

UNIVERSITÄTSKLINIKUM HAMBURG-EPPENDORF

Klinik und Poliklinik für Psychiatrie und Psychotherapie

Prof. Dr. med. Jürgen Gallinat

Auswirkungen einer definierten sozialen Stresssituation in der virtuellen Umgebung auf gesunde Probanden/innen mit der Neigung zu sozialer Phobie

Dissertation

zur Erlangung des Grades eines Doktors der Medizin an der Medizinischen Fakultät der
Universität Hamburg.

vorgelegt von:

Melik Berk Balci
aus
Hamburg, Deutschland

Hamburg 2021

Angenommen von der

Medizinischen Fakultät der Universität Hamburg am: 29.03.2022

Veröffentlicht mit Genehmigung der

Medizinischen Fakultät der Universität Hamburg.

Prüfungsausschuss, der/die Vorsitzende: Prof. Dr. Martin Scherer

Prüfungsausschuss, zweite/r Gutachter/in: Prof. Dr. Jürgen Gallinat

Die Daten, auf denen die vorliegende Arbeit basiert, wurden teilweise publiziert:

Fariba Mostajeran, Melik Berk Balci, Frank Steinicke, Simone Kühn, Jürgen Galinat (2020): The Effects of Virtual Audience Size on Social Anxiety during Public Speaking in Proceedings of IEEE Virtual Reality (VR) 2020

Inhaltsverzeichnis

Abbildungsverzeichnis	6
Tabellenverzeichnis.....	7
Abkürzungsverzeichnis	8
1 Einleitung.....	10
1.1 Soziale Phobie.....	10
1.1.1 Epidemiologie	10
1.1.2 Diagnostik und Differenzialdiagnosen.....	10
1.1.3 Verlaufsdiagnostik	11
1.1.4 Therapie	11
1.2 Expositionstherapie.....	12
1.2.1 Wirksamkeit von Expositionstherapien bei sozialer Phobie	12
1.2.2 Schwächen der in vivo-Expositionstherapie	13
1.3 Virtuelle Realität Expositionstherapie (VRET)	13
1.4 Forschungsschwerpunkt dieser Arbeit.....	15
1.5 Hypothesen	16
2 Material und Methoden.....	17
2.1 Aufbau des Experiments.....	17
2.2 Probanden/innen	18
2.3 Materialien	20
2.3.1 Technisches Equipment	20
2.3.2 Cortisolspeichelproben.....	20
2.3.3 Hautleitwiderstand (GSR).....	21
2.3.4 EKG	21
2.3.5 Fragebögen	21
2.3.5.1 Self-Assessment Manikin (SAM).....	22
2.3.5.2 State-Trait-Angst-Depressions-Inventar State (STADI-S).....	22
2.3.5.3 Situativer sozialer Angstfragebogen (SSA)	22
2.3.5.4 Graphisch numerischer Fragebogen	22
2.3.5.5 Co- and Social-Presence Fragebogen	23
2.3.5.6 Igroup Presence Questionnaire (IPQ)	23
2.3.5.7 Simulator Sickness Questionnaire (SSQ).....	23
2.4 Statistisches Vorgehen bei der Datenanalyse.....	24
2.4.1 Allgemeines Vorgehen bei der statistischen Auswertung	24
2.4.2 Grafische Darstellung der Ergebnisse.....	24
2.4.3 Cortisol.....	25
2.4.4 Hautleitwiderstand (GSR).....	25
2.4.5 Herzfrequenz.....	26
2.4.6 Herzfrequenzvariabilität	27
2.4.7 Fragebögen	27
3 Ergebnisse	28
3.1 Cortisol.....	28
3.2 Hautleitwiderstand (GSR)	31

3.3 Herzfrequenz.....	33
3.4 Herzfrequenzvariabilität.....	36
3.4.1 RMSSD.....	36
3.4.2 SDNN	38
3.5 Fragebögen.....	40
3.5.1 Self-Assessment Manikin	40
3.5.2 State-Trait-Angst-Depressions-Inventar State.....	42
3.5.3 Situativer sozialer Angstfragebogen	44
3.5.4 Graphisch numerischer Fragebogen	44
3.5.5 Co- and Social-Presence Questionnaire.....	45
3.5.6 Igroup Presence Questionnaire.....	46
3.5.7 Simulator Sickness Questionnaire.....	46
4 Diskussion.....	47
4.1 Methoden.....	47
4.1.1 Aufbau des Experiments.....	47
4.1.2 Probanden/innen	47
4.1.3 Technisches Equipment und digitale Visualisierung des TSSTs	48
4.2 Ergebnisse.....	49
4.2.1 Cortisol.....	49
4.2.2 Hautleitwiderstand (GSR).....	49
4.2.3 Herzfrequenz.....	50
4.2.4 Herzfrequenzvariabilität	50
4.2.5 Fragebögen	51
4.3 Ausblick für weitere Forschungen.....	52
5 Zusammenfassung	54
6 Zusammenfassung auf Englisch.....	55
7 Literaturverzeichnis.....	56
8 Danksagung	59
9 Lebenslauf	60

Abbildungsverzeichnis

Abbildung 1: Vergleich der Cortisolwerte zwischen den VR-Konditionen.....	29
Abbildung 2: Cortisolverlauf der in vivo- und N3-Gruppe.....	29
Abbildung 3: Cortisolverlauf der Responder und Non-Responder im Vergleich.....	30
Abbildung 4: Vergleich des Hautleitwiderstandes zwischen den VR-Konditionen	32
Abbildung 5: GSR-Verlauf der in vivo- und N3-Gruppe.....	33
Abbildung 6: Vergleich der Herzfrequenzen zwischen den VR-Konditionen.....	34
Abbildung 7: Herzfrequenzverlauf der in vivo- und N3-Gruppe	35
Abbildung 8: RMSSD für die einzelnen VR-Konditionen	37
Abbildung 9: RMSSD-Verlauf der in vivo- und N3-Gruppe.....	38
Abbildung 10: SDNN für die einzelnen VR-Konditionen.....	39
Abbildung 11: SDNN-Verlauf der in vivo- und N3-Gruppe	40

Tabellenverzeichnis

Tabelle 1: Charakteristika der Probanden/innen	20
Tabelle 2: Vergleich der Cortisolwerte für die gesamte Testungspopulation	28
Tabelle 3: Vergleich der Cortisolwerte zwischen der in vivo- und N3-Gruppe.....	30
Tabelle 4: Vergleich des Hautleitwiderstandes für die gesamte Testungspopulation.....	31
Tabelle 5: Vergleich des Hautleitwiderstandes zwischen der in vivo- und N3-Gruppe	33
Tabelle 6: Vergleich der Herzfrequenzen für die gesamte Testungspopulation.....	34
Tabelle 7: Vergleich der Herzfrequenzen zwischen der in vivo- und N3-Gruppe.....	36
Tabelle 8: Vergleich der RMSSD-Werte für die gesamte Testungspopulation	36
Tabelle 9: Vergleich der SDNN-Werte für die gesamte Testungspopulation	39
Tabelle 10: SAM-Fragebogen mit den Subskalen für die gesamte Testungspopulation.....	41
Tabelle 11: SAM-Fragebogen im Vergleich zwischen der in vivo- und N3-Gruppe	42
Tabelle 12: STADI-S für die gesamte Testungspopulation	43
Tabelle 13: STADI-S Fragebogen im Vergleich zwischen der in vivo- und N3-Gruppe.....	44
Tabelle 14: SSA für die gesamte Testungspopulation.....	44
Tabelle 15: Graphischer Fragebogen für die gesamte Testungspopulation.....	45
Tabelle 16: Co- and Social-Presence Questionnaire für die VR-Konditionen	46
Tabelle 17: Igroup Presence Questionnaire	46
Tabelle 18: Vergleich zwischen Prä- und Post-Werten des SSQ.....	46

Abkürzungsverzeichnis

Arithm.	Arithmetisch
Bpm	Beats per minute
CI	Konfidenzintervall
DSM	Diagnostic and Statistical Manual of Mental Disorders
Engl.	Englisch
ES	Effektstärke
FIR	Filter mit endlicher Impulsantwort
GSR	Galvanic skin response
H	Hypothese
HRV	Herzfrequenzvariabilität
ICD	International Statistical Classification of Diseases and Related Health Problems
IPQ	Igroup Presence Questionnaire
LSAS	Liebowitz Social Anxiety Scale
M	Mean
MINI	Mini-International Neuropsychiatric Interview
N3	Drei virtuelle Avatare in der VR-Umgebung
N6	Sechs virtuelle Avatare in der VR-Umgebung
N15	15 virtuelle Avatare in der VR-Umgebung
Okkulomotor.	Okkulomotorisch
RMSSD	Root Mean Square of Successive Differences
RR-Intervall	Zeitlicher Abstand zwischen den R-Zacken zweier Herzschläge im EKG
SCID	Structured Clinical Interview for DSM-IV Disorders

SD	Standardabweichung
SDNN	Standard Deviation of the NN Intervals
SE	Standarderror
SNRI	Selektive Serotonin-Noradrenalin-Wiederaufnahmehemmer
SPDQ	Social Phobia Diagnostic Questionnaire
SPSS	IBM SPSS Statistics 26
SSA	Situativer sozialer Angstfragebogen
SSQ	Simulator Sickness Questionnaire
SSRI	Selektive Serotonin-Wiederaufnahmehemmer
STADI-S	State-Trait-Angst-Depression-Inventar State (Zustand)
STAI-T	State-Trait-Angst-Depression-Inventar Trait (Eigenschaft)
TSST	Trier Social Stress Test
TSST-G	Trier Social Stress Test für Gruppen
Unabhäng.	Unabhängig/e

1 Einleitung

1.1 Soziale Phobie

1.1.1 Epidemiologie

Spezifische Phobien, soziale Phobie, generalisierte Angststörung, Agoraphobie und Panikstörungen zählen zu dem Krankheitsspektrum der Angststörungen. Dabei weisen Angststörungen mit 15,3% die höchste 12-Monats-Prävalenz unter den psychischen Störungen auf. Gefolgt werden die Angststörungen von unipolaren Depressionen mit 7,7% und Alkohol- bzw. Medikamentenabhängigkeit mit 5,7% [1].

Unter den Angststörungen ist die Häufigste die spezifische Phobie mit einer 12-Monats-Prävalenz von 10,3% bei 18-79-Jährigen, gefolgt von Agoraphobie mit 4% und sozialer Phobie mit 2,7% [1]. Somit handelt es sich bei sozialer Phobie um die dritthäufigste Angststörung, wobei Frauen deutlich stärker betroffen sind als Männer. Im Vergleich zu Männern, 1,9%, beträgt die 12-Monats-Prävalenz für soziale Phobien bei Frauen 3,6% [1]. Insgesamt sind 1,7 Millionen Menschen in Deutschland betroffen [2]. Zudem beeinflusst das Krankheitsbild den Alltag und die Arbeitsproduktivität: Laut dem Bundes-Gesundheitssurvey aus dem Jahr 1998 können Patienten/innen mit sozialer Phobie im Durchschnitt drei Mal häufiger ihren Alltagsaktivitäten nicht nachgehen als Personen ohne psychische Störungen. Gegenüber berufstätigen Frauen und Männern ohne psychische Störungen, deren monatliche Ausfalltage durchschnittlich 0,6 bzw. 0,9 betragen, weisen Betroffene mit 1,9 respektive 6,9 Ausfalltagen eine deutliche Einschränkung in ihrer Arbeitsproduktivität auf [3].

1.1.2 Diagnostik und Differenzialdiagnosen

Die Diagnosekriterien für die soziale Phobie sind im ICD-10 im Kapitel V unter der Gruppe phobischer Störungen vorzufinden [4]. Dabei muss für die Diagnose eine deutliche Furcht oder Vermeidung von Situationen, in denen man im Zentrum der Aufmerksamkeit steht, vorhanden sein. Es müssen mindestens zwei Angstsymptome, wie z.B. Schweißausbrüche, Herzrasen, Engegefühl in der Brust, in den gefürchteten Situationen auftreten. Darüber hinaus muss mindestens eines der folgenden Symptome auftreten: Erröten oder Zittern, Angst zu vomieren, Miktions- oder Defäkationsdrang oder die Angst davor. Die Angst oder das Vermeidungsverhalten stellen für den/die Betroffene/n eine deutliche emotionale Belastung dar. Außerdem beschränken sich die oben genannten Symptome ausschließlich auf die gefürchteten Situationen oder bei Gedanken an diese. Als Ausschlussdiagnose gelten andere Erkrankung, die ursächlich für die Symptome sein können, wie z.B. Wahn, Halluzinationen oder organische psychische Störungen.

Für die Praxis haben sich daher kurze Fragen zur Ersteinschätzung und Screening-Fragebögen bewährt. Ein bekanntes und international verwendetes strukturiertes Interview ist das

Mini-International Neuropsychiatric Interview, kurz MINI. Es wurde in Anlehnung an das Diagnostic and Statistical Manual of Mental Disorders (DSM) und die International Statistical Classification of Diseases and Related Health Problems (ICD) von amerikanischen und europäischen Psychiatern/innen und Klinikern/innen entwickelt und ermöglicht in etwa 15 Minuten eine Erfassung von Symptomen und die Diagnose von psychischen Erkrankungsbildern.

Bei der Diagnose müssen internistische, endokrinologische, neurologische und andere psychiatrische Diagnosen ausgeschlossen werden. Beispielsweise können Hypoglykämien mit Tremor, Ängsten und Schweißausbrüchen oder Myokardinfarkte mit retrosternalen Schmerzen und einem Engegefühl, einer Unruhe, Atemnot und Kaltschweißigkeit einhergehen. Als mögliche Differenzialdiagnosen aus dem Bereich phobischer Erkrankungen gelten die Panikstörung und Agoraphobie [5]. Im Gegensatz zu Panikstörungen sind die Panikattacken von Patienten/innen, die an einer sozialen Phobie leiden, weniger intensiv und gehen nicht mit Vernichtungsängsten einher. Darüber hinaus ist der Auslöser einer Panikattacke im Rahmen einer sozialen Phobie i.d.R. sehr spezifisch und bezieht sich auf die möglicherweise kritischen bzw. beurteilenden Beobachtungen Anderer. Bei der Agoraphobie wiederum handelt es sich um eine Angst vor größeren Menschenansammlungen und vor Reisen, d.h. das insbesondere Situationen in öffentlichen Verkehrsmitteln, Einkaufsmalls und/oder auf öffentlichen Plätzen Auslöser von Ängsten und Panikattacken sein können. Die Angst im Rahmen einer sozialen Phobie bezieht sich häufig auf kleinere Menschenmengen und ist nicht mit größeren anonymen Menschenansammlungen auf großen Plätzen assoziiert [5]. Als weitere nicht-phobische Differenzialdiagnose gilt die Depression, die Symptome wie depressive Verstimmung, Interessensverlust, Freudlosigkeit und Antriebsmangel aufweisen kann. Die erwähnten Symptome können zum sozialen Rückzug führen und müssen von der sozialen Phobie abgegrenzt werden.

1.1.3 Verlaufsdagnostik

Für die Verlaufsbeurteilung der sozialen Phobie wird der Liebowitz Social Anxiety Scale, LSAS, genutzt [6]. Der Fragebogen beinhaltet insgesamt 24 Items, in denen diverse Situationen beschrieben werden. Je Item wird die Angst und das Vermeidungsverhalten mit einer Likert-Skala von 0-3 erfragt und aus den Antworten ein Summenwert gebildet. In der Literatur gibt es Hinweise, dass mit dem LSAS auch Patienten/innen mit einer sozialen Phobie identifizierbar sind [7].

1.1.4 Therapie

Zur Therapie von sozialer Phobie stehen mehrere psycho- und pharmatherapeutische Verfahren zur Verfügung. Im Rahmen von Psychotherapien wird die kognitive Verhaltenstherapie empfohlen (Empfehlungskategorie A) [8]. Alternativ werden psychoanalytische Verfahren bei Unwirksamkeit bzw. Nichtverfügbarkeit von kognitiven Verhaltenstherapien oder

bei Patientenpräferenz angewandt. Die Empfehlungskategorie beträgt hierbei B [8]. Im Rahmen der kognitiven Verhaltenstherapie soll der/die Betroffene dysfunktionale Vorstellungen, Gedanken und Erwartungen identifizieren, um anschließend neue, sich positiv auf die Erkrankung auswirkende Verhaltensmuster zu erlernen oder um belastende Verhaltensmuster zu verändern. Die Wirksamkeit von kognitiven Verhaltenstherapien wurde in einer Metaanalyse mit 24 randomisierten Studien mit einer Effektstärke von 0,71 [9] angegeben.

Für die medikamentöse Therapie von sozialer Phobie werden zwei Medikamentenklassen empfohlen (Empfehlungskategorie A) [8]: Selektive Serotonin-Wiederaufnahmehemmer, SSRI, und selektive Serotonin-Noradrenalin-Wiederaufnahmehemmer, SNRI [8]. Metaanalysen konnten zeigen, dass beide Medikamentenklassen bei sozialer Phobie effektiv waren [10, 11]. In Deutschland zugelassen sind aus der Klasse der SSRIs Escitalopram, Paroxetin und Sertralin und aus der Klasse der SNRIs Venlafaxin. Abgesehen von den oben erwähnten Medikamentenklassen wird ein drittes Medikament, Moclobemid, aus der Klasse der reversiblen Monoaminoxidase-A Inhibitoren, RIMA, zur Therapie von sozialer Phobie genutzt. Aufgrund von inkonsistenten Studienergebnisse bezüglich Moclobemid soll es nur bei Unwirksamkeit und/oder Unverträglichkeit von Medikamenten aus der Kategorie A genutzt werden (klinischer Konsenspunkt) [8].

1.2 Expositionstherapie

Der Begriff „Exposition“ kommt aus dem Lateinischen, *expositio*, und bedeutet Aussetzung. Im Rahmen von Expositionstherapien werden Betroffene mit Situationen konfrontiert, die mit Ängsten assoziiert sind und vermieden werden. Es kommen je nach zugrundeliegender Phobie unterschiedliche Expositionsarten und -modelle zum Einsatz. Beispielsweise können Spinnenarten bei Arachnophobie genutzt werden, ein öffentlicher Platz bei Agoraphobie besucht werden oder Bewerbungsgespräche bei sozialer Phobie durchgeführt werden. Ziel der Exposition ist es, bewusst mithilfe von auslösenden externen und internen Faktoren die unangenehmen, vermiedenen Verhaltensweisen bzw. Reaktionen zu induzieren, identifizieren und verändern. Die Wirkung von Expositionen geht vermutlich auf eine Habituation und einen Umbau von kognitiven Schemata zurück. Durch repetitive Expositionen gewöhnt sich der/die Betroffene an die induzierende Situation. Zudem wird angenommen, dass assoziative Verknüpfungen zwischen Reiz und Reizreaktion gelockert werden. Die Expositionen können in der Realität, als Vorstellung oder in virtueller Realität durchgeführt werden und sollten, um den Effekt der Habituation zu maximieren, über einen längeren Zeitraum wiederholt werden [12].

1.2.1 Wirksamkeit von Expositionstherapien bei sozialer Phobie

Mehrere Metaanalysen haben die Wirksamkeit von Expositionstherapien nachgewiesen. In der Metaanalyse von Gould et al. wurden 24 Studien und 1079 Probanden/innen inkludiert.

Die alleinige Anwendung der Expositionstherapie hatte mit 0,89 die größte Effektstärke, gefolgt von der Kombination der Expositions- und kognitiver Verhaltenstherapie (ES: 0,8) [13]. Eine weitere Metaanalyse von Fedoroff et al. zeigte bei sieben eingeschlossenen Studien eine Effektstärke von 1,1 für die Expositionstherapie und bei 21 eingeschlossenen Studien eine Effektstärke von 0,84 für die Kombination aus Expositions- und kognitiver Verhaltenstherapie [14]. Eine Metaanalyse, die ausschließlich randomisierte Studien inkludierte, zeigte für die Expositionstherapie eine Effektstärke von 0,8 und für die kognitive Expositionstherapie eine Effektstärke von 0,8 [9].

Insgesamt ist die Studienlage bezüglich Expositionstherapien bei sozialer Phobie relativ eindeutig. Die Wirksamkeit wurde in diversen Studien und Metaanalysen aufgezeigt und ist integraler Bestandteil der Therapie von sozialer Phobie. Klassischerweise werden in der Klinik Expositionstherapien in vivo durchgeführt.

