinary approach. *Urban Forestry & Urban Greening*, 43:126353, 2019.

[282] David J Ready, Robert J Gerardi, Andrea G Backscheider, Nathan Mascaro, and Barbara Olasov Rothbaum. Comparing virtual reality exposure therapy to present-centered therapy with 11 us vietnam veterans with ptsd. *Cyberpsychology, Behavior, and Social Networking*, 13(1):49–54, 2010.

[283] Greg M Reger, Patricia Koenen-Woods, Kimberlee Zetocha, Derek J Smolenski, Kevin M Holloway, Barbara O Rothbaum, JoAnn Difede, Albert A Rizzo, Amanda Edwards-Stewart, Nancy A Skopp, et al. Randomized controlled trial of prolonged exposure using imaginal exposure vs. virtual reality exposure in active duty soldiers with deployment-related posttraumatic stress disorder (ptsd). *Journal of Consulting and Clinical Psychology*, 84(11):946, 2016.

[284] Leonard Reinecke. Games and recovery: The use of video and computer games to recuperate from stress and strain. *Media Psychology*, 21(3):126–142, 2009.

[285] Leonard Reinecke, Ron Tamborini, Matthew Grizzard, Robert Lewis, Allison Eden, and Nicholas David Bowman. Characterizing mood management as need satisfaction: The effects of intrinsic needs on selective exposure and mood repair. *Journal of Communication*, 62(3):437–453, 2012.

[286] World Health Organization. Who global report on falls prevention in older age. *World Health Organization*, 2008.

[287] Falko Rheinberg, Stefan Engeser, and Regina Vollmeyer. Measuring components of flow: the flow-short-scale. *Proceedings of the 1st International Positive Psychology Summit*, 2002.

[288] Falko Rheinberg, Regina Vollmeyer, and Stefan Engeser. Die erfassung des flow-erlebens. 2003.

[289] Diana Rieger, Lena Frischlich, Tim Wulf, Gary Bente, and Julia Kneer. Eating ghosts: The underlying mechanisms of mood repair via interactive and noninteractive media. *Psychology of Popular Media Culture*, 4(2):138, 2015.

[290] Diana Rieger, Tim Wulf, Julia Kneer, Lena Frischlich, and Gary Bente. The winner takes it all: The effect of in-game success and need satisfaction on mood repair and enjoyment. *Computers in Human Behavior*, 39:281–286, 2014.

[291] Melvin G Rigg. The mood effects of music: A comparison of data from four investigators. *The Journal of Psychology*, 58(2):427–438, 1964.

[292] Albert Rizzo, Sebastian Thomas Koenig, et al. Is clinical virtual reality ready for primetime? *Neuropsychology*, 31(8):877, 2017.

[293] Albert Skip Rizzo and Gerard Jounghyun Kim. A swot analysis of the field of virtual reality rehabilitation and therapy. *Presence: Teleoperators & Virtual Environments*, 14(2):119–146, 2005.

[294] M Clare Robertson, A John Campbell, Melinda M Gardner, and Nancy Devlin. Preventing injuries in older people by preventing falls: A meta-analysis of individual-level data. *Journal of the American Geriatrics Society*, 50(5):905–911, 2002.

[295] Geneviève Robillard, Stéphane Bouchard, Stéphanie Dumoulin, Tanya Guitard, and E Klinger. Using virtual humans to alleviate social anxiety: preliminary report from a comparative outcome study. *Stud Health Technol Inform*, 154:57–60, 2010.

[296] Christina Röcke and Daniel Grühn. German translation of the panas-x. *Unpublished manuscript, Free University Berlin*, 2003.

[297] Karin Roelofs, Jacobien van Peer, Ed Berretty, Paula de Jong, Philip Spinhoven, and Bernet M Elzinga. Hypothalamus–pituitary–adrenal axis hyperresponsiveness is associated with increased social avoidance behavior in social phobia. *Biological Psychiatry*, 65(4):336–343, 2009.

[298] Mary Beth Rosson, John M Carroll, and Natalie Hill. *Usability engineering: scenario-based development of human-computer interaction.* Morgan Kaufmann, 2002.