1.2.2 Schwächen der in vivo-Expositionstherapie

Trotz des Nachweises der Wirksamkeit von Expositionstherapien weisen in vivo-Expositionstherapien Schwächen auf:

1. Die sozialen Situationen lassen sich schwierig graduieren und in aufbauende Einheiten organisieren [12]. Soziale Situationen sind in der Regel von sehr vielen Faktoren abhängig und sowohl der Beginn als auch Ausgang einer sozialen Situation sind schwierig abzuschätzen.
2. Um soziale Situationen zu erzeugen, ist ein großer Aufwand notwendig, da die Betroffenen auf unterschiedliche Reize, wie z.B. öffentliche Reden, Restaurantbesuche, öffentliche Telefonate, mit als unangenehm empfundenen körperlichen Symptomen reagieren. Eine Nachstellung dieser unterschiedlichen Situation kann unter Umständen sehr aufwendig sein.
3. Für die Nachstellung von sozialen Situationen und für das Monitoring des Patienten/der Patientin sind viele Angestellte notwendig. Bei einer großen Fallzahl kann dies sich zu einem relevanten Kostenfaktor entwickeln, da sich in vivo-Expositionen nicht gut skalieren lassen. Darüber hinaus ist der Zeitaufwand für längere soziale Kontaktsituationen nicht zu vernachlässigen.

1.3 Virtuelle Realität Expositionstherapie (VRET)

Aufgrund der im Abschnitt 1.2.2. erwähnten Schwächen bzw. Schwierigkeiten der in vivo-Expositionstherapie wurde die Entwicklung von Alternativen in Form von virtueller Realitätsexposition vorangetrieben. Durch die virtuelle Realität ist eine kostengünstige, standardisierte und gut skalierbare Möglichkeit zur Durchführung von Expositionstherapien realisierbar:

1. Die Entwicklung von standardisierten Programmen, die aufeinander aufbauend gestaltet sind, ist möglich.

2. Die Darstellung von sozialen Situationen ist mithilfe von Programmen vergleichsweise einfach möglich. Abgesehen von einem initialen Programmieraufwand ist das Programm autark und mit wenig Personal lauffähig.
3. VR-Lösungen sind gut skalierbar und inzwischen preisgünstig zu erwerben. Im Jahr 1990 hat Tart von einem „neuen Millennium“ gesprochen, in der ein signifikanter Anteil der westlichen Welt die Möglichkeit haben wird, freiwillig in die virtuelle Realität einzusteigen [15]. Tart erwähnte mögliche Optionen für die Diagnostik und die Therapie im Zusammenhang von Psychotherapien. Die damals hohen Kosten von etwa 430000\$ sind inzwischen auf ein Bruchteil gefallen und VR-fähige Systeme inklusive VR-Brille sind für etwa 2000-3000€ zu erwerben.
4. Abgesehen von Kostenersparnissen hat sich die Technologie in den letzten Jahren rasant entwickelt, sodass sich die Darstellung der virtuellen Realität stets verbessert hat.
5. Die Akzeptanz für virtuelle Realitätsexpositionstherapie, VRET, ist höher als für in vivo-Expositionen. Die Teilnehmer/innen der Studie bevorzugten auf einer Skala von 1-7 mit einem Mittelwert von 6,08 die virtuelle Exposition. Der Mittelwert für eine in vivo-Exposition betrug 3,9 und unterschied sich signifikant von dem Wert für die virtuelle Exposition. Zudem war die Ablehnungsquote mit 3% für die virtuelle Exposition im Vergleich zu 27% für die in vivo-Exposition deutlich niedriger [16].
6. Das Monitoring ist für den/die Therapeuten/Therapeutin auf einer einfachen Art und Weise möglich. Der/die Betreuende kann alle wichtigen Vitalfunktionen am Monitor beobachten und bei Notwendigkeit intervenieren.

Die Vorteile der virtuellen Expositionstherapie in Bezug auf die Therapie von sozialer Phobie ist Bestandteil aktueller Forschung. Ein systemisches Review bezüglich der Wirksamkeit von VRET betonte die Unterschiede innerhalb der Phobien und lehnte eine Generalisation der Effekte auf alle Phobien ab. Insbesondere bei zwei spezifischen Phobien, der Aviophobie und der Akrophobie, sei die Wirksamkeit von VRET durch Studien erwiesen. Hinsichtlich der klinischen Relevanz für andere Phobien und der posttraumatischen Belastungsstörung wäre ein Fazit verfrüht [17]. Zu diesem Zeitpunkt, 2010, waren kontrollierte klinische Studien zum Einsatz von VRET bei sozialen Phobien im Vergleich zur Warteliste bzw. in vivo-Exposition eine Seltenheit. Inzwischen haben sich mehrere Forschungsgruppen mit dieser Thematik befasst und ihre Ergebnisse publiziert. Eine Metaanalyse von Carl et al. hat 2019 den Effekt der VRET auf Phobien, u.a. sozialer Phobie, untersucht [18]. Es wurden insgesamt acht Studien zu sozialer Phobie und Lampenfieber, engl. „performance anxiety“, in die Metaanalyse inkludiert. Von den acht Studien bezogen sich sechs auf die VRET bei sozialer Phobie und zwei auf die VRET bei Lampenfieber. Die Analyse stellte eine große Effektstärke, 0.97, für die VRET im Vergleich zu der Warteliste fest. Im Vergleich zu einer in vivo-Exposition gab die Analyse einen nicht-signifikanten Effekt zugunsten der VRET an (ES: 0.06).

1.4 Forschungsschwerpunkt dieser Arbeit

Die Forschungsarbeit beschäftigt sich mit der Induktion von Stress durch ein standardisiertes Verfahren mit einer variablen Anzahl von Avataren in der virtuellen Realität. Darüber hinaus sollen in dieser Studie die Reaktionen in der virtuellen Realität mit denen in der physischen Realität verglichen werden. Für die Induktion von Stress eignet sich der bekannte und validierte „Trier Social Stress Test“, kurz TSST, der an der Universität Trier entwickelt wurde [19, 20]. Bei diesem Test handelt es sich um ein standardisiertes Verfahren zur Induktion von „moderatem psychologischen Stress unter Laborbedingungen“ [19]. Der eigentliche Test dauert 10 Minuten und beinhaltet jeweils eine fünfminütige Rede und eine fünfminütige Rechenaufgabe vor einer dreiköpfigen Jury. Zuvor wird dem/r Probanden/in eine Vorbereitungszeit von fünf Minuten eingeräumt. Nach dem Test erfolgt eine Beobachtungsphase.

Die „Social-Impact-Theorie“ wurde von Bibb Latané im Jahr 1981 postuliert und beschreibt den Einfluss einer Gruppe auf ein Individuum in Abhängigkeit von unterschiedlichen Faktoren [21]. Folgende Faktoren spielen laut Latané hierbei eine Rolle: Die Stärke, die Unmittelbarkeit und die Anzahl der Gruppenmitglieder. Die drei Faktoren seien multiplikative Faktoren, die man mit der Formel $(I = f(S * I * N))^1$ darstellen könne. Bei der Stärke handle es sich um die Wichtigkeit und die Macht der Person/en, wobei die Stärke vom Alter, Status, sozioökonomischen Status, von vorausgehenden Beziehungen und zukünftigen Einflusstärken abhängt. Die Unmittelbarkeit beschreibe sowohl die zeitliche als auch örtliche Nähe zur Person und das Nichtvorhandensein von Barrieren. Mit der Anzahl der Gruppenmitglieder sei die Anzahl der Personen, die anwesend sind, gemeint. Zur Veranschaulichung verglich Latané die Faktoren mit einer Glühbirne, die auf eine Oberfläche strahlt. Dabei sei die Stärke mit der Wattzahl oder Intensität der Glühbirne vergleichbar, die Nähe der Glühbirne zur Oberfläche spiegle die Unmittelbarkeit wider und die Anzahl der aktiven Glühbirnen stelle die Anzahl der Gruppenmitglieder dar.

Ein weiteres Gesetz, das psychosoziale Gesetz, beschreibe den Zusammenhang zwischen der Anzahl der Personen und den Einfluss auf die betroffene Person [21]. Das Gesetz sagt aus, dass mit der zunehmenden Anzahl an Personen der Einfluss zwar zunehme, jedoch in einem geringeren Maße. D.h., dass der Unterschied zwischen keiner und einer Person größer ist als der Unterschied zwischen 99 und 100 Personen.

Im Rahmen dieser Forschungsarbeit wird der Frage nachgegangen, ob dieser Zusammenhang zwischen Personenanzahl und Einfluss auch auf die Exposition in der virtuellen Realität zutrifft.

¹ I: Einfluss, engl. impact, S: Stärke, engl. strength, I: Unmittelbarkeit, engl. immediacy, N:Anzahl, engl. Numbers.

1.5 Hypothesen

In Anbetracht der zuvor angeführten Forschungsziele wurden folgende vier Hypothesen formuliert:

- Hypothese 1:** Die VR-Exposition durch eine herausfordernde Situation (Bewerbungsgespräch) induziert Angst (H1).
- Hypothese 2:** Die Anzahl der virtuellen Avatare in dieser Situation beeinflusst die Stärke der erzeugten Angst (H2).
- Hypothese 3:** Die in vivo-Exposition in einer herausfordernden Situation induziert Angst (H3).
- Hypothese 4:** Die durch die VR-Exposition erzeugte Reaktion ist stärker als die der in vivo-Exposition (H4).

2 Material und Methoden

2.1 Aufbau des Experiments

Das Experiment bestand aus zwei Teilen: Einer in vivo-Testung und einer Testung in der virtuellen Umgebung. Das Ziel dieser Studie war es die beiden Testungsumgebungen hinsichtlich der Induktion von Angst zu vergleichen. Die beiden Testungen wurden zur identischen Uhrzeit an unterschiedlichen Wochentagen durchgeführt, da die Cortisolexpression einer zirkadianen Rhythmik unterliegt [22]. Jede/r Proband/in führte beide Testungen durch und somit lässt sich unsere Studie als eine prospektive Cross-over-Studie kategorisieren.

Bei der in vivo-Testung handelt es sich um eine leicht modifizierte Form des Trier Social Stress Test (TSST) [19]. Der Unterschied zum ursprünglichen TSST bestand darin, dass die Probanden/innen die Testung im Sitzen durchführten. Diese Modifikation wurde zum Zwecke der Vergleichbarkeit mit der Testung in der virtuellen Umgebung implementiert, da die Testung in der virtuellen Umgebung im Sitzen durchgeführt wurde.

Zudem wurde für die Testung in der virtuellen Umgebung eine variable Anzahl an Personen, virtuelle Avatare, in die Jury eingeführt. Die Anzahl betrug hierbei drei (N3), sechs (N6) bzw. 15 (N15). Diese Modifikation des TSSTs diente der Beurteilung des Einflusses der variablen Anzahl an virtuellen Avataren auf die Induktion von Angst. Die drei Testungen mit einer variablen Anzahl an virtuellen Avataren erfolgten konsekutiv am selben Tag.

Die Probanden/innen durchliefen folgende Abschnitte während der Testung: Ausgangssituation (Baseline), Vorbereitungsphase, TSST, Ruhephase. Vor dem Beginn der eigentlichen Testung erhielten die Probanden/innen Zeit, um sich zu entspannen und den Einwilligungsbogen durchzulesen und auszufüllen. Anschließend wurden den Probanden/innen – bei Zustimmung zur Teilnahme an der Studie – die erste Cortisolspichelprobe in Form einer Salivette abgenommen und die notwendige Verkabelung für die Durchführung des Experiments angelegt. Dabei handelte es sich um zwei Elektroden für die Messung des Hautleitwiderstandes, eine Elektrode für die Messung des Pulses und fünf Elektroden für die Ableitung des EKGs. In der Vorbereitungsphase hatten die Probanden/innen fünf Minuten Zeit sich auf einen vorgegebenen Beruf vorzubereiten. Für die Vorbereitung erhielten die Probanden/innen von dem/der Experimentleiter/in einen Zettel und Stift und durften sich für die anstehende Rede Notizen machen. Nach der Vorbereitungsphase erfolgte die eigentliche Testung in Form des zehnminütigen TSSTs. Der TSST bestand aus zwei gleichlangen, jeweils fünf minütigen Teilen: Einer freien Rede und der Lösung einer arithmetischen Aufgabe. Die freie Rede erfolgte vor einer Jury mit drei Personen, die diesen Testungsabschnitt mit der Aufforderung, die Rede zu beginnen, initiierte. Falls der/die Proband/in während der fünf Minuten stoppte, wurde er/sie von der Jury gebeten, fortzufahren, da die Zeit nicht abgelaufen war. Bei einem weiteren Stopp ist die Jury befugt gewesen, katalogisierte Fragen zu stellen. Danach erfolgte der zweite Abschnitt des TSSTs in Form einer arithmetischen Aufgabe.

Die Jury bat den/die Probanden/in von der Zahl 1022, 1333, 1555 oder 1777 die Zahl 13 zu subtrahieren, wobei die Zahl 1022 für die in vivo-Testung genutzt wurde. Die anderen erwähnten Zahlen wurden im Rahmen der Testung in der virtuellen Umgebung verwendet. Bei einem Fehler wurde der/die Proband/in gebeten, erneut von der Ursprungszahl zu beginnen. Am Ende der Testung erfolgte die Ruhephase. Diese betrug zwischen den Testungen in der virtuellen Umgebung zehn Minuten und am Ende der in vivo und VR-Testung 20 Minuten.

2.2 Probanden/innen

Im Rahmen der vorliegenden Arbeit wurden die Daten von 24 Probanden/innen auf der Station für Angst- und Zwangsstörungen der Klinik für Psychiatrie und Psychotherapie der Universitätsklinik Hamburg Eppendorf im Zeitraum vom Februar 2018 bis Februar 2019 erhoben.

Die Bewerbung für die Studie erfolgte per Online-Formular. Es bewarben sich 379 mögliche Probanden/innen, von denen 24 Prozent für die Studie geeignet waren. Folgende Kriterien wurden für die Auswahl der Probanden/innen gewählt: Die Probanden/innen sollten ein Alter zwischen 18 und 55 Jahren aufweisen, keine psychiatrischen Erkrankungen haben und keine Medikation zu sich nehmen, die das Bewusstsein oder die Stressreaktion beeinflussen könnte. Zudem wurden für die Durchführung des Experimentes ausreichende Deutschkenntnisse in Form von mindestens vier von fünf Punkten auf einer 5-Punkt Likert-Skala vorausgesetzt.

Das Screening der Probanden/innen erfolgte mittels folgender Fragebögen: Mini-International Neuropsychiatric Interview (MINI) [23] und Social Phobia Diagnostic Questionnaire (SPDQ) [24].

Es wurde eine gekürzte Form des MINI-Fragebogens mit Fokus auf die wesentlichen und akuten Erkrankungen verwendet. Hierbei berücksichtigte die gekürzte Form folgende Fragebereiche: Major Depression, Dysthymie, Panik/Agoraphobie, soziale Phobie, generalisierte Angststörung, Alkohol- und Drogenprobleme. Bejahungen bezüglich vergangener Lebensabschnitte wurden ausgeschlossen. Voraussetzung für die Teilnahme an der Studie war eine Verneinung der Fragen zu den oben genannten Erkrankungsbereichen.

Der SPDQ-Fragebogen diente uns zur groben Einschätzung des Grades der Neigung zu sozialer Phobie bei den Probanden/innen. Der SPDQ-Fragebogen wurde zur Beurteilung von sozialer Phobie übereinstimmend mit den Kriterien des Diagnostic and Statistical Manual of Mental Disorders-IV (DSM-IV) entwickelt. Der Fragebogen beinhaltet 25 Fragen, die sowohl spezifisch als auch unspezifisch verschiedene Bereiche, in denen Angst bzw. Vermeidungsverhalten auftreten können, abdeckt. Die Fragen werden zum Teil mit „Ja“ und „Nein“ und zum anderen Teil in Abhängigkeit der Schwere der Symptomatik mithilfe einer Likert-Skala beantwortet. Im Anschluss wird aus vorliegenden Antworten ein Summenwert

mit unterschiedlicher Gewichtung der Einzelfragen gebildet. Als Grenzwert diente uns ein Summenwert von ≥ 1 und stellte somit die Bedingung für die Teilnahme an dieser Studie.

Vor dem Beginn der Experimente wurden zudem folgende Fragebögen von den Probanden/innen erhoben: Liebowitz Social Anxiety Scale (LSAS) [6] und State-Trait-Angst-Depression Inventar-Trait (STADI-T) [25].

Die LSAS-Skala wurde von Liebowitz im Jahr 1987 [6] entwickelt und beinhaltet 24 Situationen. Von jenen Situationen sind 13 leistungsbezogen (bspw. öffentliche Rede halten) und 11 beziehen sich auf soziale Interaktionen (bspw. mit einer fremden Person sprechen). Die Situationen werden sowohl mit dem Grad der ausgelösten Angst als auch mit dem Vermeidungsverhalten bewertet. Zur Auswertung wird der kumulierte Gesamtwert aus den Angaben bezüglich der Angst bzw. dem Vermeidungsverhalten gebildet.

Der STADI-T ist ein Fragebogen zur Erfassung und Bewertung der Eigenschaft (Trait) Angst bzw. Depression [25]. Der Fragebogen besteht aus 20 Fragen, die in vier Subbereiche eingeteilt werden können: Als Komponenten der Angst werden Aufgeregtheit und Besorgnis und als Komponenten der Depression Euthymie und Dysthymie erfasst. Je Subbereich werden fünf Fragen gestellt, die mit „fast nie“ bis „fast immer“ beantwortet werden können. Die Antworten werden mit den Zahlen eins bis vier kodiert. Daraus errechnet sich der jeweilige Wert für den jeweiligen Subbereich und durch die Addition der Subbereiche wird der Wert für die Angst bzw. Depression gebildet. Da es sich bei der Euthymie um einen positiven Affekt handelt, wird hierbei der invertierte Wert verwendet. Zudem wird der Wert an das Alter und Geschlecht anhand von T-Wert-Tabellen angepasst.

Die teilnehmenden Probanden/innen waren überwiegend weiblich (75%), hatten ein durchschnittliches Alter von 28,17 Jahren ($SD = 7,54$), wiesen ausreichende Deutschkenntnisse vor (Mittel: 4,92, $SD = 0,28$), hatten keine psychiatrischen Vorerkrankung, waren Nicht-Raucher und litten nicht unter Gleichgewichtsstörungen. Zudem hatten 15 Probanden/innen (64%) vor der Teilnahme am Experiment eine VR-Brille genutzt. Bezüglich der Screening-Fragebögen wiesen die Probanden/innen keine psychiatrische Erkrankung auf. Der SPDQ-Durchschnittswert betrug 4,86 ($SD = 3,25$). Der durchschnittliche Gesamtwert der LSAS-Skala betrug 31,21 ($SD = 19,41$) und dies würde einer milden, klinisch insignifikanten, sozialen Phobie entsprechen [26]. Der STADI-T T-Wert für die Angst und für die Depression betrug 49,35 ($SD = 10,65$) respektive 43,13 ($SD = 12,5$). Beide T-Werte befanden sich innerhalb der Spanne von 40 bis 60 und sind somit als durchschnittlich zu werten.

Tabelle 1: Charakteristika der Probanden/innen

	Gesamt	Männer	Frauen
Anzahl der Teilnehmer (%)	24 (100%)	6 (25%)	18 (75%)
Durchschnittsalter in Jahren (SD)	28,2 (7,5)	31,8 (9,5)	26,9 (6,6)
SPDQ-Score (SD)	4,9 (3,3)	4,8 (2,1)	4,9 (3,6)
LSAS (SD)	31,2 (19,4)	45 (32,4)	27,4 (13)
STADI-T Angst (SD)	49,35 (10,65)	47,4 (14)	49,9 (9,9)
Erfahrung mit VR-Brillen (%)	15 (62,5)	5 (83,3)	10 (55,6)

2.3 Materialien

2.3.1 Technisches Equipment

Für die Ausführung der Experimente in virtueller Umgebung wurden ein Rechner, eine VR-Brille, ein zusätzlicher Monitor und ein Notebook benötigt.

Bei dem Rechner handelte es sich um einen zusammengestellten Rechner mit einer Intel CPU (Core i7- 7700K), 32 GB DDR4 Arbeitsspeicher und einer MSI GeForce GTX1080 Ti 11GB Founder Edition. Das VR-Programm für die Testung in virtueller Umgebung wurde mithilfe von der „Unity 3D game engine“ und der Programmiersprache C# erstellt. Die Jurymitglieder der virtuellen Umgebung wurden mit Adobe Fuse CC, Mixamo, und Blender erstellt. Der in Textform vorliegende Inhalt des Gespräches wurde mit einem online Text-in-Sprache-Programm [27] in Audio konvertiert. Um den Ablauf der VR-Testung zu monitoren, wurde ein zusätzlicher Monitor für den/die Experimentleiter/in zur Verfügung gestellt. Des Weiteren kam für die Darstellung der virtuellen Realität die VR-Brille *Vive* der Marke HTC zum Einsatz. Für das Ausfüllen der Fragebögen wurde ein Notebook der Marke Asus mit folgenden Komponenten verwendet: Intel CPU (i5), 8GB RAM, Nvidia Geforce GTX1050 mit einem 15,6" großen Bildschirm. Dieses Notebook diente uns im Rahmen der in vivo-Testung ebenso für das Aufzeichnen des Hautleitwiderstandes und des Pulses.

2.3.2 Cortisolspichelproben

Für die Bestimmung des Speichelcortisols wurden Salivetten der Firma Sarstedt, die zur Steigerung der Speichelproduktion mit Zitronensäure versetzt wurden, verwendet. Die Salivetten wurden durch unser internes Labor bei 4000 Umdrehungen pro Minute (UpM) und 4°C zentrifugiert. Das Zentrifugat wurde bei -80°C in der Tiefkühltruhe eingelagert. Bis zur end-

gültigen Bestimmung des Cortisols vergingen maximal 3 Wochen. Die Auswertung der Cortisolkits der Firma DRG (RIA-0189) erfolgte am RIA-Star der Firma Packard (Perkin Elmer). Als Qualitätskontrollen dienten hierbei Lypocheck Immonoassays Plus (Firma Bio Rad).

Die Abnahme des Speichelcortisols erfolgte zu festgelegten Zeitpunkten. Kurz nach der Ankunft der Probanden/innen erfolgte die erste Bestimmung als Ausgangswert. Dazu wurde der/die Proband/in angeleitet, die Watterolle für eine Minute zu zerkaugen und nach Ablauf der Minute wieder in die Salivette hineinzuführen. Es erfolgten mehrere Cortisolbestimmungen jeweils vor und nach der Vorbereitungsphase bzw. TSST-Testung. Um die Dynamik und den Verlauf des Cortisolwertes auch nach Beendigung der Testung zu beobachten, erfolgte eine 20-minütige Nachbeobachtungsphase mit zwei weiteren Messungen. Insgesamt wurden somit sechs Cortisolwerte bei der *in vivo*-Testung und zwölf Werte bei der VR-Testung bestimmt. Diese Diskrepanz ist durch die drei konsekutiven Testungen in der VR-Testung bedingt.

2.3.3 Hautleitwiderstand (GSR)

Der Hautleitwiderstand wurde mithilfe des „Neulog GSR logger sensor“ während der Testung aufgezeichnet. Dabei wurden die zwei Elektroden an der nicht dominanten Hand am Zeige- und Ringfinger via Klettverschlüsse fixiert. Die Samplingrate der Aufzeichnung betrug 60 Hz.

2.3.4 EKG

Für die Bestimmung des EKGs wurde das Modell Schiller Medilog® AR12 verwendet. Vor dem Beginn der Testungen wurden die Elektroden des Gerätes durch den medizinisch geschulten Experimentleiter/in angebracht und auf Fehlerfreiheit hin überprüft. Bei fehlerfreien, d.h. artefaktarmen, EKG-Ableitungen wurde die Aufzeichnung initiiert. Die Aufzeichnung der EKG-Ableitungen erfolgte bis zum Ende des Experimentes und entsprechende Zeitmarker, die die festgelegten Abschnitte des Testungsvorgangs kennzeichneten, wurden schriftlich fixiert.

2.3.5 Fragebögen

Die Erfassung und Beurteilung der Emotionsdynamik der Probanden/innen erfolgte mittels folgender Fragebögen: Self-Assessment Manikin (SAM), State-Trait-Angst-Depression-Inventar State und Trait (STADI-S und -T), State Social Anxiety (SSA) und ein graphisch numerischer Selbstbewertungsfragebogen zur Angst. Zudem wurden zur Einschätzung der virtuellen Umgebung der Co- and Social-Presence Questionnaire und der Igroup Presence Questionnaire (IPQ) verwendet. Für die Beurteilung der mit der virtuellen Umgebung möglicherweise einhergehenden Beschwerden, wie z.B. Übelkeit, wurde der Simulator Sickness Questionnaire (SSQ) genutzt.

2.3.5.1 Self-Assessment Manikin (SAM)

Der Self-Assessment Manikin, kurz SAM, ist ein non-verbaler Selbstfragebogen zur Messung der Dimensionen Valenz, Arousal und Dominanz [28]. Die Messung erfolgte mithilfe von Piktogrammen, die in Form eines Männchens verschiedene Gemütslagen in Bilderreihen darstellten: Bei der Valenz von glücklich bis unglücklich, bei dem Arousal von aufgeregt bis ruhig und bei der Dominanz von dominiert bis dominierend. Die Antworten wurden in Zahlenwerte umkodiert und anschließend die Summe gebildet. In unserer Studie wurde der Fragebogen von den Probanden/innen vor dem eigentlichen Testungsbeginn, nach der Vorbereitungsphase und nach der Testung in digitaler Form ausgefüllt.

2.3.5.2 State-Trait-Angst-Depressions-Inventar State (STADI-S)

Beim STADI-S handelt es sich um einen im Jahr 2013 veröffentlichten Fragebogen zur Beurteilung der Angst und Depression als Zustand (State) [25]. Der Fragebogen beinhaltet 20 Fragen und erfasst den aktuellen Zustand der Angst und Depression der befragten Person. Die 20 Fragen lassen sich mit jeweils 5 Fragen in vier Kategorien einordnen: Aufgeregtheit, Besorgnis, Euthymie und Dysthymie. Die ersten beiden Kategorien dienen zur Erfassung der Angst und die letzten beiden zur Erfassung der Depression. Analog zum STADI-T wurden die Fragen mit „fast nie“ bis „fast immer“ beantwortet, in Zahlenwerte umkodiert und anschließend die Summe für die jeweilige Kategorie gebildet. Aus den Summenwerten wurde dann anhand des Manuals der T-Wert bestimmt [25]. In unserer Studie wurde der Fragebogen von den Probanden/innen vor dem eigentlichen Testungsbeginn, nach der Vorbereitungsphase und nach der Testung in digitaler Form ausgefüllt.

2.3.5.3 Situativer sozialer Angstfragebogen (SSA)

Der SSA wurde zur genauen Beurteilung der Angst für die Studie von unserer Arbeitsgruppe entwickelt. Der Fragebogen besteht aus acht Items, wobei zwei Items 16 bzw. sieben Unterfragen beinhalten. Insgesamt setzt sich der Fragebogen somit aus 29 Fragen zusammen. Zudem lassen sich die einzelnen Items in drei Kategorien einordnen: Fragen zur Rede, zu den Symptomen und zu den Befürchtungen. Die Fragen wurden in Form einer Likert-Skala mit „gar nicht“ bis „sehr stark“ beantwortet. Die Antworten wurden dann in Zahlenwerte umkodiert und der Summenwert und der Wert für die jeweilige Kategorie gebildet, wobei der Maximalwert 116 beträgt. Der Fragebogen wurde von den Probanden/innen nach jeder Testung ausgefüllt.

2.3.5.4 Graphisch numerischer Fragebogen

Der graphisch numerische Fragebogen wurde den Probanden/innen nach jeder Testung in Papierform überreicht. Der Fragebogen besteht aus einem X-Y-Diagramm: Auf der X-Achse ist die Zeit in Minuten in Bezug auf die TSST-Testung aufgetragen und auf der Y-Achse befindet sich eine Skala von 0-10 für die Einschätzung der empfundenen Angst. Für jede volle Minute trug der/die Proband/in einen entsprechenden Angstwert ein.

2.3.5.5 Co- and Social-Presence Fragebogen

Der Co- and Social-Presence Questionnaire wurde zur Beurteilung der Ko-Präsenz und sozialen Präsenz für öffentliche Reden in virtueller Realität entwickelt [29]. Der Fragebogen beinhaltet vier Subkategorien: Die Reaktion des/r Redners/in auf die virtuellen Avatare, die wahrgenommene Reaktion der virtuellen Avatare durch den/die Redner/in, der Eindruck von Interaktionsmöglichkeiten und die Ko-Präsenz von anderen Personen. Jede Subkategorie beinhaltet vier Unterfragen, die in Form einer Likert-Skala von „stimme überhaupt nicht zu“ bis „stimme voll zu“ beantwortet wurden. Die Antworten wurden in Zahlenwerte umkodiert und die jeweilige Summe der Kategorie gebildet. Der Fragebogen wurde nach jeder VR-Testung von den Probanden/innen digital ausgefüllt.

2.3.5.6 Igroup Presence Questionnaire (IPQ)

Der Igroup Presence Questionnaire, IPQ, ist eine Skala zur Messung des Präsenzerlebens in der virtuellen Umgebung [30]. Der Fragebogen beinhaltet 15 Fragen, die in drei Subskalen eingeordnet werden: Räumliche Präsenz, Involviertheit und Realitätsurteil. In Abhängigkeit zur Subskala sind es für die räumliche Präsenz fünf, für die Involviertheit und für das Realitätsurteil jeweils vier Fragen. Zudem ist die Einschätzung zur Aussage „In der computererzeugten Welt hatte ich den Eindruck, dort gewesen zu sein...“ zur Beurteilung des allgemeinen Eindrucks der virtuellen Realität inkludiert. Die Fragen wurden in Form einer Likert-Skala von „überhaupt nicht“ bis zu „sehr stark“ beantwortet und in Zahlenwerte umkodiert. Daraus erfolgte die Summenwertbildung für die jeweilige Subskala. Dabei wurden die dritte, neunte und elfte Frage gemäß des Auswertungsschemas invertiert [30]. Der Igroup Presence Questionnaire wurde nach jeder VR-Sitzung einmalig in digitaler Form ausgehändigt.

2.3.5.7 Simulator Sickness Questionnaire (SSQ)

Der Simulator Sickness Questionnaire, SSQ, diente uns zur Einschätzung der mit der virtuellen Simulation möglicherweise einhergehenden Symptome. Der Fragebogen wurde auf Basis des Pensacola Motion Sickness Questionnaires [31] entwickelt und beinhaltet 16 Fragen zu diverseren Symptomen wie z.B. Kopfschmerzen, Übelkeit und Überanstrengung der Augen. Die Symptome werden in vier Kategorien zusammengefasst: 1. Übelkeit, 2. Okulomotorische Symptome, 3. Desorientierung und 4. Gesamtsumme. Die Fragen wurden mithilfe einer 4-Punkte Likert-Skala, deren Antwortmöglichkeiten von „keine“ bis „stark“ reichen, beantwortet. Die entsprechenden Antworten wurden von uns in Zahlenwerte umkodiert. Die entsprechenden Punkte für die einzelnen Kategorien wurden anhand der Auswertungsanleitung von Kennedy et al. [32] gebildet. Der Fragebogen wurde vor und nach der VR-Testungssession in digitaler Form beantwortet.

2.4 Statistisches Vorgehen bei der Datenanalyse

2.4.1 Allgemeines Vorgehen bei der statistischen Auswertung

Für die Auswertung der Daten erfolgte zunächst eine graphische Überprüfung auf Normalverteilung mithilfe von Histogrammen mit Normalverteilungskurven. Anschließend fiel die Wahl auf einen entsprechenden Test auf Lageunterschiede. Es wurden hierbei folgende Tests verwendet: Friedman Test, einfaktorielle Varianzanalyse mit Messwiederholung, Mann-Whitney Test, Wilcoxon Test, den gepaarten und ungepaarten t-Test und den Kruskal-Wallis Test. Für den paarweisen Vergleich wurde, wenn nicht anders angegeben, die Bonferroni-Korrektur verwendet. Das Signifikanzniveau betrug global $p=0,05$.

Eine besondere Rolle spielte die Analyse der mit der N3-Kondition beginnenden Gruppe, da sie von den Bedingungen am ehesten der in vivo-Testung entspricht. Demgemäß wurden zusätzliche Analysen explizit für diese Gruppe ausgeführt und ein Vergleich mit der in vivo-Testung beginnenden Gruppe durchgeführt.

2.4.2 Grafische Darstellung der Ergebnisse

Die Daten und die Ergebnisse der Testungen wurden mithilfe von MatLab 2020a visualisiert. Die Darstellung der Daten erfolgte mittels Boxplots und Liniendiagrammen mit einfachen Standardfehlern. Die Boxplot-Diagramme bestehen aus Rechtecken, die als Box bezeichnet werden, mit einer querverlaufenden etwas breiteren Linie. Diese Linie stellt den Median dar. Jeweils 50% der Werte liegen ober- bzw. unterhalb dieses Medians. Das untere und obere Ende der Box stellen das untere und obere Quartil dar. 25% der Daten befinden sich unterhalb oder sind gleich dem unteren Quartil. Beim oberen Quartil befinden sich somit analog dazu 75% der Werte unterhalb oder auf dem Niveau des Quartils und 25% der Werte oberhalb. Der Abstand zwischen dem unteren und oberen Quartil wird als Interquartilabstand bezeichnet. 50% der Daten befinden sich in diesem Bereich. Von der Box selbst gehen zwei senkrechte Linien hervor, die mit einer querverlaufenden Linie abgeschlossen werden. Diese Antennen, auch Whisker genannt, stellen das Minimum und Maximum dar. Im Fall des Minimums wurde vom unteren Quartil das 1.5-fache des Interquartilabstands subtrahiert. Hinsichtlich des Maximums wurde das 1.5-fache des Interquartilabstands zum oberen Quartil addiert.

Bezüglich der Signifikanzdarstellung in Form von Hilfszeichen wurde folgende gängige Definition verwendet:

- * $P \leq 0,05$
- ** $P \leq 0,01$
- *** $P \leq 0,001$
- **** $P \leq 0,0001$

2.4.3 Cortisol

Die Rohdaten der Cortisolspichelproben lagen in Form einer csv-Datei vor, die anschließend in IBM SPSS Statistics 26 (SPSS) importiert wurden. Zunächst erfolgte eine Prüfung der Werte nach Ausreißern mit dem Grubbs-Test in Kombination mit einer optischen Überprüfung der konsekutiven Cortisolwerte der jeweiligen Testperson. Dabei zeigten sich drei Ausreißer: Ausgangswert (ID: 20, 10 ng/ml) und zwei Werte für die Ruhephase (ID:21, 24,2 und 19,7 ng/ml). Die Daten der Cortisolmessungen waren nicht normalverteilt. Für die Überprüfung der Signifikanz zwischen den einzelnen Abschnitten der in vivo- respektive VR-Testung wurde der Friedman Test genutzt. Darüber hinaus wurden der Friedman Test und der Kruskal-Wallis Test zum Vergleich der VR-Konditionen untereinander genutzt. Für den Vergleich zwischen der N3-Konditionsgruppe und der in vivo-Gruppe wurden sowohl der Mann-Whitney Test als auch der Wilcoxon Test genutzt.

In der Literatur fanden sich Hinweise [33, 34], dass es Personen gibt, die auf Stresssituationen mit einem schwachen bzw. nicht vorhanden Anstieg des Speichelcortisols reagieren. Dementsprechend wurden im Rahmen unserer Studie eine Subgruppenanalyse und -darstellung durchgeführt. Die Einteilung in Responder und Non-Responder erfolgte anhand der Einzelverlaufskurven der Probanden/innen.

2.4.4 Hautleitwiderstand (GSR)

Die Daten des Hautleitwiderstandes lagen uns als CSV-Datei vor. Für die Berechnung erfolgte eine Übertragung der Daten in MatLab 2020a.

Mittels eines low-pass Filters mit endlicher Impulsantwort (FIR) wurde das hochfrequente Rauschen eliminiert [35]. Die „Cut-off“-Frequenz betrug 1 Hz und die Umsetzung erfolgte mit der integrierten Filterfunktion von MatLab:

$$fir1((n - 1), Wn, 'low')$$

Anschließend erfolgte eine Normalisierung der Daten [36] mithilfe folgender Formel:

$$GSR_{norm} = \frac{gSR_x - gSR_{min}}{gSR_{max} - gSR_{min}}$$

Dabei wurde von jedem GSR-Wert (gSR_x) der minimale GSR-Wert der gesamten Aufzeichnung subtrahiert und durch die Differenz zwischen dem maximalen und minimalen GSR-Wert dividiert.

Für die Berechnung der durchschnittlichen GSR-Werte für die einzelnen Abschnitte erfolgte eine Indexierung der entsprechenden GSR-Werte in den vorgegebenen Zeitabschnitten. Die indexierten Werte wurden durch die Anzahl der Werte im Abschnitt dividiert:

$$\begin{aligned} idx = \\ find(vi_time_0 > vi_GSR_markers(1,1) \& vi_time_0 \leq vi_GSR_markers(1,2)); \\ vi_results(1,1) = sum(vi_GSR_0(idx))/length(idx); \end{aligned}$$

Die errechneten Werte wurden anschließend für die weitere statistische Analyse nach SPSS importiert. Die vorliegenden Daten waren nicht normalverteilt. Die Untersuchung nach Lageunterschieden erfolgte für die Abschnitte der VR- bzw. in vivo-Testung mit dem Friedman Test. Der Friedman Test und der Kruskal-Wallis Test wurden zum Vergleich der VR-Konditionen untereinander genutzt. Beim Vergleich der N3-Konditionsgruppe mit der in vivo-Gruppe wurden der Mann-Whitney Test und der Wilcoxon Test verwendet.

2.4.5 Herzfrequenz

Die Daten des EKGs wurden zunächst in der Software Schiller medilog Darwin 2 Basic (Version 2.5.2) auf Artefakte hin überprüft. Die Überprüfung der Daten erfolgte manuell. Nach der initialen Artefaktbereinigung wurden die Daten zur Berechnung der Herzfrequenzen als MatLab-Datei exportiert. In MatLab 2020a erfolgte die Berechnung der Herzfrequenz in mehreren Schritten:

1. Zunächst wurde aus den entsprechenden absoluten Zeiten der Herzschläge das dazugehörige RR-Intervall berechnet.

$$RR_n = t_{n+1} - t_n$$

2. Aus den RR-Abständen erfolgte die Berechnung der durchschnittlichen Herzfrequenz wie folgt:

$$HF_n = 1/RR_n$$

3. Die Zeitmarker für die einzelnen Testungsabschnitte wurden in eine Matrix übertragen. Es erfolgte eine Indexierung der zugehörigen Herzfrequenzen innerhalb der vorgegebenen Zeitmarker. Die Werte wurden anschließend durch die Anzahl der Werte dividiert, um die durchschnittliche Herzfrequenz in diesem Zeitabschnitt zu erhalten. Die Berechnung der durchschnittlichen Herzfrequenz für einen Abschnitt geschah wie folgt:

$$\begin{aligned} idx = & \\ & find(vi_time_0 > vi_ecg_markers(1,1) \& vi_time_0 < \\ & = vi_ecg_markers(1,2)); \\ vi_results(1,1) = & sum(vi_HF_0(idx))/length(idx); \end{aligned}$$

Zur weiteren Analyse wurden die Daten nach SPSS übertragen. Die Daten der Herzfrequenzen waren, bis auf zwei Ausnahmen, normalverteilt. Zur Prüfung der Signifikanzen für die einzelnen Abschnitte der Testungen wurde die einfaktorielle Varianzanalyse verwendet. Für die zwei Ausnahmen, die nicht normalverteilt waren, wurde der Friedman Test verwendet. Zum Vergleich der VR-Konditionen untereinander wurden der Friedman Test und der Kruskal-Wallis Test verwendet. Der Vergleich zwischen der N3-Konditionsgruppe und der in vivo-Gruppe erfolgte mittels des ungepaarten t-Tests. Für den Vergleich innerhalb der gesamten Testkohorte kam der gepaarte t-Test zum Einsatz.

2.4.6 Herzfrequenzvariabilität

Analog zur Berechnung der durchschnittlichen Herzfrequenz der einzelnen Abschnitte wurde die Berechnung der Standardabweichung der RR-Intervalle (SDNN) und des quadratischen Mittelwertes der Differenzen aufeinanderfolgender RR-Intervalle (RMSSD) durchgeführt [37]. Dabei erfolgte die Berechnung anhand der folgenden Formeln:

$$RR_{mean} = \frac{1}{n-1} * \sum_{k=1}^n (RR_k)$$

$$SDNN = \sqrt{\frac{1}{n-1} \sum_{k=1}^n (RR_k - RR_{mean})^2}$$

$$RMSSD = \sqrt{\frac{1}{n-1} \sum_{k=1}^n (RR_{k+1} - RR_k)^2}$$

Die errechneten Daten wurden zur statistischen Auswertung nach SPSS übertragen. Die Normalverteilungsprüfung ergab, dass die Daten nicht normalverteilt sind. Zur Prüfung, ob signifikante Unterschiede für die einzelnen Abschnitte der Testungen vorliegen, wurde der Friedman Test genutzt. Überdies wurden der Friedman Test und Kruskal-Wallis Test zum Vergleich der VR-Konditionen untereinander genutzt. Zudem erfolgte der Vergleich zwischen der N3-Konditionsgruppe und der in vivo-Gruppe mit dem Mann-Whitney und Wilcoxon Test.