[299] Barbara Olasov Rothbaum, Larry Hodges, Samantha Smith, Jeong Hwan Lee, and Larry Price. A controlled study of virtual reality exposure therapy for the fear of flying. *Journal of Consulting and Clinical Psychology*, 68(6):1020, 2000.

[300] Mikail Rubinov and Olaf Sporns. Complex network measures of brain connectivity: uses and interpretations. *Neuroimage*, 52(3):1059–1069, 2010.

[301] Ana Santos Ruiz, Maria Isabel Peralta-Ramirez, Maria Carmen Garcia-Rios, Miguel A Muñoz, Nuria Navarrete-Navarrete, and Antonia Blazquez-Ortiz. Adaptation of the trier social stress test to virtual reality: psycho-physiological and neuroendocrine modulation. *Journal of CyberTherapy and Rehabilitation*, 3:405–415, 2010.

[302] Michael A Rupp, Richard Sweetman, Alejandra E Sosa, Janan A Smither, and Daniel S McConnell. Searching for affective and cognitive restoration: Examining the restorative effects of casual video game play. *Human Factors*, 59(7):1096–1107, 2017.

[303] James A Russell. A circumplex model of affect. *Journal of Personality and Social Psychology*, 39(6):1161, 1980.

[304] Carmen V Russoniello, Matthew Fish, and Kevin O'Brien. The efficacy of casual videogame play in reducing clinical depression: a randomized controlled study. *Games for Health*, 2(6):341–346, 2013.

[305] Carmen V Russoniello, Matthew T Fish, and Kevin O'Brien. The efficacy of playing videogames compared with antidepressants in reducing treatment-resistant symptoms of depression. *Games for Health Journal*, 8(5):332–338, 2019.

[306] Carmen V Russoniello, Kevin O'Brien, and Jennifer M Parks. Eeg, hrv and psychological correlates while playing bejeweled ii: A randomized controlled study. *Annual review of cybertherapy and telemedicine*, 7(1):189–192, 2009.

[307] Carmen V Russoniello, Kevin O'Brien, and Jennifer M Parks. The effectiveness of casual video games in improving mood and decreasing stress. *Journal of CyberTherapy & Rehabilitation*, 2(1):53–66, 2009.

[308] Richard M Ryan and Edward L Deci. Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist*, 55(1):68, 2000.

[309] William S Helton and Katharina Naswall. Short stress state questionnaire. *European Journal of Psychological Assessment*, 1:1–11, 01 2014.

[310] Michael Sailer. *Die Wirkung von Gamification auf Motivation und Leistung*. Springer, 2016.

[311] Rafael Saracchini, Carlos Catalina-Ortega, and Luca Bordoni. A mobile augmented reality assistive technology for the elderly. *Comunicar*, 23(45):65–74, 2015.

[312] Morgan Faith Schebella, Delene Weber, Lisa Schultz, and Philip Weinstein. The nature of reality: Human stress recovery during exposure to biodiverse, multisensory virtual environments. *International Journal of Environmental Research and Public Health*, 17(1):56, 2020.

[313] Björn Schlier, Steffen Moritz, and Tania M Lincoln. Measuring fluctuations in paranoia: Validity and psychometric properties of brief state versions of the paranoia checklist. *Psychiatry Research*, 241:323–332, 2016.

[314] Dieter Schmalstieg and Tobias Hollerer. *Augmented reality: principles and practice*. Addison-Wesley Professional, 2016.

[315] Daniel Schoene, Stephen R Lord, Kim Delbaere, Connie Severino, Thomas A Davies, and Stuart T Smith. A randomized controlled pilot study of home-based step training in older people using videogame technology. *PloS One*, 8(3):e57734, 2013.

[316] Thomas Schubert, Frank Friedmann, and Holger Regenbrecht. The experience of presence: Factor analytic insights. *Presence: Teleoperators & Virtual Environments*, 10(3):266–281, 2001.