2.4.7 Fragebögen

Die Eingaben für die Fragebögen erfolgten digital und lagen uns daher – mit Ausnahme des graphischen Fragebogens, der uns in Papierform vorlag – als CSV-Datei vor. Die Übertragung des graphischen Fragebogens in eine CSV-Datei erfolgte manuell. Anschließend wurde die CSV-Datei nach SPSS importiert. Vor der Auswahl der richtigen statistischen Tests erfolgte eine graphische Überprüfung der Daten mithilfe von Histogrammen mit Normalverteilungskurve auf Normalverteilung. Die Daten aller Fragebögen waren nicht normalverteilt. Die einzelnen Abschnitte der jeweiligen Testungen wurden mit dem Friedman Test verglichen. Die Untersuchung nach Lageunterschieden zwischen der N3-Konditionsgruppe und der in vivo-Gruppe erfolgte mithilfe des Mann-Whitney und Wilcoxon Tests.

Ausnahmen stellen der IPQ und SSQ dar. Beim IPQ lagen uns nur einmalige Werte nach der VR-Testung vor. Beim SSQ lagen uns Prä- und Posttest-Werte vor, die allesamt mit dem Wilcoxon Test untersucht wurden.

3 Ergebnisse

3.1 Cortisol

Bei der Betrachtung der einzelnen Abschnitte der VR-Testung für die Gesamtgruppe zeigte sich ein signifikanter Unterschied für alle Konditionen (N3: $\chi^2=27,80$, $p<0,001$; N6: $\chi^2=28,58$, $p<0,001$; N15: $\chi^2=33,93$, $p<0,001$). Der paarweise Vergleich der einzelnen Abschnitte zeigte einen signifikanten Unterschied zwischen der Vorbereitungs- und Ruhephase (N3: $p=0,001$; N6: $p=0,014$; N15: $p<0,001$), zwischen der Testungs- und Ruhephase für die Konditionen N3 und N15 (N3: $p=0,017$; N6: $p=0,54$; N15: $p=0,004$) und zwischen der Ausgangs- und Ruhephase (N3, N6 und N15: $p<0,001$). Zwischen den VR-Testungsabschnitten für die mit der N3-Kondition beginnenden VR-Gruppe zeigte sich kein signifikanter Unterschied ($\chi^2=5,4$, $p=0,15$). Die Cortisolwerte waren im Schnitt während der Testungs- und Ruhephase zwar signifikant höher als der Ruhewert, jedoch sind die Werte in Anbetracht des hohen Cortisolwertes für die Ausgangsphase kritisch zu sehen (siehe Tabelle 2). Diese Erkenntnisse würden die H1-Hypothese nicht unterstützen, da ein klarer Anstieg des Cortisolwertes durch die VR-Exposition für die VR-Gesamtgruppe nicht gezeigt werden konnte.

Tabelle 2: Vergleich der Cortisolwerte für die gesamte Testungspopulation

Mean in ng/ml (SD)	Vivo	N3	N6	N15
Ausgangswert	3,25 (1,53)	2,93 (1,54)	2,93 (1,54)	2,93 (1,54)
Vorbereitung	3,75 (1,81)	2,51 (1,7)	2,17 (1,29)	2,37 (1,21)
TSST/Testung	5,03 (3,01)	2,24 (1,66)	1,87 (1,0)	2,13 (1,03)
Ruhewert	4,12 (3,07)	1,47 (0,86)	1,47 (0,86)	1,47 (0,86)

Beim Vergleich der Testungswerte der einzelnen Konditionen (N3, N6, N15) der Gesamtgruppe untereinander zeigte sich kein signifikanter Unterschied ($\chi^2=1,31$, $p=0,52$). Auch bei der Untersuchung der einzelnen Konditionen unter der Bedingung, dass mit der jeweiligen Kondition die erste Testung durchgeführt worden ist, konnte kein signifikanter Unterschied festgestellt werden ($p=0,41$, siehe Abbildung 1). Ein Einfluss der Avatare auf Cortisolinduktion ließ sich aus den oben genannten Ergebnissen nicht eruieren. Die Ergebnisse untermauern somit die H2-Hypothese nicht.

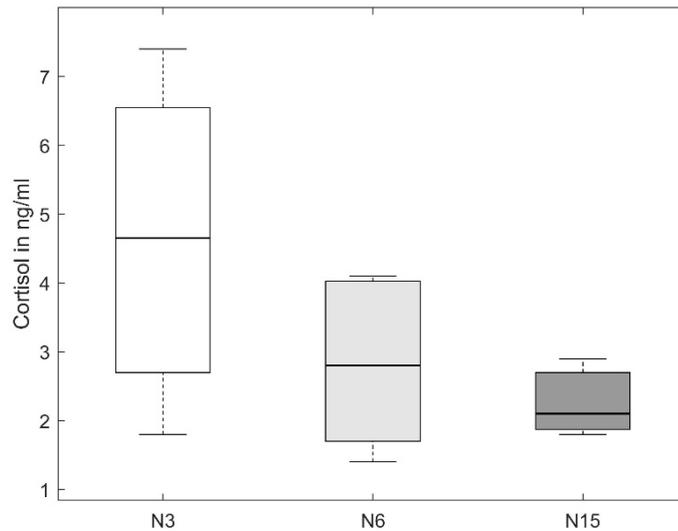


Abbildung 1: Vergleich der Cortisolwerte zwischen den VR-Konditionen. Es sind die jeweils mit der Kondition Beginnenden dargestellt.

Zwischen den einzelnen *in vivo*-Testungsabschnitten der gesamten Gruppe konnte ein signifikanter Unterschied ($\chi^2=12,04$, $p=0,007$) aufgezeigt werden. Beim paarweisen Vergleich zeigte sich ein signifikanter Unterschied zwischen der Testungs- und Ruhephase ($p=0,005$). Zwischen den anderen Phasen zeigte sich dagegen kein signifikanter Unterschied. Bei der Betrachtung der mit der *in vivo*-Testung beginnenden Gruppe konnte für die einzelnen Abschnitte der Testung kein signifikanter Unterschied festgestellt werden ($\chi^2=4,77$, $p=0,19$, siehe Abbildung 2). Die beiden Untersuchungen und die Grafik zeigen, dass die Cortisolkonzentration im Speichel angestiegen ist, jedoch die Unterschiede zwischen den einzelnen

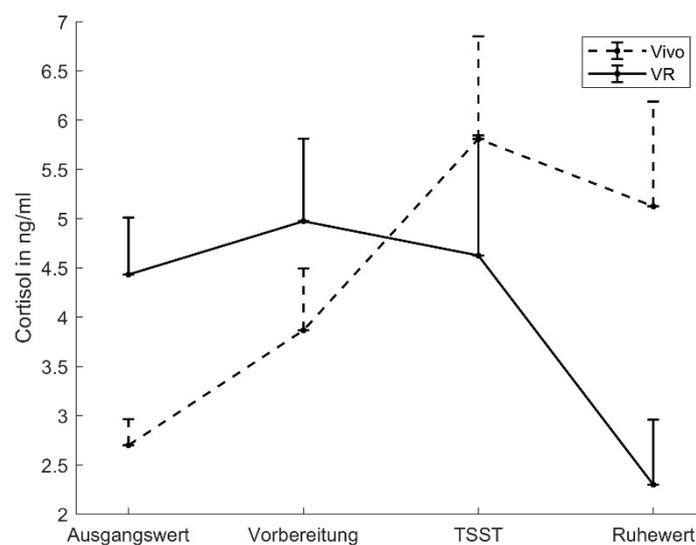


Abbildung 2: Cortisolverlauf der *in vivo*- und N3-Gruppe. Es sind die jeweils mit der Kondition Beginnenden dargestellt.

Abschnitten bis auf eine Ausnahme, nicht signifikant sind. Somit wird die H3-Hypothese nicht eindeutig gestützt.

Beim Vergleich zwischen der mit N3 bzw. in vivo beginnenden Gruppe konnte kein signifikanter Unterschied für den Testungswert ($p=0,6$) festgestellt werden. Bei der Betrachtung der gesamten N3-Gruppe zeigte sich im Vergleich zu der in vivo-Gruppe, jeweils unabhängig der Testungsreihenfolge, für den Wert der Testungsphase ein signifikanter Unterschied ($p=0,001$). Die absoluten Werte für die Testung sind bei der in vivo-Testung tendenziell höher als bei der VR-Testung. Insbesondere die Cortisolwerte für die gesamte Gruppe, unabhängig der Reihenfolge, sind in der in vivo-Gruppe signifikant höher und dies würde die H4-Hypothese widerlegen (siehe Tabelle 3).

Tabelle 3: Vergleich der Cortisolwerte zwischen der in vivo- und N3-Gruppe

Mean in ng/ml (SD)	Vivo	N3
Beginnend	5,81 (3,61)	4,62 (2,44)
Unabhäng. Reihenfolge	5,03 (3,0)	2,24 (1,66)

Bei der Betrachtung der Einzelverlaufskurven wurden jeweils 12 Probanden/innen in die Gruppe der Responder bzw. Non-Responder kategorisiert. Es zeigte sich ein signifikanter Unterschied zwischen den Abschnitten der in vivo-Testung ($\chi^2=20,2$, $p<0,001$). Beim paarweisen Vergleich stellte sich ein signifikanter Unterschied zwischen dem Testungswert und dem Ausgangswert ($p<0,001$) heraus. Darüber hinaus unterschied sich der Vorbereitungsvom Testungswert signifikant ($p=0,03$). Im Gegensatz zu der Betrachtung der gesamten in vivo-Gruppe konnten durch die Separierung in Responder und Non-Responder signifikante Unterschiede gezeigt werden (siehe Abbildung 3).

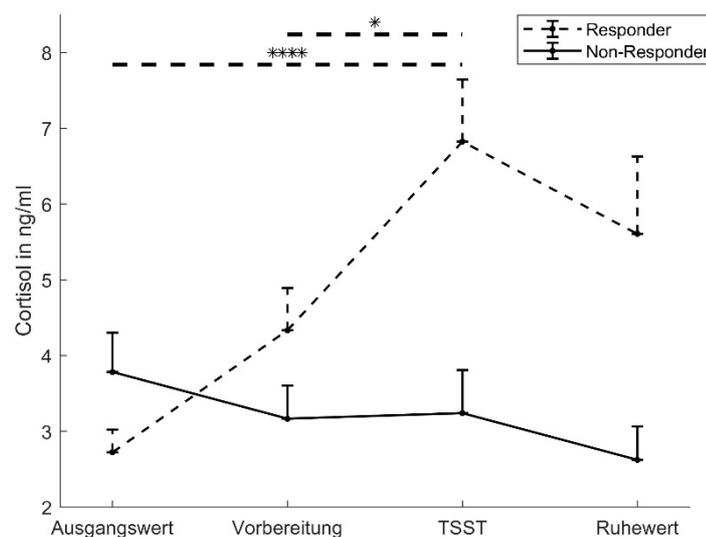


Abbildung 3: Cortisolverlauf der Responder und Non-Responder im Vergleich

3.2 Hautleitwiderstand (GSR)

Für die einzelnen Abschnitte der VR-Testung für die Gesamtgruppe zeigte sich ein signifikanter Unterschied (N3: $\chi^2=29,9$, $p<0,001$; N6: $\chi^2=25,55$, $p<0,001$; N15: $\chi^2=27,45$, $p<0,001$). Beim paarweisen Vergleich stellte sich für alle Konditionen ein signifikanter Unterschied zwischen dem Wert der Ausgangs- und Vorbereitungs- bzw. Testungsphase heraus (N3: $p=0,005$, $p<0,001$; N6: $p=0,007$, $p<0,001$; N15: $p=0,007$, $p<0,001$). Darüber hinaus zeigte sich ein signifikanter Unterschied zwischen der Testungs- und Ruhephase (N3: $p=0,005$; N6: $p=0,031$; N15: $p=0,031$). Die Untersuchung der mit der N3-Kondition beginnenden Gruppe zeigte einen signifikanten Unterschied zwischen den Abschnitten der VR-Testung auf ($\chi^2=11,1$ $p=0,011$). Beim paarweisen Vergleich konnte ein signifikanter Unterschied zwischen dem Testungs- und Ausgangswert festgestellt werden ($p=0,02$). Die vorliegenden signifikanten Steigerungen der GSR-Werte zwischen dem Testungs- bzw. Vorbereitungswert und Ausgangswert und der anschließende Abfall der GSR-Werte stützen die H1-Hypothese (siehe Tabelle 4).

Tabelle 4: Vergleich des Hautleitwiderstandes für die gesamte Testungspopulation

Mean in μS (SD)	Vivo	N3	N6	N15
Ausgangswert	0,35 (0,20)	0,30 (0,15)	0,30 (0,15)	0,30 (0,15)
Vorbereitung	0,66 (0,12)	0,44 (0,15)	0,43 (0,17)	0,46 (0,17)
TSST/Testung	0,73 (0,13)	0,55 (0,18)	0,52 (0,17)	0,50 (0,21)
Ruhewert	0,40 (0,20)	0,38 (0,14)	0,38 (0,14)	0,38 (0,14)

Beim Vergleich der einzelnen Konditionen untereinander konnte für die Gesamtgruppe kein signifikanter Unterschied aufgezeigt werden ($\chi^2=0,75$, $p=0,69$). Bei der Untersuchung zwischen den mit der jeweiligen Kondition beginnenden Gruppen konnte kein signifikanter Unterschied festgestellt werden ($p=0,57$, siehe Abbildung 4). Unabhängig davon, ob die Testungsreihenfolge beachtet wurde oder nicht, konnte kein signifikanter Unterschied festgestellt werden. Die H2-Hypothese wird durch die Ergebnisse nicht bekräftigt.

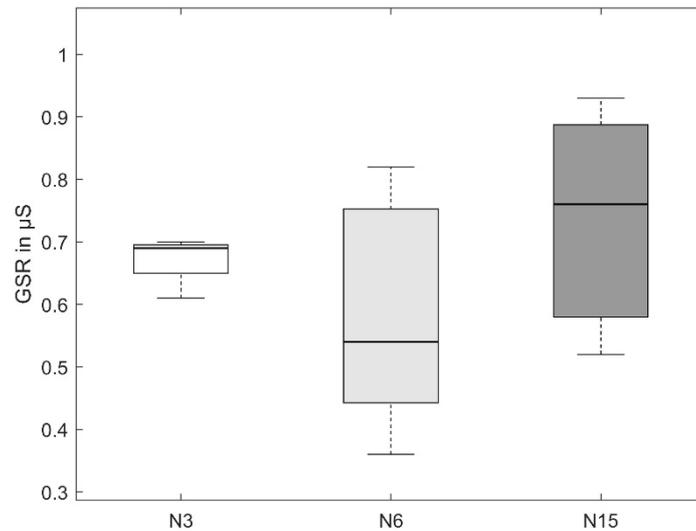


Abbildung 4: Vergleich des Hautleitwiderstandes zwischen den VR-Konditionen. Es sind jeweils die mit der Kondition Beginnenden dargestellt.

Bei der Betrachtung der einzelnen Abschnitte der *in vivo*-Testung konnte ein signifikanter Unterschied zwischen den einzelnen Abschnitten aufgezeigt werden ($\chi^2=52,25$, $p<0,001$). Der paarweise Vergleich zeigte zwischen dem Ausgangswert und dem Wert für die Vorbereitungs- bzw. Testungsphase einen signifikanten Unterschied ($p<0,001$). Zudem zeigte sich auch ein signifikanter Unterschied zwischen der Ruhephase und der Vorbereitungs- bzw. Testungsphase ($p<0,001$). Bei der Analyse der mit der *in vivo*-Testung beginnenden Gruppe konnte zwischen den Abschnitten der Testung ein signifikanter Unterschied festgestellt werden ($\chi^2=24,7$, $p<0,001$). Der paarweise Vergleich zeigte einen signifikanten Unterschied zwischen der Vorbereitungs- bzw. Testungsphase und dem Ausgangswert ($p=0,002$; $p<0,001$). Darüber hinaus konnte ein signifikanter Unterschied zwischen der Testungsphase und dem Ruhewert festgestellt werden ($p=0,02$, siehe Abbildung 5). Durch die oben genannten Ergebnisse konnte gezeigt werden, dass durch die *in vivo*-Exposition eine signifikante Erhöhung des GSR induziert werden kann, wodurch die H3-Hypothese bestätigt wird.

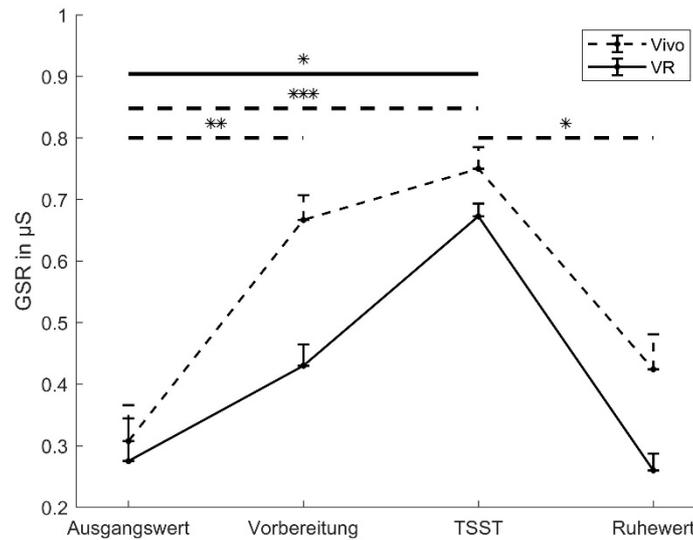


Abbildung 5: GSR-Verlauf der in vivo- und N3-Gruppe. Es sind die jeweils mit der Kondition Beginnenden dargestellt.

Beim Vergleich zwischen der Gruppe, die mit der VR-Testung in der N3-Kondition begann und der mit der in vivo-Testung beginnenden Gruppe konnte kein signifikanter Unterschied für den Wert der Testungsphase ($p=0,21$) aufgezeigt werden. Bei nicht Berücksichtigung der Testreihenfolge war ein signifikanter Unterschied für den Wert der Testungsphase ($p<0,001$) festzustellen. Bezüglich des GSRs wies die in vivo-Gruppe höhere Werte als die N3-Konditionsgruppe auf, jedoch sind die Unterschiede nur bei der Betrachtung der gesamten Gruppe ohne Berücksichtigung der Testreihenfolge signifikant (siehe Tabelle 5). Somit wird die H4-Hypothese entkräftigt.

Tabelle 5: Vergleich des Hautleitwiderstandes zwischen der in vivo- und N3-Gruppe

Mean in μS (SD)	Vivo	N3
Beginnend	0,75 (0,12)	0,67 (0,04)
Unabhäng. Reihenfolge	0,73 (0,13)	0,55 (0,18)

3.3 Herzfrequenz

Für die einzelnen Abschnitte der VR-Testung konnte ein signifikanter Unterschied gezeigt werden (N3: $p<0,001$; N6: $p<0,001$, N15: $p<0,001$). Beim paarweisen Vergleich der einzelnen Abschnitte konnte für jede Kondition ein signifikanter Unterschied zwischen dem Ausgangswert und dem Ruhewert aufgezeigt werden ($p<0,05$). Zudem unterschied sich der Vorbereitungs- respektive der Testungswert vom Ruhewert signifikant ($p<0,05$). Die einzelnen Abschnitte der mit der N3-Kondition beginnenden Gruppe unterschieden sich signifikant ($\chi^2=10,8$, $p=0,013$) und der paarweise Vergleich zeigte ausschließlich einen signifikanten Un-

terschied zwischen dem Testungs- und Ruhewert ($p=0,006$). In Anbetracht des relativen hohen Ausgangswertes für die gesamte Gruppe kann keine Steigerung der Herzfrequenz im Rahmen der Testung festgestellt werden (siehe Tabelle 6). Ungeachtet der fehlenden Steigerung ist ein signifikanter Abfall der Herzfrequenz nach der Testung zu beobachten. Für die mit N3-Kondition beginnenden Gruppe konnte ein signifikanter Unterschied zwischen dem Testungswert und Ruhewert aufgezeigt werden ($p=0,01$). Insgesamt konnte kein eindeutiger Effekt der VR-Exposition aufgezeigt werden, sodass die H1-Hypothese tendenziell abzulehnen wäre.

Tabelle 6: Vergleich der Herzfrequenzen für die gesamte Testungspopulation

Mean in bpm (SD)	Vivo	N3	N6	N15
Ausgangswert	83,51 (10,82)	79,43 (8,61)	79,43 (8,61)	79,43 (8,61)
Vorbereitung	96,06 (11,62)	77,29 (9,43)	77,90 (8,96)	77,60 (9,22)
TSST/Testung	93,99 (12,60)	77,93 (9,49)	78,81 (9,87)	78,03 (8,92)
Ruhewert	78,05 (8,70)	73,44 (8,75)	73,44 (8,75)	73,44 (8,75)

Der Vergleich der Testungswerte zwischen den einzelnen VR-Konditionen zeigte keinen signifikanten Unterschied ($p=0,76$). Die Aufteilung in die jeweils mit der Kondition beginnenden Gruppen zeigte keinen signifikanten Unterschied zwischen den Konditionen ($p=0,46$, siehe Abbildung 6). In Hinblick auf die H2-Hypothese lässt sich der Einfluss der Avatare auf die Herzfrequenz verneinen.

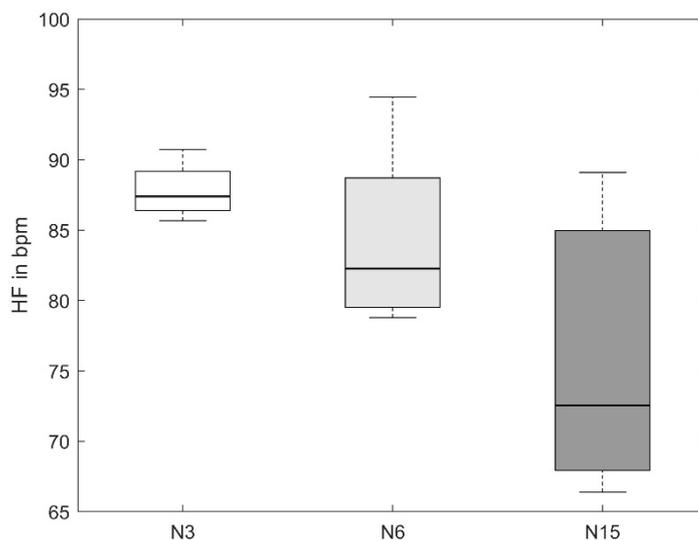


Abbildung 6: Vergleich der Herzfrequenzen zwischen den VR-Konditionen. Es sind die jeweils mit der Kondition Beginnenden dargestellt.

Bei der Betrachtung der *in vivo*-Testung konnte zwischen den einzelnen Abschnitten der Testung ein signifikanter Unterschied aufgezeigt werden ($p < 0,001$). Der paarweise Vergleich zeigte einen signifikanten Unterschied zwischen dem Testungs- und Ausgangswert ($p < 0,001$) bzw. Ruhewert ($p < 0,001$). Zudem zeigte sich ein signifikanter Unterschied zwischen dem Vorbereitungswert und Ausgangswert ($p < 0,001$) bzw. Ruhewert ($p < 0,001$). Es konnte kein signifikanter Unterschied zwischen der Testungs- und Vorbereitungsphase aufgezeigt werden ($p = 1,0$). Zwischen den Testungsabschnitten für die mit *in vivo* beginnende Gruppe zeigte sich ein signifikanter Unterschied ($\chi^2 = 29,8$, $p < 0,001$). Der paarweise Vergleich zeigte einen signifikanten Unterschied zwischen dem Vorbereitungs- bzw. Testungswert und dem Ausgangs- ($p = 0,003$; $p = 0,009$) und Ruhewert ($p < 0,001$; $p < 0,001$, siehe Abbildung 7). Die Ergebnisse unterstützen die H3-Hypothese, da die Untersuchungen einen signifikanten Einfluss auf die Herzfrequenz durch die *in vivo*-Testung zeigen.

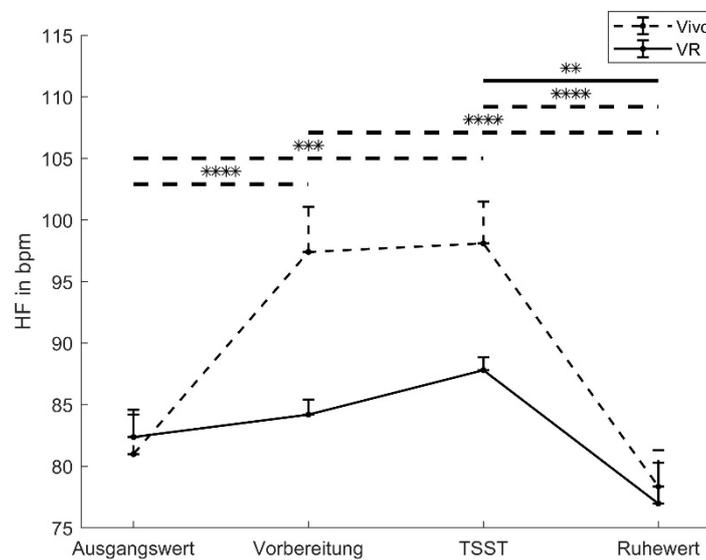


Abbildung 7: Herzfrequenzverlauf der *in vivo*- und VR-Gruppe.
Es sind die jeweils mit der Kondition Beginnenden dargestellt.

Der Vergleich der Testungswerte zwischen der Konditionsgruppe, die mit N3 begann, und der mit der *in vivo*-Testung beginnenden Gruppe mithilfe des ungepaarten t-Tests zeigte einen signifikanten Unterschied auf ($p = 0,013$, Varianzheterogenität, Levene Test $p = 0,018$). Bei der Betrachtung der gesamten N3-Gruppe zeigte der gepaarte t-Test im Vergleich zu der *in vivo*-Gruppe einen signifikanten Unterschied auf ($p < 0,001$). Hinsichtlich der Herzfrequenz im Vergleich zwischen der N3- und *in vivo*-Gruppe zeigten sich signifikant höhere Werte für die *in vivo*-Testung (siehe Tabelle 7), sodass die H4-Hypothese falsifiziert wird.

Tabelle 7: Vergleich der Herzfrequenzen zwischen der in vivo- und N3-Gruppe

Mean in bpm (SD)	Vivo	N3
Beginnend	98,1 (11,73)	87,8 (2,13)
Unabhäng. Reihenfolge	93,99 (12,6)	77,93 (9,49)

3.4 Herzfrequenzvariabilität

3.4.1 RMSSD

Es zeigten sich signifikante Unterschiede zwischen den Testungsabschnitten für jede Kondition der Gesamtgruppe (N3: $\chi^2=15,95$, $p=0,001$, N6: $\chi^2=13,57$, $p=0,004$; N15: $\chi^2=18,95$, $p<0,001$). Beim paarweisen Vergleich zeigte sich für die N3-Kondition ein signifikanter Unterschied zwischen dem Ausgangs- und Ruhewert ($p=0,001$) sowie zwischen dem Vorbereitungs- und Ruhewert ($p=0,02$). Bei der N6-Kondition zeigte sich nur ein signifikanter Unterschied zwischen dem Ausgangs- und Ruhewert ($p=0,002$). Für die N15-Kondition konnte der paarweise Vergleich einen signifikanten Unterschied zwischen dem Ausgangs- und Testungswert ($p=0,02$), Ausgangs- und Ruhewert ($p=0,001$) und dem Vorbereitungs- und Ruhewert ($p=0,03$) aufzeigen. Für die mit der N3-Kondition beginnenden Gruppe konnte ein signifikanter Unterschied zwischen den Testungsabschnitten festgestellt werden ($\chi^2=8,1$, $p=0,044$). Der paarweise Vergleich zeigte einen signifikanten Unterschied zwischen dem Vorbereitungswert und dem Ruhewert ($p=0,037$). Die absoluten Werte für die Gesamtgruppen sind aufgrund des hohen Ausgangswertes schwierig einzuordnen, da die Reduktion des RMSSDs während der Testung ausfällt (siehe Tabelle 8). Auch bei der Betrachtung der mit der N3-Kondition beginnenden Gruppe konnte nur eine geringfügige und nicht signifikante Absenkung festgestellt werden. Unabhängig der Betrachtungsweise (Gesamtgruppe oder N3-Kondition) konnte als Zeichen der Entspannung ein Anstieg des RMSSDs am Ende des Tests beobachtet werden. Die oben genannten Ergebnisse stützen die H1-Hypothese nur teilweise.

Tabelle 8: Vergleich der RMSSD-Werte für die gesamte Testungspopulation

Mean in ms (SD)	Vivo	N3	N6	N15
Ausgangswert	40,82 (31,19)	42,77 (21,85)	42,77 (21,85)	42,77 (21,85)
Vorbereitung	42,95 (38,41)	50,52 (40,51)	47,77 (27,03)	48,69 (36,44)
TSST/Testung	42,20 (26,49)	48,98 (26,95)	62,66 (76,55)	52,13 (30,20)
Ruhewert	57,07 (35,86)	55,73 (38,77)	55,73 (38,77)	55,73 (38,77)

Zwischen den einzelnen Konditionen der VR-Testung konnte für den Testungswert kein signifikanter Unterschied aufgezeigt werden ($\chi^2=0,75$, $p=0,69$). Bei der Analyse der mit der jeweiligen Kondition beginnenden Gruppe war kein signifikanter Unterschied feststellbar

($p=0,35$, siehe Abbildung 8). Die Ergebnisse der einzelnen Konditionen unterschieden sich nicht signifikant und somit ist die H2-Hypothese abzulehnen.

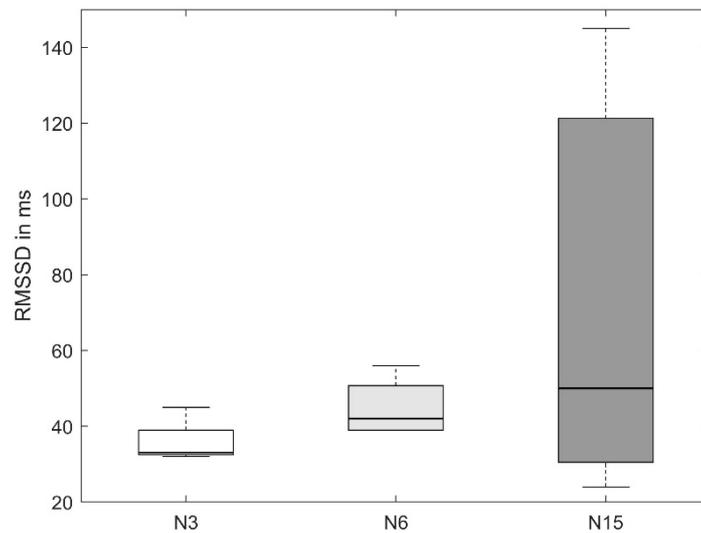


Abbildung 8: RMSSD für die einzelnen VR-Konditionen. Es sind die jeweils mit der Kondition Beginnenden dargestellt.

Für die *in vivo*-Testung attestierte der Friedman Test einen signifikanten Unterschied zwischen den einzelnen Abschnitten ($\chi^2=24,65$, $p<0,001$). Der paarweise Vergleich zeigte einen signifikanten Unterschied zwischen dem Testungs- und Ruhewert ($p=0,002$), dem Vorbereitungs- und Ruhewert ($p<0,001$) sowie zwischen dem Ausgangs- und dem Ruhewert ($p=0,001$). Bei der Untersuchung der mit der *in vivo*-Testung beginnenden Gruppe zeigte sich ein signifikanter Unterschied zwischen den Testungsabschnitten ($\chi^2=11,7$, $p=0,01$). Beim paarweisen Vergleich konnten signifikante Unterschiede zwischen dem Vorbereitungs- bzw. Testungswert und Ruhewert festgestellt werden ($p=0,009$; $p=0,04$). Die Ergebnisse der *in vivo*-Gruppe zeigen eine signifikante Senkung während der Vorbereitungs- und Testungsphase im Vergleich zum Ruhewert, jedoch nicht im Vergleich zum Ausgangswert. Analog zur VR-Testung zeigte sich eine signifikante Steigerung des RMSSDs als Zeichen der Entspannung mit Beendigung der Testung. Zusammenfassend stützen die Ergebnisse die H3-Hypothese teilweise.

Beim Vergleich zwischen der mit der N3-Kondition beginnenden Gruppe und der mit *in vivo* beginnenden Gruppe konnte kein signifikanter Unterschied für den Testungswert aufzuzeigen werden ($p=0,86$). Unter der Annahme der Nichtberücksichtigung der Testreihenfolge konnte zwischen den beiden Gruppen kein signifikanter Unterschied festgestellt werden ($p=0,1$). Die Ergebnisse hinsichtlich des Vergleiches zwischen den beiden Gruppen sind nicht klar einzuordnen, da bei beiden Gruppen ein starker und signifikanter Abfall des RMSSDs während der Testung nicht vorhanden ist (siehe Abbildung 9). Dennoch lässt sich feststellen, dass es nicht zu einer stärkeren Reaktion in der VR-Gruppe kam und somit ist die H4-Hypothese abzulehnen.

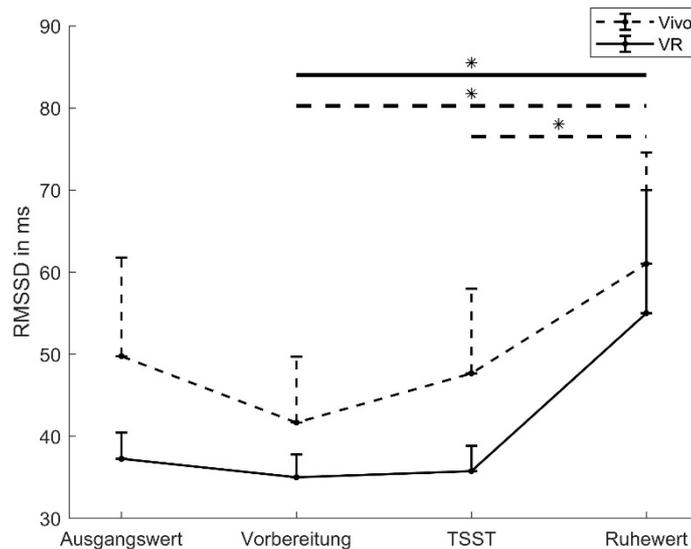


Abbildung 9: RMSSD-Verlauf der in vivo- und N3-Gruppe. Es sind die jeweils mit der Kondition Beginnenden dargestellt.

3.4.2 SDNN

Zwischen den VR-Testungsabschnitten zeigten sich signifikante Unterschiede für jede Kondition (N3: $\chi^2=25,1$, $p<0,001$; N6: $\chi^2=24,72$, $p<0,001$; N15: $\chi^2=28,1$, $p<0,001$). Der paarweise Vergleich für die N3-Kondition zeigte einen signifikanten Unterschied zwischen dem Ausgangs- und Testungswert ($p=0,04$), dem Ausgangs- und Ruhewert ($p<0,001$) sowie zwischen dem Vorbereitungs- und Ruhewert ($p=0,01$). Für die N6-Kondition zeigte der paarweise Vergleich einen signifikanten Unterschied zwischen dem Ausgangs- und Ruhewert ($p<0,001$), dem Vorbereitungs- und Ruhewert ($p=0,02$) sowie zwischen dem Testungs- und Ruhewert ($p=0,03$). Bezüglich der N15-Kondition zeigte sich ein signifikanter Unterschied zwischen dem Ausgangs- und Ruhewert ($p<0,001$), dem Vorbereitungs- und Ruhewert ($p<0,01$) sowie zwischen dem Testungs- und dem Ruhewert ($p=0,01$). Bei der Betrachtung der mit der N3-Kondition beginnenden Gruppe konnte ein signifikanter Unterschied zwischen den Testungsabschnitten festgestellt werden ($\chi^2=10,2$, $p=0,02$). Beim paarweisen Vergleich zeigte sich ein signifikanter Unterschied zwischen dem Vorbereitungs- und Ruhewert ($p=0,04$). Bei der Betrachtung der Ergebnisse für die gesamte Gruppe lässt sich trotz signifikanter Unterschiede kein Trend, d.h. ein Abfall des SDNNs während der Testung, eruieren (siehe Tabelle 9), sodass die H1-Hypothese entkräftigt wird. Darüber hinaus konnte die mit der N3-Kondition beginnende Gruppe keinen klaren Trend aufweisen.

Tabelle 9: Vergleich der SDNN-Werte für die gesamte Testungspopulation

Mean in ms (SD)	Vivo	N3	N6	N15
Ausgangswert	74,13 (41,39)	65,32 (26,93)	65,32 (26,93)	65,32 (26,93)
Vorbereitung	84,74 (33,10)	74,80 (33,22)	73,14 (29,24)	71,81 (31,91)
TSST/Testung	67,64 (17,64)	74,01 (26,98)	79,53 (50,88)	73,91 (27,43)
Ruhewert	87,83 (37,42)	83,17 (32,48)	83,17 (32,48)	83,17 (32,48)

Der Vergleich der einzelnen VR-Konditionen (N3, N6, N15) untereinander zeigte für den Testungswert keinen signifikanten Unterschied ($\chi^2=1,0$, $p=0,61$). Bei Aufteilung in die mit den jeweiligen VR-Konditionen beginnenden Gruppen konnte kein signifikanter Unterschied zwischen den Konditionen ($p=0,75$, siehe Abbildung 10) festgestellt werden. In Hinblick auf die H2-Hypothese lässt sich der Einfluss der Avatare auf SDNN verneinen.

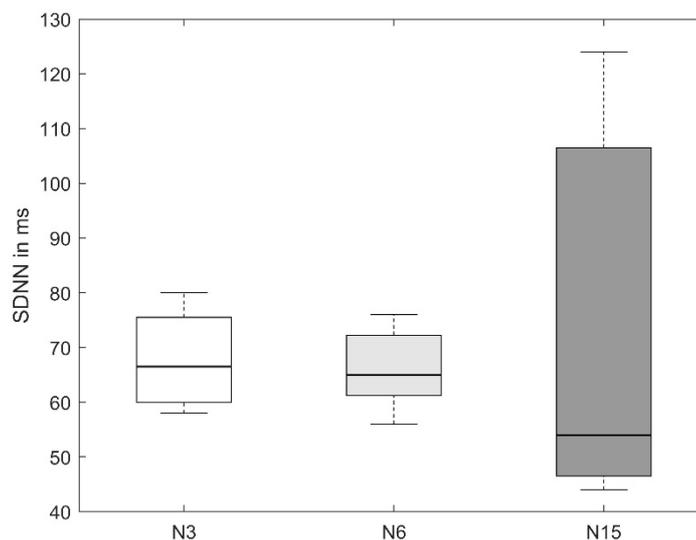


Abbildung 10: SDNN für die einzelnen VR-Konditionen. Es sind die jeweils mit der Kondition Beginnenden dargestellt.

Für die in vivo-Testung lässt sich ein signifikanter Unterschied zwischen den verschiedenen Abschnitten ($\chi^2=16,15$, $p=0,001$) attestieren. Beim paarweisen Vergleich zeigte sich ein signifikanter Unterschied zwischen dem Testungs- und Ruhewert ($p=0,002$) sowie zwischen dem Ausgangs- und Ruhewert ($p=0,01$). Die Untersuchung der Gruppe, die mit der in vivo-Testung begann, zeigte keinen signifikanten Unterschied zwischen den Testungsabschnitten auf ($\chi^2=6,7$, $p=0,082$). Abgesehen von einem nicht signifikanten Abfall des SDNNs während der Testungsphase (siehe Abbildung 11) lässt sich keine eindeutige Reaktion auf die in vivo-Exposition feststellen, sodass die H3-Hypothese entkräftigt wird.

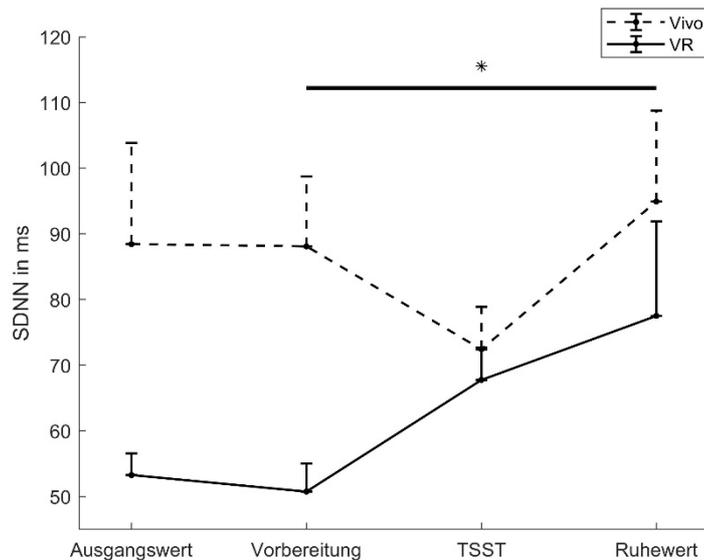


Abbildung 11: SDNN-Verlauf der in vivo- und N3-Gruppe. Es sind die jeweils mit der Kondition Beginnenden dargestellt.

Beim Vergleich der mit der N3-Kondition beginnenden Gruppe mit der in vivo beginnenden Gruppe zeigte sich für den Testungswert kein signifikanter Unterschied ($p=0,68$). Bei Betrachtung der beiden Gruppen ohne Berücksichtigung der Reihenfolge konnte kein signifikanter Unterschied aufgezeigt werden ($p=0,44$). Eine stärkere Reaktion auf eine VR-Exposition konnte nicht gezeigt werden, weshalb die H4-Hypothese zu falsifizieren ist.