[317] TW Schubert, Frank Friedmann, and HT Regenbrecht. Decomposing the sense of presence: Factor analytic insights. In *2nd International Workshop on Presence*, volume 1999, 1999.

[318] Lars Schulze, Babette Renneberg, and Janek S Lobmaier. Gaze perception in social anxiety and social anxiety disorder. *Frontiers in Human Neuroscience*, 7, 2013.

[319] Nicola S Schutte, Navjot Bhullar, Emma J Stilinović, and Katheryn Richardson. The impact of virtual environments on restorativeness and affect. *Ecopsychology*, 9(1):1–7, 2017.

[320] Vicky Scott, Marty Pearce, and Cate Pengelly. Technical report: Injury resulting from falls among canadians age 65 and over on the analysis of data from the canadian community health survey, cycle 2.1. *Report on Seniors' Falls in Canada*, 2005.

[321] Bruno Seraglia, Luciano Gamberini, Konstantinos Priftis, Pietro Scatturin, Massimiliano Martinelli, and Simone Cutini. An exploratory fnirs study with immersive virtual reality: a new method for technical implementation. *Frontiers in Human Neuroscience*, 5:176, 2011.

[322] Saya Shacham. A shortened version of the profile of mood states. *Journal of Personality Assessment*, 47(3):305–306, 1983.

[323] Lubna Bte Iskhandar Shah, Samantha Torres, Premarani Kannusamy, Cecilia Mui Lee Chng, Hong-Gu He, and Piyanee Klainin-Yobas. Efficacy of the virtual reality-based stress management program on stress-related variables in people with mood disorders: the feasibility study. *Archives of Psychiatric Nursing*, 29(1):6–13, 2015.

[324] Jason M Sharbanee, Litje Hu, Werner GK Stritzke, Reinout W Wiers, Mike Rinck, and Colin MacLeod. The effect of approach/avoidance training on alcohol consumption is mediated by change in alcohol action tendency. *PloS One*, 9(1):e85855, 2014.

[325] Youssef Shiban, Julia Diemer, Simone Brandl, Rebecca Zack, Andreas Mühlberger, and Stefan Wüst. Trier social stress test in vivo and in virtual reality: Dissociation of response domains. *International Journal of Psychophysiology*, 110:47–55, 2016.

[326] Hiroyuki Shimada, Megumi Suzukawa, Anne Tiedemann, Kumiko Kobayashi, Hideyo Yoshida, and Takao Suzuki. Which neuromuscular or cognitive test is the optimal screening tool to predict falls in frail community-dwelling older people? *Gerontology*, 55(5):532–538, 2009.

[327] Won Sop Shin, Poung Sik Yeoun, Rhi Wha Yoo, and Chang Seob Shin. Forest experience and psychological health benefits: the state of the art and future prospect in korea. *Environmental Health and Preventive Medicine*, 15(1):38, 2010.

[328] Tiffany E Shubert. Evidence-based exercise prescription for balance and falls prevention: a current review of the literature. *Journal of Geriatric Physical Therapy*, 34(3):100–108, 2011.

[329] Paula Alexandra Silva, Francisco Nunes, Ana Vasconcelos, Maureen Kerwin, Ricardo Moutinho, and Pedro Teixeira. Using the smartphone accelerometer to monitor fall risk while playing a game: the design and usability evaluation of dance! don't fall. In *International Conference on Augmented Cognition*, pages 754–763. Springer, 2013.

[330] Raquel A Silva, Kim Rogers, and Timothy J Buckley. Advancing environmental epidemiology to assess the beneficial influence of the natural environment on human health and well-being. *Environmental Science & Technology*, 52(17):9545–9555, 2018.

[331] Patrícia Silveira, Rolf Van De Langenberg, Eva Van Het Reve, Florian Daniel, Fabio Casati, and Eling D De Bruin. Tablet-based strength-balance training to motivate and improve adherence to exercise in independently living older people: a phase ii preclinical exploratory trial. *Journal of Medical Internet Research*, 15(8):e159, 2013.