3.5 Fragebögen

3.5.1 Self-Assessment Manikin

Zwischen den VR-Testungsabschnitten zeigte sich für die Valenz (N3: $\chi^2=13,75$, $p=0,01$; N6: $\chi^2=19,63$, $p<0,001$; N15: $\chi^2=18,82$, $p<0,001$), für das Arousal in der N15-Kondition (N15: $\chi^2=6,71$, $p=0,04$) und für die Dominanz in der N6-Kondition (N6: $\chi^2=7,12$, $p=0,03$) ein signifikanter Unterschied. Die anderen Konditionen des Arousal (N3: $\chi^2=2,44$, $p=0,30$; N6: $\chi^2=5,74$, $p=0,06$) und der Dominanz (N3: $\chi^2=2,98$, $p=0,23$; N15: $\chi^2=2,98$, $p=0,23$) unterschieden sich nicht signifikant. Die paarweise Prüfung zeigte für die Valenz einen signifikanten Unterschied zwischen der Vorbereitungs- respektive Testungsphase und dem Ausgangswert (N3: $p=0,02$, $p=0,04$; N6: $p=0,01$, $p=0,01$; N15: $p=0,01$, $p=0,02$). Für das Arousal und die Dominanz konnten im paarweisen Vergleich keine signifikanten Unterschiede festgestellt werden. Bei der Betrachtung der mit der N3-Kondition beginnenden Gruppe konnte für die Valenz und das Arousal kein signifikanter Unterschied zwischen den Testungsabschnitten aufgezeigt werden (Valenz: $\chi^2=3,86$, $p=0,15$; Arousal: $\chi^2=2,6$, $p=0,27$). Für die Dominanz zeigte sich ein signifikanter Unterschied (Dominanz: $\chi^2=6,62$, $p=0,04$), jedoch waren die Ergebnisse der paarweisen Prüfung nicht signifikant. Zusammenfassend ist nur

bei der Betrachtung der Gesamtgruppe für die Valenz ein signifikanter Unterschied festzustellen. D.h., dass die Emotionslage der Probanden/innen nach der Testung etwas „negativer“ ist. Für die anderen Aspekte des SAM konnten keine signifikanten Unterschiede aufgezeigt werden. Die Ergebnisse (siehe Tabelle 10) lassen nicht auf eine starke Veränderung im Laufe der VR-Testung schließen, sodass die H1-Hypothese entkräftigt wird.

Tabelle 10: SAM-Fragebogen mit den Subskalen für die gesamte Testungspopulation

Mean (SD)	Vivo	N3	N6	N15
Valenz				
Ausgangswert	3,04 (1,08)	3,04 (1,00)	3,04 (1,00)	3,04 (1,00)
Vorbereitung	4,33 (1,17)	4,25 (1,45)	4,04 (1,37)	4,08 (1,50)
TSST/Testung	3,54 (1,41)	4,00 (1,64)	4,25 (1,36)	3,88 (1,19)
Arousal				
Ausgangswert	6,71 (1,46)	7,04 (1,60)	7,04 (1,60)	7,04 (1,60)
Vorbereitung	4,79 (2,04)	6,67 (1,76)	6,38 (1,88)	6,42 (1,69)
TSST/Testung	6,00 (2,09)	6,71 (1,88)	6,54 (1,84)	6,58 (1,74)
Dominanz				
Ausgangswert	5,04 (1,49)	5,25 (1,39)	5,25 (1,39)	5,25 (1,39)
Vorbereitung	4,38 (1,28)	4,96 (1,57)	5,00 (1,44)	5,00 (1,67)
TSST/Testung	4,96 (1,43)	4,92 (1,74)	4,83 (1,52)	4,96 (1,57)

Beim Vergleich der einzelnen Konditionen (N3, N6, N15) untereinander konnte für keine Komponente des SAMs ein signifikanter Unterschied aufgezeigt werden (Valenz: $\chi^2=0,3$, $p=0,86$; Arousal: $\chi^2=0,96$, $p=0,62$; Dominanz: $\chi^2=0,21$, $p=0,9$). Die jeweils mit einer VR-Kondition beginnenden Gruppen zeigten keinen signifikanten Unterschied (Valenz: $p=0,83$; Arousal: $p=0,32$; Dominanz: $p=0,11$). Ein Einfluss der Anzahl der virtuellen Avatare auf den SAM war nicht erkennbar, sodass die H2-Hypothese hinsichtlich des Fragebogens falsifiziert wird.

Zwischen den Abschnitten der in vivo-Testung zeigten sich signifikante Unterschiede (Valenz: $\chi^2=17,94$, $p<0,001$; Arousal: $\chi^2=14,14$, $p=0,001$; Dominanz: $\chi^2=6,53$, $p=0,04$). Der paarweise Vergleich attestierte einen signifikanten Unterschied zwischen dem Ausgangs- und Vorbereitungswert für die Valenz und das Arousal ($p=0,001$). Für die Dominanz zeigte sich kein signifikanter Unterschied.

Für die Gruppe, die mit der in vivo-Testung begann, konnte ein signifikanter Unterschied zwischen den Testungsabschnitten für die Valenz ($\chi^2=9,39$, $p=0,01$) und für das Arousal ($\chi^2=13,91$, $p=0,001$) gezeigt werden. Beim paarweisen Vergleich zeigte sich ein signifikanter Unterschied zwischen dem Vorbereitungs- und Ausgangswert (Valenz: $p=0,02$; Arousal: $p=0,001$). Für die Dominanz konnte kein signifikanter Unterschied gezeigt werden ($\chi^2=5,52$, $p=0,06$). Bei der Betrachtung der in vivo-Testung waren nur signifikante Unterschiede be-

züglich der Vorbereitungsphase festzustellen, sodass ein Einfluss auf SAM durch die eigentliche Testung nicht nachgewiesen werden konnte. Die H3-Hypothese wird daher nur partiell gestützt.

Beim Vergleich der mit der N3-Kondition beginnenden Gruppe mit der in vivo beginnenden Gruppe war kein signifikanter Unterschied für die einzelnen Komponenten des SAMs festzustellen (Valenz: $p=0,38$; Arousal: $p=0,38$; Dominanz: $p=0,21$). Auch bei der Betrachtung der beiden Gruppen ohne Berücksichtigung der Testreihenfolge zeigte sich kein signifikanter Unterschied (Valenz: $p=0,18$; Arousal: $p=0,07$; Dominanz: $p=0,87$). Insgesamt ist keine klare Tendenz bezüglich der Reaktion auf die VR-Exposition zu erkennen (siehe Tabelle 11).

Tabelle 11: SAM-Fragebogen im Vergleich zwischen der in vivo- und N3-Gruppe

Mean (SD)	Vivo	N3
Valenz		
Beginnend	3,58 (1,62)	4,75 (2,22)
Unabhäng. Reihenfolge	3,54 (1,41)	4,00 (1,64)
Arousal		
Beginnend	5,50 (2,11)	6,75 (2,63)
Unabhäng. Reihenfolge	6,00 (2,09)	6,71 (1,88)
Dominanz		
Beginnend	4,75 (1,54)	3,50 (1,29)
Unabhäng. Reihenfolge	4,96 (1,43)	4,92 (1,74)

3.5.2 State-Trait-Angst-Depressions-Inventar State

Zwischen den einzelnen Abschnitten der VR-Testungen konnte für die Subskala Angst in der N6-Kondition ($\chi^2=8,74$, $p=0,13$) und für die Subskala Depression in der N15-Kondition ($\chi^2=6,04$, $p=0,049$) ein signifikanter Unterschied nachgewiesen werden. Der paarweise Vergleich wies für die Subskala Angst der N6-Kondition einen signifikanten Unterschied zwischen dem Vorbereitungs- und Testungswert nach ($p=0,04$). Für die Subskala Depression in der N15-Kondition konnte im paarweisen Vergleich kein signifikanter Unterschied aufgezeigt werden. Bei der Betrachtung der mit der N3-Kondition beginnenden Gruppe konnte kein signifikanter Unterschied für die Angst ($\chi^2=5,29$, $p=0,07$) und Depression ($\chi^2=1,27$, $p=0,53$) gezeigt werden. Die absoluten Ergebnisse (siehe Tabelle 12) zeigen eine geringe, nicht signifikante Steigerung der Werte für die Subskalen, sodass von einer Angstinduktion durch die VR-Testung nicht ausgegangen werden kann und die H1-Hypothese somit nicht gestützt wird.

Tabelle 12: STADI-S für die gesamte Testungspopulation

Mean (SD)	Vivo	N3	N6	N15
Angst				
Ausgangswert	53,71 (9,30)	47,38 (9,77)	47,38 (9,77)	47,38 (9,77)
Vorbereitung	62,17 (10,36)	50,71 (11,77)	52,67 (11,04)	50,67 (12,25)
TSSST/Testung	50,46 (9,02)	46,83 (10,56)	48,17 (12,58)	48,92 (11,17)
Depression				
Ausgangswert	48,88 (9,85)	48,17 (11,51)	48,17 (11,51)	48,17 (11,51)
Vorbereitung	52,04 (11,34)	51,67 (11,59)	51,29 (10,37)	51,38 (11,82)
TSSST/Testung	47,50 (11,50)	50,13 (11,82)	51,50 (12,25)	50,17 (11,29)

Beim Vergleich der einzelnen VR-Konditionen untereinander konnte weder für die Subskala Angst ($\chi^2=1,1$, $p=0,58$) noch für die Depression ($\chi^2=1,37$, $p=0,5$) ein signifikanter Unterschied nachgewiesen werden. Bei einer Aufteilung in die mit der jeweiligen VR-Kondition beginnenden Gruppen konnte kein signifikanter Unterschied aufgewiesen werden (Angst: $p=0,25$; Depression: $p=0,35$). Ein Einfluss der virtuellen Avatare auf die empfundene Angst konnte nicht gezeigt werden. Daher ist die H2-Hypothese abzulehnen.

Es konnte ein signifikanter Unterschied zwischen den einzelnen Testungsabschnitten der in vivo-Gruppe für die Subskala Angst ($\chi^2=21,67$, $p<0,001$) und Depression ($\chi^2=15,51$, $p<0,001$) attestiert werden. Der paarweise Vergleich zeigte einen signifikanten Unterschied zwischen dem Ausgangs- und Vorbereitungswert ($p=0,001$) für die Subskala Angst, dem Vorbereitungs- und der Testungswert ($p<0,001$) für die Subskala Angst und einen signifikanten Unterschied zwischen dem Vorbereitungs- und Testungswert ($p=0,03$) für die Subskala Depression. Bei Betrachtung der Gruppe, die mit der in vivo-Testung begann, konnte ein signifikanter Unterschied gezeigt werden ($\chi^2=15,62$, $p<0,001$). Es zeigte sich ein signifikanter Unterschied zwischen dem Vorbereitungs- bzw. Testungswert und dem Ausgangswert ($p=0,001$; $p=0,01$). Für die Subskala Depression konnte kein signifikanter Unterschied gezeigt werden ($\chi^2=1,27$, $p=0,53$). Bei der Betrachtung der Zahlen sticht die Vorbereitungsphase in Bezug auf die Angstinduktion hervor (siehe Tabelle 12). Für die Testungsphase konnte kein Anstieg beobachtet werden, sodass die H3-Hypothese nur partiell bestätigt wird.

Der Vergleich zwischen der Konditionsgruppe, die mit N3 begann, und der mit der in vivo-Testung beginnenden Gruppe zeigte einen signifikanten Unterschied für die Subskala Depression ($p=0,04$). Für die Subskala Angst konnte dies nicht gezeigt werden ($p=0,68$). Bei Nichtberücksichtigung der Testreihenfolge zeigte sich kein signifikanter Unterschied zwischen den beiden Gruppen für die Subskala Angst ($p=0,14$) und Depression ($p=0,33$). Trotz der signifikant höheren Werte für die Subskala Depression der mit der N3-Kondition beginnenden Gruppe, konnte dies für die Subskala Angst nicht gezeigt werden, sodass die H4-Hypothese teilweise bestätigt wird.

Tabelle 13: STADI-S Fragebogen im Vergleich zwischen der in vivo- und N3-Gruppe

Mean (SD)	Vivo	N3
Angst		
Beginnend	50,08 (8,84)	49,75 (9,29)
Unabhäng. Reihenfolge	50,46 (9,02)	46,83 (10,56)
Depression		
Beginnend	44,33 (11,48)	58,00 (10,10)
Unabhäng. Reihenfolge	47,50 (11,50)	50,13 (11,82)

3.5.3 Situativer sozialer Angstfragebogen

Der Vergleich zwischen den einzelnen Konditionen der VR-Testung zeigte keinen signifikanten Unterschied ($\chi^2=1,36$, $p=0,51$). Auch bei der Betrachtung der Komponenten des Fragebogens zeigte sich kein signifikanter Unterschied (Fragen zur Rede: $\chi^2=3,65$, $p=0,16$; Symptome: $\chi^2=2,64$, $p=0,27$; Befürchtungen: $\chi^2=0,72$, $p=0,7$). Es konnte beim Vergleich der mit der jeweiligen Kondition beginnenden Gruppe, mit Ausnahme der Subskala Befürchtungen ($p=0,03$), kein signifikanter Unterschied festgestellt werden (Gesamtsumme: $p=0,45$; Fragen zur Rede: $p=0,90$; Symptome: $p=0,69$). Für die Subskala Befürchtungen stellten sich beim paarweisen Vergleich keine signifikanten Unterschiede heraus. Die H2-Hypothese wird durch die oben genannten Untersuchungen somit nicht gestützt.

Es konnte kein signifikanter Unterschied zwischen der mit N3 beginnenden Konditionsgruppe und der in vivo-Gruppe aufgezeigt werden ($p=0,68$). Zudem konnte bei Betrachtung der beiden Gruppen, ohne Berücksichtigung der Reihenfolge, nur für die Befürchtungen ein signifikanter Unterschied ($p=0,04$) festgestellt werden. Die anderen Komponenten unterschieden sich nicht signifikant (Gesamtsumme: $p=0,06$; Fragen zur Rede: $p=0,36$; Symptome: $p=0,14$). Die in vivo-Exposition erzeugte vergleichsweise eine stärkere Reaktion als die VR-Exposition (siehe Tabelle 14). Somit ist die H4-Hypothese zu falsifizieren.

Tabelle 14: SSA für die gesamte Testungspopulation

Mean (SD)	Vivo	N3	N6	N15
Gesamtsumme	28,00 (14,78)	21,42 (17,61)	21,38 (18,04)	18,50 (14,73)
Fragen zur Rede	8,04 (4,69)	7,25 (5,30)	7,71 (5,35)	6,42 (4,47)
Symptome	9,33 (6,44)	7,08 (7,35)	7,71 (7,74)	6,38 (6,25)
Befürchtungen	10,63 (6,34)	7,08 (6,88)	5,96 (6,17)	5,71 (5,74)

3.5.4 Graphisch numerischer Fragebogen

Es zeigte sich kein signifikanter Unterschied zwischen den Konditionen der VR-Testung ($\chi^2=1,56$, $p=0,46$). Auch bei der separaten Betrachtung der freien Rede und des arithmetischen Teils des TSSTs konnte kein signifikanter Unterschied zwischen den einzelnen Konditionen nachgewiesen werden (Freie Rede: $\chi^2=0,84$, $p=0,66$; Arithmetischer Teil: $\chi^2=0,17$,

$p=0,92$). Beim Vergleich zwischen den mit der jeweiligen VR-Kondition beginnenden Gruppen zeigte sich kein signifikanter Unterschied (Gesamtsumme: $p=0,27$; Freie Rede: $p=0,41$; Arithmetischer Teil: $p=0,15$). Resümierend konnten zwischen den VR-Konditionen keine signifikanten Unterschiede gefunden werden. Auch die absoluten Werte unterscheiden sich zwischen den Konditionen nur minimal (siehe Tabelle 15), sodass die H2-Hypothese abzulehnen ist.

Beim Vergleich der mit N3 beginnenden Konditionsgruppe und der mit in vivo beginnenden Gruppe konnte kein signifikanter Unterschied attestiert werden (Gesamtsumme: $p=0,86$; Freie Rede: $0,86$; Arithmetischer Teil: $0,52$). Bei der Betrachtung der gesamten N3-Gruppe im Vergleich zu der in vivo-Gruppe zeigte der Wilcoxon Test einen signifikanten Unterschied für die Gesamtsumme ($p=0,03$) und für die freie Rede ($p=0,005$) auf. Der Unterschied für den arithmetischen Teil war nicht signifikant ($p=0,17$). Die in vivo-Exposition erzeugte eine stärkere Reaktion bei den Probanden/innen (siehe Tabelle 15), sodass die H4-Hypothese durch die Untersuchungen nicht untermauert wird.

Tabelle 15: Graphischer Fragebogen für die gesamte Testungspopulation

Mean (SD)	Vivo	N3	N6	N15
Gesamtsumme	41,92 (21,66)	31,04 (24,70)	32,08 (23,20)	32,17 (22,52)
Freie Rede	21,75 (11,28)	14,25 (11,77)	15,46 (11,81)	14,42 (11,17)
Arithm. Teil	20,17 (11,97)	16,79 (14,05)	16,63 (12,72)	17,75 (13,38)

3.5.5 Co- and Social-Presence Questionnaire

Die einzelnen Kategorien des Fragebogens wiesen zwischen den einzelnen VR-Konditionen keinen signifikanten Unterschied auf (1. Kategorie: $\chi^2=1,1$, $p=0,58$; 2. Kategorie: $\chi^2=0,09$, $p=0,96$; 3. Kategorie: $\chi^2=2,51$, $p=0,29$; 4. Kategorie: $\chi^2=0,9$, $p=0,64$).

Zwischen den jeweils mit den einzelnen VR-Konditionen beginnenden Gruppen konnte für die erste Kategorie ein signifikanter Unterschied gezeigt werden ($p=0,02$). Für die anderen Kategorien konnte dies nicht gezeigt werden (2. Kategorie: $p=0,8$; 3. Kategorie: $p=0,16$; 4. Kategorie: $p=0,38$). Der paarweise Vergleich wies für die erste Kategorie keine signifikanten Unterschiede nach.

Bei der Betrachtung der Subskalen der mit der jeweiligen Kondition beginnenden Gruppe konnte die N3- und die N15-Kondition im Vergleich zu der N6-Kondition höhere Werte aufweisen (siehe Tabelle 16). Ein Zusammenhang zwischen der Anzahl der virtuellen Avatare und dem Realitätsempfinden der Probanden/innen lässt sich daraus aber nicht ableiten.

Tabelle 16: Co- and Social-Presence Questionnaire für die VR-Konditionen

Mean (SD)	N3	N6	N15
Reaktion d. Redners auf die VA	3,31 (1,34)	1,65 (0,60)	3,42 (0,52)
Wahrgenommene Reaktion d. VA	2,31 (0,69)	2,00 (0,75)	2,08 (1,01)
Eindruck d. Interaktionsmöglichkeiten	2,63 (0,97)	1,40 (0,52)	2,50 (1,39)
Ko-Präsenz von anderen Personen	2,67 (1,19)	3,20 (0,87)	3,78 (0,51)

3.5.6 Igroup Presence Questionnaire

Der Igroup Presence Questionnaires wurde nur einmalig nach der Testung erhoben. Insbesondere der Wert für die Realität war im Vergleich zu den anderen Werten deutlich niedriger (siehe Tabelle 17).

Tabelle 17: Igroup Presence Questionnaire

	M (SD)
Eindruck, da gewesen zu sein	3,21 (1,89)
Räumliche Präsenz	15,88 (7,74)
Involviertheit	10,88 (3,67)
Realität	7,08 (4,18)
Gesamtsumme	37,04 (15,93)

3.5.7 Simulator Sickness Questionnaire

Der Wilcoxon Test zeigte zwischen den Prä- und Post-Werten des SSQs einen signifikanten Unterschied für die Gesamtsumme ($p=0,03$). Die Unterschiede bezüglich der Übelkeit ($p=0,1$), der okulomotorischen Symptome ($p=0,06$) und der Desorientierung ($p=0,11$) waren nicht signifikant.

Tabelle 18: Vergleich zwischen Prä- und Post-Werten des SSQ

Mean (SD)	Prä-VR	Post-VR
Gesamtsumme	11,13 (12,15)	21,19 (18,71)
Übelkeit	11,13 (12,15)	20,27 (23,92)
Okkulomotor. Symptome	12,95 (15,72)	20,21 (14,94)
Desorientierung	6,38 (17,88)	12,18 (15,53)

4 Diskussion

Das Ziel unserer Cross-over Studie war es, die virtuelle Exposition mit der in vivo-Exposition hinsichtlich der Auslösung von Angst zu vergleichen. Insgesamt wurden 24 Probanden/innen im Alter zwischen 21 und 50 Jahren, wovon 75% weiblich waren, im Rahmen unserer Studie getestet. Jeweils die Hälfte der Probanden begann mit der in vivo-Testung bzw. VR-Testung. Anschließend erfolgte ein Wechsel in die andere Expositionsgruppe. Diverse physiologische Parameter und Fragebögen wurden im Zuge der Testung erhoben. Im folgenden Abschnitt werden zunächst die methodische Vorgehensweise und anschließend die Ergebnisse diskutiert. Am Ende der Diskussion wird ein Ausblick für weitere Forschungsmöglichkeiten skizziert.

4.1 Methoden

4.1.1 Aufbau des Experiments

Die Testung teilte sich in zwei Teile auf: Die in vivo-Testung und die VR-Testung. Die in vivo-Testung entsprach weitestgehend dem TSST. Der einzige Unterschied bestand darin, dass die Probanden/innen – zum Zwecke der Vergleichbarkeit mit der VR-Testung – während der Testung saßen. Die VR-Testung bestand aus drei Untertestungen mit jeweils drei, sechs und 15 virtuellen Avataren. Aufgrund der niedrigen Fallzahl ($n=24$) begannen jeweils 12 Probanden/innen mit der in vivo- bzw. VR-Testung. Bei detaillierter Betrachtung der VR-Testung ergaben sich insgesamt sechs Kombinationsmöglichkeiten (Pfade) für die VR-Testung, sodass pro Pfad nur vier Probanden getestet worden sind. Grundsätzlich haben geringere Fallzahlen Konsequenzen für die Ergebnisse der Testung nach Lageunterschieden [38]. Entweder ist der erwartete Unterschied zwischen den Gruppen groß und es kann auf eine große Stichprobengröße verzichtet werden oder der Unterschied ist vergleichsweise gering und es besteht die Notwendigkeit, eine große Stichprobengröße anzustreben. Die Mehrheit unserer Ergebnisse unterschied sich nicht signifikant. Dies könnte an dem Aufbau des Experimentes und der kleinen Stichprobengröße liegen. Insbesondere in Anbetracht der erwartungsgemäß [39] eher geringen Unterschiede mit einer Tendenz zu niedrigeren Werten für die VR-Exposition liegt die angeführte Erklärung nahe.

4.1.2 Probanden/innen

Für die Testung im Rahmen der Studie waren 20 bis 30 Probanden/innen angestrebt. Für die Teilnahme der Studie wurde vorausgesetzt, dass die Teilnehmer/innen keine psychiatrischen Erkrankungen und trotz dessen eine Neigung zu sozialer Phobie aufweisen. Die Kombination der beiden Voraussetzungen führte dazu, dass die Neigung zu sozialer Phobie, die mittels SPDQ-Fragebogens erfasst wurde, niedrig angesetzt worden ist, um ausreichend Probanden/innen rekrutieren zu können. Im Vergleich zu dem von Newman et al. empfohlenen

Cut-off-Wert von 7,38 zur Detektion von sozialer Phobie lag unser Einschlusskriterium bei ≥ 1 [24]. Unsere Probanden/innen wiesen einen Durchschnittswert von 4,9 (SD: 3,3) auf. In Studien konnte gezeigt werden, dass Patienten/Patientinnen mit sozialer Phobie im Vergleich zur Kontrollgruppe keine signifikant erhöhten physiologischen Parameter nach dem TSST aufwiesen [34, 40]. Im Gegensatz dazu zeigten andere Studien wiederum Unterschiede zwischen den beiden Gruppen [41, 42]. Insgesamt ist das Bild hinsichtlich der Stressreaktion in Relation zu der Diagnose soziale Phobie inkonsistent.

Unsere Studienpopulation war im Durchschnitt 28,2 Jahre (SD: 7,5) alt und überwiegend weiblich (75%). Eine Metaanalyse von Helminen et al. untersuchte den Einfluss vom Alter und Geschlecht auf die Cortisolkonzentration [43]. Für das Alter zeigte sich ein signifikanter Unterschied zwischen unter und über 25-Jährigen. Die Effektstärke betrug für die Gruppe unter 25 Jahren 0,83 (SE: 0,1) und für die älter als 25 Jahren Gruppe 0,47 (SE: 0,1). Zudem zeigte sich, dass die Effektstärke bei Testungspopulationen, die ausschließlich männlich waren (ES: 0,96, SE: 0,09), sich signifikant von gemischten oder ausschließlich weiblichen Testungspopulationen unterschied (ES: 0,54, SE: 0,08). Da unsere Studie im Schnitt etwas älter und der Anteil an weiblichen Probandinnen relativ hoch war, könnte die Cortisolmessung dadurch beeinflusst worden sein.

4.1.3 Technisches Equipment und digitale Visualisierung des TSSTs

Für die Testung wurde die VR-Brille *Vive* von dem Hersteller HTC verwendet. Zu dem damaligen Zeitpunkt waren die HTC Vive und Oculus Rift die bekanntesten für den Verbrauchermarkt vorgesehenen VR-Brillen. Die HTC Vive bot im Vergleich zu der Oculus Rift ein etwas größeres Sichtfeld (Vive: 110°, Oculus Rift 100°). Unsere Probanden/innen gaben während der Testung an, dass das Bild ein Pixelraster aufwies und nicht scharf sei. Dieses Phänomen, Fliegengitter-Effekt, ist ein bekanntes Problem von VR-Brillen und entsteht durch eine zu geringe Pixeldichte der Bildschirme, sodass das menschliche Auge Pixelraster wahrnimmt. Inzwischen sind neuere Modelle auf dem Markt, die eine höhere Pixeldichte als die von uns genutzte Brille bieten [44]. In Anbetracht der damaligen technischen Möglichkeiten war die Wahl auf die HTC Vive Brille naheliegend. In der Metaanalyse von Helminen et al. wurden immersive und nicht-immersive virtuelle Lösungen des TSSTs miteinander verglichen. Es konnte ein signifikanter Unterschied zwischen den beiden Gruppen gezeigt werden, wobei die Effektstärke der immersiven Lösungen (ES: 0,83, SE: 0,07) höher war als die der nicht-immersiven virtuellen Lösungen (ES: 0,46, SE: 0,07) [43]. In Anbetracht dessen wären weiterführende Studien mit neueren VR- Brillen, die eine bessere Bildqualität aufweisen, erstrebenswert, da dadurch das Immersions- und Realitätsempfinden gesteigert werden könnten.

4.2 Ergebnisse

4.2.1 Cortisol

Für alle Konditionen der VR-Testung zeigte sich ein signifikanter Unterschied zwischen dem Vorbereitungs- bzw. Testungswert und dem Ruhewert. Eine Ausnahme stellte die N6-Kondition dar, die nur einen signifikanten Unterschied zwischen dem Vorbereitungs- und Ruhewert aufwies. Beispielsweise sank die Cortisolkonzentration nach der Testung signifikant von 2,24 ng/ml (SD: 1,66) auf 1,47 ng/ml (SD: 0,86). Da der Ausgangswert mit 2,93 ng/ml (SD: 1,54) relativ hoch war, ist der Abfall kritisch zu betrachten. Im Vergleich zu den in vivo-Werten, die nur einen signifikanten Unterschied zwischen dem Testungs- und Ruhewert aufwiesen, sind die Werte der VR-Testungen vergleichsweise deutlich geringer (vivo: 5,03 ng/ml [SD: 3,01]; N3: 2,24 ng/ml [SD 1,66]). Ein signifikanter Unterschied zwischen den einzelnen Konditionen der VR-Testung konnte nicht gezeigt werden.

Im Gegensatz zu unseren Ergebnissen konnten andere Studien auch für die VR-Exposition einen deutlichen höheren Anstieg des Cortisols feststellen [39, 45]. Hierbei verfolgte insbesondere die Studie von Shibani et al. einen interessanten Ansatz, indem sie eine Aufteilung in Responder und Non-Responder durchführten. Analog dazu wurde in unserer Studie eine Aufteilung für die in vivo-Testung durchgeführt. Es konnte gezeigt werden, dass die Hälfte unserer Probanden/innen keinen Anstieg bzw. eher einen leichten Abfall während des gesamten Monitorings aufwiesen. Leider war eine Übertragung des methodischen Vorgehens auf die VR-Testung nicht möglich, da keine eindeutige Zuordnung in Responder und Non-Responder möglich war. Darüber hinaus war ein hoher Ausgangswert für beide Testungsgruppen auffällig. Eine mögliche Ursache kann die zeitliche Nähe zwischen dem Ausgangswert und dem Wert der Vorbereitungsphase sein. Die Zeitspanne betrug in etwa 10-15 Minuten. In vergleichbaren Studien betrug die Spanne 20 bzw. 30 Minuten [39, 45]. Eine weitere Schwäche unserer Studie war eine nicht ausreichend lange Nachbeobachtungsphase. Bezüglich des Cortisols konnte eine Angleichung des Post-Test-Wertes auf das Niveau des Prä-Test-Wertes nicht gemessen werden. Aufgrund der kurzen Nachbeobachtungsdauer konnte ein Gefälle nur in Ansätzen beobachtet werden.

4.2.2 Hautleitwiderstand (GSR)

Für den Hautleitwiderstand zeigte sich ein signifikanter Unterschied zwischen dem Vorbereitungs- bzw. Testungswert und dem Ausgangswert für alle Konditionen. Darüber hinaus unterschied sich der Testungs- vom Ruhewert. Der Hautleitwiderstand für die N3-Konditionsgruppe stieg signifikant von 0,3 μ S (SD: 0,15) auf 0,55 μ S (SD: 0,18) und fiel in der Ruhephase auf 0,38 μ S (SD: 0,14). Bei der in vivo-Gruppe unterschieden sich der Vorbereitungs- bzw. Testungswert vom Ausgangswert und der Testungswert vom Ruhewert signifikant. Die Reaktion war jedoch im Vergleich zu der VR-Testung ausgeprägter. Der Hautleitwiderstand stieg signifikant von 0,35 μ S (SD: 0,2) auf 0,73 μ S (SD: 0,13). Nach der Testung

fiel der Wert wieder auf $0,4 \mu\text{S}$ (SD: 0,2). Einen Unterschied zwischen den VR-Konditionen konnte diese Studie nicht zeigen. Eine Studie von Zimmer et al. hingegen zeigte einen signifikanten Anstieg des Hautleitwiderstandes im Zeitverlauf des VR-TSSTs, jedoch wurden in der Publikation keine Werte erwähnt [39]. Abgesehen von dieser Studie sind Publikationen, die die Veränderung des Hautleitwiderstandes im virtuellen TSST untersuchen eine Seltenheit. Unsere Werte zeigten einen klaren signifikanten Anstieg und Abfall des Hautleitwiderstandes im Verlauf der Testung. Wie im Abschnitt zuvor erwähnt, könnte bei einer längeren Nachbeobachtungsdauer eine komplette Normalisierung des Hautleitwiderstandes beobachtet werden.

4.2.3 Herzfrequenz

Bei der Betrachtung der Herzfrequenz konnte für die VR-Gruppe ein signifikanter Unterschied zwischen dem Ausgangs-, Vorbereitungs- bzw. Testungswert und dem Ruhewert gezeigt werden. Einen signifikanten Unterschied zwischen dem Testungswert und dem Ausgangswert gab es jedoch nicht. Die durchschnittliche Herzfrequenz war zu Beginn mit 79,43 bpm (SD: 8,61) relativ hoch, sank dann im Testungsverlauf auf 77,93 bpm (SD: 9,49) und lag anschließend in der Ruhephase bei 73,44 bpm (SD: 8,75). Ein klarer Anstieg und Abfall ließen sich aus den Werten nicht ableiten. Ein möglicher Erklärungsversuch für den hohen Ausgangswert wurde zuvor im Abschnitt Cortisol erörtert. Als mögliche Ursache wäre die kurze Zeitspanne zwischen dem Ausgangswert und dem Vorbereitungswert zu vermuten. Da es sich bei der Herzfrequenz jedoch um eine kontinuierliche Messung handelte, müsste dieses Problem – den Ergebnissen der Hautwiderstandsmessung entsprechend – vernachlässigt werden können. Bei der *in vivo*-Testung zeigte sich jedoch ein anderes Bild: Die Herzfrequenz stieg von ursprünglichen 83,51 bpm (SD: 10,82) auf 93,99 bpm (SD: 12,6) und fiel anschließend auf 78,05 bpm (SD: 8,7). Zudem unterschied sich hier der Testungswert signifikant vom Ausgangs- und Ruhewert. Bezüglich der Herzfrequenz zeigte die Studie von Zimmer et al. analog zu unserer Studie einen signifikanten Unterschied zwischen der VR- und *in vivo*-Exposition mit der Tendenz zu höheren Werten für die *in vivo*-Testung. Im Kontrast hierzu gab die Studie von Shiban et al. [45] vergleichbare Werte für beide Expositionsformen an. Beide Studien wiesen im Gegensatz zu unserer einen signifikanten Anstieg vom Ausgangswert zum Testungswert sowohl für die VR-Exposition als auch *in vivo*-Exposition auf.

4.2.4 Herzfrequenzvariabilität

In diesem Abschnitt werden beide Zeit-Domänen Parameter, RMSSD und SDNN, diskutiert. Bei der Betrachtung der gesamten Gruppe zeigte sich für den RMSSD für alle VR-Konditionen ein signifikanter Unterschied zwischen dem Ausgangs- und Ruhewert. Für die N3- und N15-Kondition konnte zudem ein signifikanter Unterschied zwischen dem Vorbereitungs- und Ruhewert gezeigt werden. Darüber hinaus stellte sich ein signifikanter Unterschied zwischen dem Testungs- und Ausgangswert für die N15-Kondition heraus. Für die N3-Kondition stieg der RMSSD von 42,77 ms (SD: 21,85) auf 49,98 ms (SD: 26,95) und

anschließend auf 55,73 ms (SD: 38,77). Für den SDNN zeigten sich signifikante Unterschiede zwischen dem Ausgangs-, Vorbereitungs- bzw. Testungswert und dem Ruhewert für alle Konditionen. Für die N3-Kondition stieg der SDNN von 65,32 ms (SD: 29,93) auf 74,01 ms (SD: 26,98) und anschließend auf 83,17 ms (SD: 32,48). Für die in vivo-Gruppe zeigte sich für den RMSSD und SDNN ein signifikanter Unterschied zwischen dem Ausgangs- bzw. Testungswert und dem Ruhewert. Zudem zeigte sich für den RMSSD ein signifikanter Unterschied zwischen dem Vorbereitungs- und Ruhewert. Der RMSSD stieg von 40,82 ms (SD: 31,19) auf 42,2 ms (SD: 26,49) und abschließend auf 57,07 ms (SD: 35,86). Für den SDNN zeigte sich ein anderes Bild: Der Ausgangswert betrug 74,13 ms (SD: 41,39) und fiel auf 67,64 ms (SD: 17,64) und stieg anschließend auf 87,83 ms (SD: 37,42). Ein Einfluss der virtuellen Avatare zeigte sich weder für den RMSSD noch für den SDNN. Nur die SDNN-Werte der in vivo-Testung zeigten den erwarteten Verlauf in Form einer Absenkung während der Testung und einer Re-Normalisierung in der Ruhephase. Andere Studien, die die HRV im Rahmen des TSSTs erhoben haben, zeigten diesen beschriebenen Verlauf [46, 47]. Die Studie von Hermann et al. verglich dabei den TSST mit einem physischen Training in Form eines HIT-Trainings und konnte auch für den TSST eine Senkung des RMSSDs feststellen, auch wenn diese nicht so ausgeprägt war wie die des physischen Trainings [47]. Mögliche Ursachen für den fehlenden Stress der Probanden/innen hinsichtlich der HRV-Parameter sind schwierig zu eruieren. Für die VR-Testung zeigten weder die RMSSD- noch die SDNN-Werte einen nachvollziehbaren Verlauf. Darüber hinaus sind Studien hinsichtlich der Herzfrequenzvariabilität während des Trier Social Stress Tests rar.

4.2.5 Fragebögen

Beim SAM-Fragebogen konnten nur punktuell signifikante Unterschiede festgestellt werden. Für die Valenz konnten für jede VR-Kondition signifikante Unterschiede zwischen dem Ausgangs- und dem Vorbereitungs- bzw. Testungswert gezeigt werden. Für die in vivo-Testung zeigte sich nur ein signifikanter Unterschied zwischen dem Ausgangs- und dem Vorbereitungswert für die Valenz und für das Arousal. Somit war die Gemütslage der Probanden/innen nach der VR- und in vivo-Testung etwas negativer und in der in vivo-Vorbereitungsphase waren die Teilnehmer/innen aufgeregter. Die absoluten Unterschiede sind jedoch – mit Ausnahme der in vivo-Vorbereitungsphase – minimal, sodass insgesamt von einer subjektiv schwach empfundenen Stressreaktion ausgegangen werden kann. Beim STADI-S führt sich dieses Bild fort. In den VR-Konditionen zeigte sich ein leichter Anstieg des Angstwertes in der Vorbereitungsphase, der jedoch nur für die N6-Kondition signifikant war. Die Vorbereitungsphase der in vivo-Testung sticht hierbei wieder auf und weist mit 62,16 (SD: 10,36) einen signifikanten Unterschied zum Ausgangswert 53,71 (SD: 9,3) auf. Dieser Wert wäre als einziger als erhöht/überdurchschnittlich zu bezeichnen [25]. Die Werte für die Testung waren durchgängig auf einem vergleichbaren Niveau oder teilweise sogar niedriger als der Ausgangswert. Eine Angstinduktion lässt sich anhand der subjektiven Angaben im STADI-S nicht erkennen. Bei dem von uns entwickelten Fragebogen, dem Situative Sozialer

Angstfragebogen, zeigten sich keine signifikanten Unterschiede zwischen den VR-Konditionen. Beim Vergleich zwischen der in vivo-Gruppe mit der N3-Konditionsgruppe stellte sich nur für die Subskala Befürchtungen ein signifikanter Unterschied heraus. Der Mittelwert der in vivo-Gruppe war mit 10,63 (SD: 6,34) signifikant höher als der der N3-Konditionsgruppe (M: 7,08, SD: 6,88). Zudem wies die in vivo-Gruppe eine höhere Gesamtsumme als die die N3-Konditionsgruppe auf (Vivo: 28 [SD: 14,78], N3: 21,42 [SD: 17,61]). In Anbetracht der Maximalpunktzahl von 116 sprechen die Ergebnisse für eine tendenziell niedrige Angstinduktion bei den Probanden/innen oder für eine nicht ausreichende Validität des Fragebogens. Weitere Studien zur Untersuchung der Validität und Reliabilität des Fragebogens wären erstrebenswert. Für den graphischen Fragebogen zeigte die in vivo-Gruppe höhere Werte als die VR-Gruppe. Zwischen der in vivo-Gruppe und der N3-Konditionsgruppe gibt es für die Gesamtsumme und für die Subskala Freie Rede einen signifikanten Unterschied. Hierbei betragen die Werte für die in vivo-Gruppe 41,92 (SD: 21,66) bzw. 21,75 (SD: 11,28) und für die N3-Konditionsgruppe 31,01 (SD: 24,7) bzw. 14,25 (SD: 11,77).

Der Igroup Presence Questionnaire wurde einmalig nach drei konsekutiven VR-Testungen erhoben. Im Vergleich zu zwei anderen Studien sind die Werte in unserer Studie niedriger. Insbesondere der Wert für die Realität ist im Vergleich (15,17 [48] bzw. 13,97 [49]) mit 7,08 als deutlich niedriger einzustufen. Eine mögliche Ursache für den niedrigen Wert könnte in unserer technischen Umsetzung der virtuellen Umgebung liegen.

Bei Betrachtung der Gesamtsumme des Simulator Sickness Questionnaires liegt die Einschätzung nahe, dass es sich hierbei um eine „problematische VR-Simulation“ handle, da die Gesamtsumme mit 21,19 (SD: 18,71) leicht über dem Wert 20 liegt [50]. Jedoch betrug der Ausgangswert der Probanden/innen im Durchschnitt 12,47 (SD 15,43). Bei genauerer Betrachtung und Bildung der Differenz zwischen dem Post- und Prä-Wert ordnete sich der Wert mit 8,72 in die Kategorie „minimale Symptome“ ein. Dies entspräche auch den Ansichten der Probanden/innen, die nach der Testung keine starken Beschwerden aufwiesen und nur über leichte Symptome berichteten.

Zusammenfassend zeigen die Ergebnisse eine eher schwache Stressinduktion durch die VR- bzw. in vivo-Testung. Sowohl bei dem von uns entwickelten SSA als auch bei etablierten Fragebögen, wie z.B. dem SAM [51], konnten nur schwache Reaktionen erfasst werden, wobei die in vivo-Testung in der Regel höhere Werte aufzuweisen hatte. Ein Einfluss der virtuellen Avatare auf die Stressinduktion zeigte sich in keinem Fragebogen.