[332] Devinder Kaur Ajit Singh, BS Rajaratnam, V Palaniswamy, VP Raman, PS Bong, and H Pearson. Effects of balance-focused interactive games compared to therapeutic balance classes for older women. *Climacteric*, 16(1):141–146, 2012.

[333] Paul Skalski and Ron Tamborini. The role of social presence in interactive agent-based persuasion. *Media Psychology*, 10(3):385–413, 2007.

[334] Mel Slater. Place illusion and plausibility can lead to realistic behaviour in immersive virtual environments. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364(1535):3549–3557, 2009.

[335] Mel Slater, D-P Pertaub, and Anthony Steed. Public speaking in virtual reality: Facing an audience of avatars. *IEEE Computer Graphics and Applications*, 19(2):6–9, 1999.

[336] Mel Slater, David-Paul Pertaub, Chris Barker, and David M Clark. An experimental study on fear of public speaking using a virtual environment. *CyberPsychology & Behavior*, 9(5):627–633, 2006.

[337] Mel Slater and Martin Usoh. Representations systems, perceptual position, and presence in immersive virtual environments. *Presence: Teleoperators & Virtual Environments*, 2(3):221–233, 1993.

[338] Mel Slater and Sylvia Wilbur. A framework for immersive virtual environments (five): Speculations on the role of presence in virtual environments. *Presence: Teleoperators & Virtual Environments*, 6(6):603–616, 1997.

[339] Semyon M Slobounov, William Ray, Brian Johnson, Elena Slobounov, and Karl M Newell. Modulation of cortical activity in 2d versus 3d virtual reality environments: an eeg study. *International Journal of Psychophysiology*, 95(3):254–260, 2015.

[340] Jordan W Smith. Immersive virtual environment technology to supplement environmental perception, preference and behavior research: a review with applications. *International Journal of Environmental Research and Public Health*, 12(9):11486–11505, 2015.

[341] Stuart T Smith, Thomas A Davies, and Jamie Lennox. Step training system: An ict solution to measure and reduce fall risk in older adults. In *2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 7033–7035. IEEE, 2013.

[342] Chorong Song, Harumi Ikei, Takahide Kagawa, and Yoshifumi Miyazaki. Effects of walking in a forest on young women. *International Journal of Environmental Research and Public Health*, 16(2):229, 2019.

[343] Xiao-Wei Song, Zhang-Ye Dong, Xiang-Yu Long, Su-Fang Li, Xi-Nian Zuo, Chao-Zhe Zhu, Yong He, Chao-Gan Yan, and Yu-Feng Zang. Rest: a toolkit for resting-state functional magnetic resonance imaging data processing. *PloS One*, 6(9):e25031, 2011.

[344] David Sparrow, Daniel J Gottlieb, Deborah DeMolles, and Roger A Fielding. Increases in muscle strength and balance using a resistance training program administered via a

telecommunications system in older adults. *Journals of Gerontology Series A: Biomedical Sciences and Medical Sciences*, 66(11):1251–1257, 2011.

[345] Paul Stallard, Thomas Richardson, Sophie Velleman, and Megan Attwood. Computerized cbt (think, feel, do) for depression and anxiety in children and adolescents: outcomes and feedback from a pilot randomized controlled trial. *Behavioural and Cognitive Psychotherapy*, 39(3):273, 2011.

[346] Katarzyna Stawarz, Chris Preist, Deborah Tallon, Laura Thomas, Katrina Turner, Nicola Wiles, David Kessler, Roz Shafran, and David Coyle. Integrating the digital and the traditional to deliver therapy for depression: Lessons from a pragmatic study. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, pages 1–14, 2020.

[347] Murray B Stein and Dan J Stein. Social anxiety disorder. *The Lancet*, 371(9618):1115–1125, 2008.

[348] Frank Steinicke. *Being really virtual*. Springer, 2016.

[349] Matt P Stevenson, Theresa Schilhab, and Peter Bentsen. Attention restoration theory ii: A systematic review to clarify attention processes affected by exposure to natural environments. *Journal of Toxicology and Environmental Health, Part B*, 21(4):227–268, 2018.

[350] John R Stilgoe. Gone barefoot lately? *American Journal of Preventive Medicine*, 20(3):243–244, 2001.

[351] Bernd Strauss. Social facilitation in motor tasks: a review of research and theory. *Psychology of Sport and Exercise*, 3(3):237–256, 2002.

[352] Daniel Stucht, K Appu Danishad, Peter Schulze, Frank Godenschweger, Maxim Zaitsev, and Oliver Speck. Highest resolution in vivo human brain mri using prospective motion correction. *PloS One*, 10(7):e0133921, 2015.

[353] Xiaoqi Sun, Suzanne H So, Raymond CK Chan, Chui-De Chiu, and Patrick WL Leung. Worry and metacognitions as predictors of the development of anxiety and paranoia. *Scientific Reports*, 9(1):1–10, 2019.

[354] S Shyam Sundar, Qian Xu, and Saraswathi Bellur. Designing interactivity in media interfaces: A communications perspective. In *Proceedings of the 2010 CHI Conference on Human Factors in Computing Systems*, pages 2247–2256, 2010.

[355] Jidong Sung, Jong-Min Woo, Won Kim, Seoung-Kyeon Lim, and Eun-Joo Chung. The effect of cognitive behavior therapy-based "forest therapy" program on blood pressure, salivary cortisol level, and quality of life in elderly hypertensive patients. *Clinical and Experimental Hypertension*, 34(1):1–7, 2012.

[356] Ivan Sutherland. The ultimate display. 1965.

[357] Jamie L Tait, Rachel L Duckham, Catherine M Milte, Luana C Main, and Robin M Daly. Influence of sequential vs. simultaneous dual-task exercise training on cognitive function in older adults. *Frontiers in Aging Neuroscience*, 9:368, 2017.

[358] Sarah Tanford and Steven Penrod. Social influence model: A formal integration of research on majority and minority influence processes. *Psychological Bulletin*, 95(2):189, 1984.

[359] Lynne M Taylor, Ralph Maddison, Leila A Pfaeffli, Jonathan C Rawstorn, Nicholas Gant, and Ngaire M Kerse. Activity and energy expenditure in older people playing active video games. *Archives of Physical Medicine and Rehabilitation*, 93(12):2281–2286, 2012.

[360] Matthew JD Taylor, Teshk Shawis, Rebecca Impson, Katie Ewins, Darren McCormick, and Murray Griffin. Nintendo wii as a training tool in falls prevention rehabilitation: case studies. *Journal of the American Geriatrics Society*, 60(9):1781–1783, 2012.

[361] Carolyn M Tennessen and Bernadine Cimprich. Views to nature: Effects on attention. *Journal of Environmental Psychology*, 15(1):77–85, 1995.

[362] Thatgamecompany. *Flower.* Game [Playstation 3], February 2009. Los Angeles, USA.

[363] Viviane Thewissen, Richard P Bentall, Margreet Oorschot, Joost à Campo, Thom van Lierop, Jim van Os, and Inez Myin-Germeys. Emotions, self-esteem, and paranoid episodes: An experience sampling study. *British Journal of Clinical Psychology*, 50(2):178–195, 2011.

[364] Audrey Tse, Charlene Jennett, Joanne Moore, Zillah Watson, Jacob Rigby, and Anna L Cox. Was i there? impact of platform and headphones on 360 video immersion. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, pages 2967–2974, 2017.

[365] Kostas M Tsiouris, Dimitrios Gatsios, Vassilios Tsakanikas, Athanasios A Pardalis, Ioannis Kouris, Thelma Androutsou, Marilena Tarousi, Natasa Vujnovic Sedlar, Iason Somarakis, Fariba Mostajeran, Nenad Filipovic, Harm op den Akkeri, Dimitrios D. Koutsourisa, and Dimitrios I. Fotiadis. Designing interoperable telehealth platforms: bridging iot devices with cloud infrastructures. *Enterprise Information Systems*, pages 1–25, 2020.

[366] Christiana Tsiourti, Emilie Joly, Cindy Wings, Maher Ben Moussa, and Katarzyna Wac. Virtual assistive companions for older adults: qualitative field study and design implications. In *Proceedings of the 8th International Conference on Pervasive Computing Technologies for Healthcare*, pages 57–64. ICST (Institute for Computer Sciences, Social-Informatics and . . . , 2014.

[367] Cosimo Tuena, Elisa Pedroli, Pietro Davide Trimarchi, Alessia Gallucci, Mattia Chiappini, Karine Goulene, Andrea Gaggioli, Giuseppe Riva, Fabrizia Lattanzio, Fabrizio Giunco, et al. Usability issues of clinical and research applications of virtual reality in older people: A systematic review. *Frontiers in Human Neuroscience*, 14:93, 2020.

[368] Shilpa Tyagi, Daniel SY Lim, Wilbert HH Ho, Yun Qing Koh, Vincent Cai, Gerald CH Koh, and Helena Legido-Quigley. Acceptance of tele-rehabilitation by stroke patients: perceived barriers and facilitators. *Archives of Physical Medicine and Rehabilitation*, 99(12):2472–2477, 2018.

[369] Nathalie Tzourio-Mazoyer, Brigitte Landeau, Dimitri Papathanassiou, Fabrice Crivello, Olivier Etard, Nicolas Delcroix, Bernard Mazoyer, and Marc Joliot. Automated anatomical labeling of activations in spm using a macroscopic anatomical parcellation of the mni mri single-subject brain. *Neuroimage*, 15(1):273–289, 2002.

[370] Roger S Ulrich. Natural versus urban scenes: Some psychophysiological effects. *Environment and Behavior*, 13(5):523–556, 1981.

[371] United Nations. Global issues overview. `https://www.un.org/en/sections/issues-depth/global-issues-overview/`. (Accessed on 02/15/2021).

[372] United Nations. Health. `https://www.un.org/en/global-issues/health`. (Accessed on 03/26/2021).

[373] United Nations. World population ageing 2019. `https://www.un.org/development/desa/pd/sites/www.un.org.development.desa.pd/files/files/documents/2020/Jan/un_2019_worldpopulationageing_report.pdf`, 2020. (Accessed on 02/12/2021).

[374] Patricia Valdez and Albert Mehrabian. Effects of color on emotions. *Journal of Experimental Psychology: General*, 123(4):394, 1994.

[375] Lucia R Valmaggia, Fern Day, and Mar Rus-Calafell. Using virtual reality to investigate psychological processes and mechanisms associated with the onset and maintenance of psychosis: a systematic review. *Social Psychiatry and Psychiatric Epidemiology*, 51(7):921–936, 2016.

[376] Mark van der Gaag, Lucia R Valmaggia, and Filip Smit. The effects of individually tailored formulation-based cognitive behavioural therapy in auditory hallucinations and delusions: a meta-analysis. *Schizophrenia Research*, 156(1):30–37, 2014.

[377] Mike van Diest, Jan Stegenga, Heinrich J Wörtche, Klaas Postema, Gijsbertus J Verkerke, and Claudine JC Lamoth. Suitability of kinect for measuring whole body movement patterns during exergaming. *Journal of Biomechanics*, 47(12):2925–2932, 2014.

[378] Marieke Van Rooij, Adam Lobel, Owen Harris, Niki Smit, and Isabela Granic. Deep: A biofeedback virtual reality game for children at-risk for anxiety. In *Proceedings of the*

2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems, pages 1989–1997, 2016.

[379] Daniel Västfjäll. Emotion induction through music: A review of the musical mood induction procedure. *Musicae Scientiae*, 5(1_suppl):173–211, 2001.

[380] C Lee Ventola. Virtual reality in pharmacy: opportunities for clinical, research, and educational applications. *Pharmacy and Therapeutics*, 44(5):267, 2019.

[381] Joe Verghese, Herman Buschke, Lisa Viola, Mindy Katz, Charles Hall, Gail Kuslansky, and Richard Lipton. Validity of divided attention tasks in predicting falls in older individuals: a preliminary study. *Journal of the American Geriatrics Society*, 50(9):1572–1576, 2002.

[382] Daniela Villani, Claudia Carissoli, Stefano Triberti, Antonella Marchetti, Gabriella Gilli, and Giuseppe Riva. Videogames for emotion regulation: a systematic review. *Games for Health Journal*, 7(2):85–99, 2018.

[383] Helene S Wallach, Marilyn P Safir, and Margalit Bar-Zvi. Virtual reality cognitive behavior therapy for public speaking anxiety: a randomized clinical trial. *Behavior Modification*, 33(3):314–338, 2009.

[384] Mattias Wallergård, Peter Jönsson, Gerd Johansson, and Björn Karlson. A virtual reality version of the trier social stress test: A pilot study. *Presence*, 20(4):325–336, 2011.

[385] Xiaobo Wang, Yaxing Shi, Bo Zhang, and Yencheng Chiang. The influence of forest resting environments on stress using virtual reality. *International Journal of Environmental Research and Public Health*, 16(18):3263, 2019.

[386] David Watson and Lee Anna Clark. The panas-x: Manual for the positive and negative affect schedule-expanded form. 1999.

[387] Joanneke Weerdmeester, Marieke van Rooij, Owen Harris, Niki Smit, Rutger CME Engels, and Isabela Granic. Exploring the role of self-efficacy in biofeedback video games. In *Extended Abstracts of the Annual Symposium on Computer-Human Interaction in Play*, pages 453–461, 2017.

[388] United Kingdom Wellcome Department for Imaging Neuroscience, London. Spm - statistical parametric mapping. `https://www.fil.ion.ucl.ac.uk/spm/`. (Accessed on 02/23/2021).

[389] Stefan Westermann, Marie-Luise Kesting, and Tania M Lincoln. Being deluded after being excluded? how emotion regulation deficits in paranoia-prone individuals affect state paranoia during experimentally induced social stress. *Behavior Therapy*, 43(2):329–340, 2012.

[390] Frank White. *The overview effect: Space exploration and human evolution*. American Institute of Aeronautics and Astronautics, 2014.

[391] Brenda K Wiederhold, Dong Pyo Jang, Richard G Gevirtz, Sun I Kim, In-Young Kim, and Mark D Wiederhold. The treatment of fear of flying: a controlled study of imaginal and virtual reality graded exposure therapy. *IEEE Transactions on Information Technology in Biomedicine*, 6(3):218–223, 2002.

[392] Brenda K Wiederhold and Mark D Wiederhold. Lessons learned from 600 virtual reality sessions. *CyberPsychology & Behavior*, 3(3):393–400, 2000.

[393] Uta S Wiemers and Oliver T Wolf. Cortisol broadens memory of a non-stressful social interaction. *Psychopharmacology*, 232(10):1727–1733, 2015.

[394] Reinout W. Wiers, Carolin Eberl, Mike Rinck, Eni S. Becker, and Johannes Lindenmeyer. Retraining automatic action tendencies changes alcoholic patients' approach bias for alcohol and improves treatment outcome. *Psychological Science*, 22(4):490–497, 2011. PMID: 21389338.

[395] Reinout W Wiers, Nieske Van Woerden, Fren TY Smulders, and Peter J De Jong. Implicit and explicit alcohol-related cognitions in heavy and light drinkers. *Journal of Abnormal Psychology*, 111:648, 2002.

[396] Hans-Ulrich Wittchen and Jürgen Hoyer. *Klinische Psychologie & Psychotherapie (Lehrbuch mit Online-Materialien)*. Springer-Verlag, 2011.

[397] Marjorie Woollacott and Anne Shumway-Cook. Attention and the control of posture and gait: a review of an emerging area of research. *Gait & posture*, 16(1):1–14, 2002.

[398] World Health Organization. *Building foundations for eHealth: progress of Member States: report of the WHO Global Observatory for eHealth*. World Health Organization, 2006.

[399] World Health Organization. *National eHealth strategy toolkit*. International Telecommunication Union, 2012.

[400] World Health Organization. *Global diffusion of eHealth: making universal health coverage achievable: report of the third global survey on eHealth*. World Health Organization, 2017.

[401] World Health Organization. Classification of digital health interventions v1. 0: a shared language to describe the uses of digital technology for health. Technical report, World Health Organization, 2018.

[402] World Health Organization. Icd-11 for mortality and morbidity statistics (2018). 2018.

[403] World Health Organization. World health statistics 2019: monitoring health for the sdgs, sustainable development goals. 2019.

[404] Maja Wrzesien, Mariano Alcañiz, Cristina Botella, Jean-Marie Burkhardt, Juana Bretón-López, Mario Ortega, and Daniel Beneito Brotons. The therapeutic lamp: treating small-animal phobias. *IEEE Computer Graphics and Applications*, 33(1):80–86, 2013.

[405] Alan Yusheng Wu and Cosmin Munteanu. Understanding older users' acceptance of wearable interfaces for sensor-based fall risk assessment. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, page 119. ACM, 2018.

[406] David B Yaden, Jonathan Iwry, Kelley J Slack, Johannes C Eichstaedt, Yukun Zhao, George E Vaillant, and Andrew B Newberg. The overview effect: Awe and self-transcendent experience in space flight. *Psychology of Consciousness: Theory, Research, and Practice*, 3(1):1, 2016.

[407] Ilona S Yim, Jodi A Quas, Larry Cahill, and Cathy M Hayakawa. Children's and adults' salivary cortisol responses to an identical psychosocial laboratory stressor. *Psychoneuroendocrinology*, 35(2):241–248, 2010.

[408] Jie Yin, Shihao Zhu, Piers MacNaughton, Joseph G Allen, and John D Spengler. Physiological and cognitive performance of exposure to biophilic indoor environment. *Building and Environment*, 132:255–262, 2018.

[409] Jung Younbo. Promoting active ageing through virtual augmented reality. Jan 2019. (Accessed on 2019/9/20/).

[410] Chia-Pin Yu, Hsiao-Yun Lee, and Xiang-Yi Luo. The effect of virtual reality forest and urban environments on physiological and psychological responses. *Urban Forestry & Urban Greening*, 35:106–114, 2018.

[411] Miao Yu, Adel Rhuma, Syed Mohsen Naqvi, Liang Wang, and Jonathon Chambers. A posture recognition-based fall detection system for monitoring an elderly person in a smart home environment. *IEEE Transactions on Information Technology in Biomedicine*, 16(6):1274–1286, 2012.

[412] Patrick Zimmer, Benjamin Buttlar, Georg Halbeisen, Eva Walther, and Gregor Domes. Virtually stressed? a refined virtual reality adaptation of the trier social stress test (tsst) induces robust endocrine responses. *Psychoneuroendocrinology*, 101:186–192, 2019.

[413] Qi-Hong Zou, Chao-Zhe Zhu, Yihong Yang, Xi-Nian Zuo, Xiang-Yu Long, Qing-Jiu Cao, Yu-Feng Wang, and Yu-Feng Zang. An improved approach to detection of amplitude of low-frequency fluctuation (alff) for resting-state fmri: fractional alff. *Journal of Neuroscience Methods*, 172(1):137–141, 2008.

Erklärung der Urheberschaft

Hiermit versichere ich an Eides statt, dass ich die vorliegende Dissertation selbstständig verfasst und keine anderen als die angegebenen Hilfsmittel - insbesondere keine im Quellenverzeichnis nicht benannten Internet-Quellen – benutzt habe. Alle Stellen, die wörtlich oder sinngemäß aus Veröffentlichungen entnommen wurden, sind als solche kenntlich gemacht. Ich versichere weiterhin, dass ich die Arbeit vorher nicht in einem anderen Prüfungsverfahren eingereicht habe und die eingereichte schriftliche Fassung der auf dem elektronischen Speichermedium entspricht.

Ort, Datum Unterschrift

Erklärung zur Veröffentlichung

Ich stimme der Einstellung der Dissertation in die Bibliothek des Fachbereichs Informatik zu.

Ort, Datum Unterschrift