4.3 Ausblick für weitere Forschungen

Die Ergebnisse unserer Studie geben einen interessanten Einblick in das neue Feld der virtuellen Technologien und machen deutlich, dass weitere Forschungsarbeiten in diesem Themenkomplex vonnöten sind, um Erkenntnisse zu prüfen und Zusammenhänge zu erschließen. In unserer Studie führte die VR-Testung in der Regel zu einer vergleichsweise deutlich

schwächeren Reaktion als die in vivo-Testung. Mögliche Ursachen, wie z.B. die technischen Limitierungen, können mit neueren VR-Lösungen weitergehend untersucht werden. Zudem zeigte sich kein Einfluss der virtuellen Avatare auf die physiologischen Parameter und subjektiven Angaben in Form von Fragebögen. Somit stützen unsere Ergebnisse die Social-Impact-Theorie nicht. Tendenziell zeigte die N3-Kondition die stärkste, jedoch sich nicht signifikant von den anderen Konditionen unterscheidende, Reaktion auf die Stresssituation. Eine mögliche Ursache könnte abgesehen von technischen Limitierungen in der Art und Weise der Programmierung der virtuellen Avatare liegen. Die Verhaltensweise der Avatare entsprach nicht dem von realen Menschen. Sowohl die Mimik und Gestik als auch die Bewegungsabläufe waren nicht natürlich. Ein stärkerer Fokus auf die Darstellungsqualität und den Ablauf der Bewegungen könnte das Immersionsgefühl und die Realitätsnähe von virtuellen Umgebungen deutlich steigern. Darüber hinaus stellt sich die Frage der grundsätzlichen Vereinbarkeit der Social-Impact-Theorie mit der virtuellen Realität. Eine Metaanalyse zur Protokollvariation des in vivo-TSSTs zeigte, dass drei Jurymitglieder hinsichtlich der Cortisolantwort eine höhere Effektstärke aufwiesen als zwei Jurymitglieder, wobei das untere Konfidenzintervall der Jury aus drei Mitgliedern die Effektstärke der Jury mit zwei Mitgliedern mit einschloss [52]. Ob diese Beobachtung auch auf die virtuelle Realität und auf sechs, 15 oder eine variable Anzahl an virtuellen Avataren zutrifft, ließe sich in einer weiteren Studie untersuchen. Es wäre möglich, dass die von uns postulierte Hypothese, dass eine Korrelation zwischen der variablen Anzahl an Avataren und der induzierten Angst besteht, nicht auf die virtuelle Realität übertragbar ist.

Abgesehen von Überlegungen zu einer variablen Anzahl von digitalen Avataren könnten auch Überlegungen zu multiplen Testpersonen innerhalb der virtuellen Umgebung aufgegriffen werden. Studien zeigten bereits eine Effektivität des TSSTs für Gruppen (TSST-G) [53, 54]. Analog dazu könnten Gruppenexpositionen in virtueller Realität auf ihre Effektivität hin untersucht werden.

5 Zusammenfassung

Die Angststörungen weisen die höchste 12-Monats-Prävalenz unter den psychischen Störungen auf, wobei die soziale Phobie mit einer 12-Monats-Prävalenz von 2,7% die dritthäufigste Angststörung ist. Zur Therapie sozialer Phobie stehen im Rahmen von verhaltenstherapeutischen Ansätzen Expositionstherapien zur Verfügung. Die Vorteile einer virtuellen Expositionstherapie liegen in der Durchführung von standardisierten und aufeinander aufbauenden Situationen, der Skalierbarkeit, der Kostenersparnis und den Monitoringmöglichkeiten.

Das Ziel unserer Studie war es, die Auswirkungen des standardisierten Verfahrens in Form des Trier Social Stress Tests in virtueller Realität und in vivo zu untersuchen und miteinander zu vergleichen. Darüber hinaus wurde die Hypothese untersucht, ob in einer virtuellen, sozialen Situation mehrere Avatare eine stärkere Angstinduktion im Individuum hervorrufen.

Hierfür wurden 24 Probanden/innen im Alter zwischen 21-50 Jahren getestet. Die Testung bestand aus zwei Teilen: Der Trier Social Stress Test zum einen in virtueller Realität und zum anderen in vivo. Jede/r Proband/in durchlief beide Testungen. Im Rahmen der Testung wurden diverse physiologische Parameter und Fragebögen erhoben. Zu den physiologischen Parametern zählten Cortisol, der Hautleitwiderstand, die Herzfrequenz und die Herzfrequenzvariabilität. Der Schwerpunkt der Fragebögen lag neben der Erfassung der empfundenen Angst in der Beurteilung des Realitäts- und Immersionsempfindens in der virtuellen Realität.

In unserer Studie konnte gezeigt werden, dass die virtuelle Testung im Vergleich zur in vivo-Testung schwächere Reaktionen aufwies. Beispielsweise stieg der Hautleitwiderstand in der virtuellen Realitätsgruppe während der Testung signifikant von $0,3 \mu\text{S}$ (SD: 0,15) auf $0,55 \mu\text{S}$ (SD: 0,18). Zum Vergleich stieg der Hautleitwiderstand in der in vivo-Gruppe signifikant von $0,35 \mu\text{S}$ (SD: 0,2) auf $0,73 \mu\text{S}$ (SD: 0,13), wobei sich der Testungswert signifikant von dem der VR-Gruppe unterschied. Für die anderen physiologischen Parameter konnte ein Anstieg für die virtuelle Realität nur begrenzt gezeigt werden. In den Ergebnissen der Fragebögen zeigte sich eine relativ geringe Angstinduktion mit der Tendenz zu höheren Werten in der in vivo-Testung. Der empfundene Angstwert für die gesamte Testung war mit 41,92 (SD: 21,66) von maximal 100 Punkten in der in vivo-Gruppe signifikant höher als der Wert der virtuellen Realitätsgruppe mit 31,04 (SD: 24,7). Das Realitätsempfinden in der virtuellen Umgebung war im Vergleich zu anderen Studien eher als niedrig einzustufen. Zudem konnte ein Einfluss der Anzahl der virtuellen Avatare auf physiologische Parameter und Fragebögen nicht nachgewiesen werden.

Das Forschungsfeld der virtuellen Realität ist ein neues und interessantes Gebiet. Die Ergebnisse unserer Studie sprechen für weiterführende Studien, die den Einfluss der virtuellen Realität, Personen und Objekten auf den menschlichen Körper tiefergehend untersuchen.

6 Zusammenfassung auf Englisch

Anxiety disorders have the highest 12-month prevalence among mental disorders in Germany. The third most common anxiety disorder is the social anxiety with a 12-month prevalence of 2.7%. Behavioral treatments of social anxiety disorder include exposure therapies. The benefits of virtual exposure therapy are the implementation of standardized situations that build on one another, scalability, saving of expenses and monitoring options.

The goal of our study was to examine the effects of the standardized procedure in form of the Trier Social Stress Test in virtual reality as well as in vivo and compare them. Furthermore, the hypothesis, whether a higher number of avatars induce greater fear in individuals in a virtual, social situation, was investigated.

For this purpose, 24 participants between the age of 21-50 years were tested. The testing consisted of two parts: The Trier Social Stress Test in vivo and, on the other hand, the same test in virtual reality. Every participant performed both parts. Various physiological parameters and questionnaires were collected during the testing. The physiological parameters included cortisol, skin resistance, heart rate and heart rate variability. The focus of the questionnaires was on recording the perceived fear as well as evaluating the perception of reality and the immersion within the virtual environment.

The results of our study showed that the virtual testing induced weaker reactions compared to the in vivo testing. For example, the skin resistance of the virtual reality group increased significantly from 0.3 μ S (SD: 0.15) to 0.55 μ S (SD: 0.18) during the test. For comparison, the skin resistance of the in vivo group increased significantly from 0.35 μ S (SD: 0.2) to 0.73 μ S (SD: 0.13). The test results of these two groups were significantly different from each other. Concerning the other physiological parameters, an increase in virtual reality could only be shown to a limited extent. The results of the questionnaires showed a relatively low level of anxiety induction with the tendency for higher results for the in vivo testing. The perceived fear for the entire test was 41.92 (SD: 21.66) out of 100 points for the in vivo group which was significantly higher compared to the virtual reality group (31.04 (SD: 24.7)). The sense of reality in the virtual environment was relatively low compared to other studies. In addition, it could be shown that the number of virtual avatars had no influence on physiological parameters and questionnaires.

The virtual reality in medicine is a new and interesting research field. The results of our study are underlining the need of further studies that examine the influence of virtual reality, people and objects on the human body.

7 Literaturverzeichnis

1. Jacobi, F., et al., *Twelve-month prevalence, comorbidity and correlates of mental disorders in Germany: the Mental Health Module of the German Health Interview and Examination Survey for Adults (DEGS1-MH)*. International journal of methods in psychiatric research, 2014. **23**(3): p. 304-319.
2. Koch-Institut, R., *Gesundheit in Deutschland*. 2015: p. Kapitel: Angststörungen S. 113, Arbeitsunfähigkeit S. 158.
3. Hans-Ulrich Wittchen, F.J., *Gesundheitsberichterstattung des Bundes, Heft 21 Angststörungen*. Robert Koch Institut, 2004.
4. Dilling, H. and H.J. Freyberger, *Taschenführer zur ICD-10-Klassifikation psychischer Störungen*. Bern: Huber, 2012: p. 160-161.
5. Bandelow, B., *Panik und Agoraphobie: Diagnose, Ursachen, Behandlung*. 2001: Springer-Verlag.
6. Liebowitz, M.R., *Social phobia*. Modern problems of pharmacopsychiatry, 1987. **22**: p. 141-173.
7. Rytwinski, N.K., et al., *Screening for social anxiety disorder with the self-report version of the Liebowitz Social Anxiety Scale*. Depression and Anxiety, 2009. **26**(1): p. 34-38.
8. Bandelow, B., et al., *Deutsche S3-Leitlinie Behandlung von Angststörungen*. 2014.
9. Acarturk, C., et al., *Psychological treatment of social anxiety disorder: a meta-analysis*. Psychol Med, 2009. **39**(2): p. 241-54.
10. Hedges, D.W., et al., *The efficacy of selective serotonin reuptake inhibitors in adult social anxiety disorder: a meta-analysis of double-blind, placebo-controlled trials*. J Psychopharmacol, 2007. **21**(1): p. 102-11.
11. Hansen, R.A., et al., *Efficacy and tolerability of second-generation antidepressants in social anxiety disorder*. International Clinical Psychopharmacology, 2008. **23**(3): p. 170-179.
12. Wlazlo, Z., *Soziale Phobie: Eine Anleitung zur Durchführung einer Exposition in vivo*. 1995: p. S. 37.
13. Gould, R.A., et al., *Cognitive-Behavioral and Pharmacological Treatment for Social Phobia: A Meta-Analysis*. Clinical Psychology: Science and Practice, 1997. **4**(4): p. 291-306.
14. Fedoroff, I.C. and S. Taylor, *Psychological and pharmacological treatments of social phobia: a meta-analysis*. J Clin Psychopharmacol, 2001. **21**(3): p. 311-24.
15. Tart, C.T., *Multiple personality, altered states and virtual reality: The world simulation process approach*. Dissociation: Progress in the Dissociative Disorders, 1990. **3**(4): p. 222-233.
16. Garcia-Palacios, A., et al., *Comparing Acceptance and Refusal Rates of Virtual Reality Exposure vs. In Vivo Exposure by Patients with Specific Phobias*. Cyberpsychology & behavior : the impact of the Internet, multimedia and virtual reality on behavior and society, 2007. **10**: p. 722-4.
17. Meyerbröker, K. and P.M. Emmelkamp, *Virtual reality exposure therapy in anxiety disorders: a systematic review of process-and-outcome studies*. Depress Anxiety, 2010. **27**(10): p. 933-44.
18. Carl, E., et al., *Virtual reality exposure therapy for anxiety and related disorders: A meta-analysis of randomized controlled trials*. J Anxiety Disord, 2019. **61**: p. 27-36.

19. Kirschbaum, C., K.M. Pirke, and D.H. Hellhammer, *The 'Trier Social Stress Test'--a tool for investigating psychobiological stress responses in a laboratory setting*. Neuropsychobiology, 1993. **28**(1-2): p. 76-81.
20. Allen, A.P., et al., *The Trier Social Stress Test: Principles and practice*. Neurobiology of stress, 2016. **6**: p. 113-126.
21. Latané, B., *The psychology of social impact*. American psychologist, 1981. **36**(4): p. 343.
22. Debono, M., et al., *Modified-release hydrocortisone to provide circadian cortisol profiles*. The Journal of clinical endocrinology and metabolism, 2009. **94**(5): p. 1548-1554.
23. Sheehan, D.V., et al., *The Mini-International Neuropsychiatric Interview (M.I.N.I.): the development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10*. J Clin Psychiatry, 1998. **59 Suppl 20**: p. 22-33;quiz 34-57.
24. Newman, M.G., et al., *The social phobia diagnostic questionnaire: preliminary validation of a new self-report diagnostic measure of social phobia*. Psychol Med, 2003. **33**(4): p. 623-35.
25. Lothar, L., et al., *Das State-trait-Angst-Depressions-Inventar : STADI ; Manual*. 2013, Hogrefe: Göttingen.
26. Russell, G. and S. Shaw, *A study to investigate the prevalence of social anxiety in a sample of higher education students in the United Kingdom*. Journal of Mental Health, 2009. **18**(3): p. 198-206.
27. *From text to speech (Sprache: German, Stimme: Michael, Geschwindigkeit: fast)*. Available from: <http://www.fromtexttospeech.com/>.
28. Bradley, M.M. and P.J. Lang, *Measuring emotion: the Self-Assessment Manikin and the Semantic Differential*. J Behav Ther Exp Psychiatry, 1994. **25**(1): p. 49-59.
29. Poeschl-Guenther, S. and N. 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. **13**: p. 58-63.
30. Schubert, T.W., *The sense of presence in virtual environments: A three-component scale measuring spatial presence, involvement, and realism*. Zeitschrift für Medienpsychologie, 2003. **15**(2): p. 69-71.
31. Kennedy, R.S., G.C. Tolhurst, and A. Graybiel, *The effects of visual deprivation on adaptation to a rotating environment*. NSAM-918. Res Rep U S Nav Sch Aviat Med, 1965: p. 1-36.
32. Kennedy, R.S., et al., *Simulator Sickness Questionnaire: An Enhanced Method for Quantifying Simulator Sickness*. The International Journal of Aviation Psychology, 1993. **3**(3): p. 203-220.
33. Furlan, P.M., et al., *Abnormal salivary cortisol levels in social phobic patients in response to acute psychological but not physical stress*. Biol Psychiatry, 2001. **50**(4): p. 254-9.
34. Klumbies, E., et al., *The Reaction to Social Stress in Social Phobia: Discordance between Physiological and Subjective Parameters*. PloS one, 2014. **9**: p. e105670.
35. Fowles, D.C., et al., *Publication recommendations for electrodermal measurements*. Psychophysiology, 1981. **18**(3): p. 232-239.
36. Lykken, D., et al., *Correcting psychophysiological measures for individual differences in range*. Psychological Bulletin, 1966. **66**(6): p. 481.

37. Shaffer, F. and J.P. Ginsberg, *An Overview of Heart Rate Variability Metrics and Norms*. *Frontiers in public health*, 2017. **5**: p. 258-258.
38. Royall, R.M., *The effect of sample size on the meaning of significance tests*. *The American Statistician*, 1986. **40**(4): p. 313-315.
39. Zimmer, P., et al., *Virtually stressed? A refined virtual reality adaptation of the Trier Social Stress Test (TSST) induces robust endocrine responses*. *Psychoneuroendocrinology*, 2019. **101**: p. 186-192.
40. Martel, F., et al., *Salivary cortisol levels in socially phobic adolescent girls*. *Depression and Anxiety*, 1999. **10**(1): p. 25-27.
41. Krämer, M., et al., *Subjective, autonomic, and endocrine reactivity during social stress in children with social phobia*. *J Abnorm Child Psychol*, 2012. **40**(1): p. 95-104.
42. Beaton, E.A., et al., *Low salivary cortisol levels among socially anxious young adults: Preliminary evidence from a selected and a non-selected sample*. *Personality and Individual Differences*, 2006. **41**(7): p. 1217-1228.
43. Helminen, E.C., et al., *A meta-analysis of cortisol reactivity to the Trier Social Stress Test in virtual environments*. *Psychoneuroendocrinology*, 2019. **110**: p. 104437.
44. Vive. *HTC VIVE Pro Series*. [cited 2020 05.07.2020, 16:00]; Available from: <https://www.vive.com/uk/product/vive-pro/>.
45. Shiban, Y., et al., *Trier Social Stress Test in vivo and in virtual reality: Dissociation of response domains*. *Int J Psychophysiol*, 2016. **110**: p. 47-55.
46. Pereira, T., et al., *Heart rate variability metrics for fine-grained stress level assessment*. *Comput Methods Programs Biomed*, 2017. **148**: p. 71-80.
47. Hermann, R., et al., *Physical versus psychosocial stress: effects on hormonal, autonomic, and psychological parameters in healthy young men*. *Stress*, 2018. **22**: p. 1-10.
48. Price, M., et al., *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*, 2011. **25**(6): p. 763-770.
49. Felnhofer, A., et al., *Afraid to be there? Evaluating the relation between presence, self-reported anxiety, and heart rate in a virtual public speaking task*. *Cyberpsychol Behav Soc Netw*, 2014. **17**(5): p. 310-6.
50. Kennedy, R.S., et al., *Configural Scoring of Simulator Sickness, Cybersickness and Space Adaptation Syndrome: Similarities and Differences*. *Virtual and adaptive environments: Applications, implications, and human performance issues*, 2003: p. 256.
51. Bynion, T.-M. and M. Feldner, *Self-Assessment Manikin*. 2017. p. 1-3.
52. Goodman, W.K., J. Janson, and J.M. Wolf, *Meta-analytical assessment of the effects of protocol variations on cortisol responses to the Trier Social Stress Test*. *Psychoneuroendocrinology*, 2017. **80**: p. 26-35.
53. von Dawans, B., C. Kirschbaum, and M. Heinrichs, *The Trier Social Stress Test for Groups (TSST-G): A new research tool for controlled simultaneous social stress exposure in a group format*. *Psychoneuroendocrinology*, 2011. **36**(4): p. 514-22.
54. Boesch, M., et al., *Mood and autonomic responses to repeated exposure to the Trier Social Stress Test for Groups (TSST-G)*. *Psychoneuroendocrinology*, 2014. **43**: p. 41-51.

8 Danksagung

Ich möchte mich an dieser Stelle für die Unterstützung zahlreicher Personen und Institutionen bedanken.

Mein besonderer Dank gilt meinem Doktorvater Prof. Dr. Jürgen Gallinat für die hervorragende Betreuung und Unterstützung bei der Durchführung der Studie. Die zahlreichen Anregungen zur Umsetzung der Studie und die konstruktive Kritik waren hierbei stets hilfreich.

Auch möchte ich mich bei Prof. Dr. Simone Kühn für die hervorragende Betreuung und die zahlreichen inhaltlichen Hilfestellungen bedanken.

Außerdem bedanke ich mich bei meiner Kollegin Fariba Mostajeran für die großartige gemeinsame Zusammenarbeit.

Für die Beratung und Hilfe bei der statistischen Auswertung der Daten möchte ich mich an dieser Stelle bei Christian Wiessner bedanken.

Ebenfalls möchte ich mich bei Frances Bohnsack, Dimitrij Kugler, Peter Lanzoni und Benedict Lütken für ihre Geduld und die Durchführung der *in vivo*-Testungen bedanken.

Ein besonderer Dank gilt allen Teilnehmern und Teilnehmerinnen der Studie, ohne die diese Dissertation und die Publikation nicht hätten entstehen können.

Meiner Familie, meinen Geschwistern und meinen Freunden danke ich für ihre Unterstützung und Ermutigungen während des Studiums und der Promotion. Ein besonderer Dank gebührt meiner Mutter, Fatma Balci, für ihre unermüdliche Unterstützung.

9 Lebenslauf

Lebenslauf wurde aus datenschutzrechtlichen Gründen entfernt.

Eidesstattliche Versicherung

Ich versichere ausdrücklich, dass ich die Arbeit selbständig und ohne fremde Hilfe verfasst, andere als die von mir angegebenen Quellen und Hilfsmittel nicht benutzt und die aus den benutzten Werken wörtlich oder inhaltlich entnommenen Stellen einzeln nach Ausgabe (Auflage und Jahr des Erscheinens), Band und Seite des benutzten Werkes kenntlich gemacht habe.

Ferner versichere ich, dass ich die Dissertation bisher nicht einem Fachvertreter an einer anderen Hochschule zur Überprüfung vorgelegt oder mich anderweitig um Zulassung zur Promotion beworben habe.

Ich erkläre mich einverstanden, dass meine Dissertation vom Dekanat der Medizinischen Fakultät mit einer gängigen Software zur Erkennung von Plagiaten überprüft werden kann.

Unterschrift: