Assessing the quality of experience of virtual reality games

Vlahović, Sara

Doctoral thesis / Disertacija

2024

University of Zagreb

FACULTY OF ELECTRICAL ENGINEERING AND COMPUTING

Sara Vlahović

ASSESSING THE QUALITY OF EXPERIENCE OF VIRTUAL REALITY GAMES

DOCTORAL THESIS

Zagreb, 2024

University of Zagreb

FACULTY OF ELECTRICAL ENGINEERING AND COMPUTING

Sara Vlahović

ASSESSING THE QUALITY OF EXPERIENCE OF VIRTUAL REALITY GAMES

DOCTORAL THESIS

Supervisor: Professor Lea Skorin-Kapov, PhD

Zagreb, 2024

Sveučilište u Zagrebu

FAKULTET ELEKTROTEHNIKE I RAČUNARSTVA

Sara Vlahović

PROCJENA ISKUSTVENE KVALITETE IGARA U VIRTUALNOJ STVARNOSTI

DOKTORSKI RAD

Mentor: prof. dr. sc. Lea Skorin-Kapov

Zagreb, 2024.

The doctoral thesis was completed at the University of Zagreb Faculty of Electrical Engineering and Computing, Department of Telecommunications.

Supervisor: Professor Lea Skorin-Kapov, PhD

This thesis has: 224 pages

Thesis number: ___________

About the Supervisor

Lea Skorin-Kapov is Professor and head of the Multimedia Quality of Experience Research Lab (MUEXLab) at the University of Zagreb Faculty of Electrical Engineering and Computing. Her primary research interests include Quality of Experience assessment and modeling of immersive services and applications, and cross-layer management of QoS/QoE in networks. She teaches courses at bachelor, masters, and doctoral levels dealing with multimedia services, heuristic optimization methods, and communication networks.

She received her Dipl.-Ing., M.S., and Ph.D. degrees in Telecommunications from the Faculty of Electrical Engineering and Computing (FER) at the University of Zagreb, Croatia, in 2001, 2004, and 2007, respectively. From 2001-2009 she was employed in the Research and Development Center of Ericsson Nikola Tesla, Zagreb, Croatia, doing research on QoS signaling, negotiation, and adaptation for multimedia services. From 2002-2009 she was also an adjunct teaching and research assistant at the Department of Telecommunications, FER, University of Zagreb.

Since 2010 she has been employed at the Department of Telecommunications, FER, University of Zagreb. She is currently involved in a number of research and industry funded projects and is principal investigator for the project "Modeling and Monitoring QoE for Immersive 5G-Enabled Multimedia Services" funded by the Croatian Science Foundation. She is a member of the Croatian Centre of Research Excellence for Data Science and Advanced Cooperative Systems, and a member of the Management Board of the Center for Artificial Intelligence at FER.

She has published over 100 scientific papers, serves on the editorial board of IEEE Transactions on Network and Service Management, and has served as guest editor for a number of journals such as Elsevier Computer Networks, IEEE Journal of Selected Topics in Signal Processing, and ACM Transactions on Multimedia Computing, Communications, and Applications.

Lea Skorin-Kapov is a senior member of IEEE and from 2019-2022 served as Chapter chair of the IEEE Communications Society - Croatia Chapter.

O mentorici

Lea Skorin-Kapov je redovita profesorica na Zavodu za telekomunikacije Fakulteta elektrotehnike i računarstva (FER) Sveučilišta u Zagrebu te voditeljica istraživačkog laboratorija Multimedia Quality of Experience Research Lab (MUEXLab). Njezino glavno područje istraživačkog interesa jest procjena i modeliranje iskustvene kvalitete imerzivnih aplikacija i usluga te mehanizmi upravljanja kvalitetom usluge/iskustvenom kvalitetom u mrežama. Podučava na FER-ovom preddiplomskom, diplomskom i doktorskom studiju o višemedijskim uslugama i komunikacijama, heurističkim metodama optimizacije i komunikacijskim mrežama.

Diplomirala je 2001. godine, magistrirala 2004. godine te doktorirala 2007. godine na Sveučilištu u Zagrebu FER, smjer telekomunikacije i informatika. Od 2001.-2009. godine radila je u Istraživačkom odjelu tvrtke Ericsson Nikole Tesle d.d., Zagreb u jedinici za Istraživanje i razvoj, gdje se bavila istraživanjem područja signalizacije, pregovaranja i prilagodbe kvalitete usluge za napredne višemedijske usluge. Od 2002. do 2009. godine bila je vanjski mlađi asistent na Zavodu za telekomunikacije na FER-u, Sveučilišta u Zagrebu.

Od 2010. godine zaposlena je na Zavodu za telekomunikacije, FER, Sveučilište u Zagrebu. Uključena je u razne istraživačke i industrijske projekte. Trenutno vodi istraživački projekt “Modeliranje i praćenje iskustvene kvalitete imerzivnih višemedijskih usluga u 5G mrežama” kojeg financira Hrvatska zaklada za znanost. Članica je Znanstvenog centra izvrsnosti za znanost o podatcima i kooperativne sustave te članica Vijeća Centra za umjetnu inteligenciju na FER-u.

Autorica je i koautorica više od 100 znanstvenih radova u časopisima, zbornicima radova s međunarodnih konferencija i knjigama. Članica je uredničkog odbora časopisa IEEE Transactions on Network and Service Management te je sudjelovala kao gostujući urednik za časopise poput Elsevier Computer Networks, IEEE Journal of Selected Topics in Signal Processing i ACM Transactions on Multimedia Computing, Communications, and Applications.

Lea Skorin-Kapov je senior member sekcija IEEE te je od 2019-2022 obnašala funkciju predsjednice Odjela za komunikacije, hrvatske sekcije IEEE.

Acknowledgements

First and foremost, I would like to express my heartfelt gratitude to my supervisor, Prof. Lea Skorin-Kapov. Her guidance and dedication throughout all these years have been invaluable. Moreover, her expertise, feedback, and encouragement have not only shaped this work but will also continue to serve as a source of inspiration and motivation for me in the years to come.

I extend my sincere appreciation to Prof. Mirko Sužnjević for his essential professional input and motivational pep talks. His support played a crucial role in helping me navigate challenges and maintain focus throughout this journey.

I am grateful for the help and guidance provided by Prof. Nina Pavlin-Bernardić and Prof. Maja Matijašević. I would also like to extend my gratitude to my colleagues Ivan Slivar, Matko Šilić, and Krešimir Bačić, as well as to students Emilia Haramina, Neda Kušurin, Filip Gustetić, Mateo Paladin, and Fran Posarić, for their assistance in administering user studies. Special thanks go to students Monika Matokanović and Filip Nemec for their contributions to the implementation of the specialized platform for evaluating interaction mechanics quality. Additionally, I wish to express my appreciation to Prof. Katrien De Moor and Prof. Željka Car for generously dedicating their time and effort to reviewing this dissertation and providing valuable feedback.

To all of the MUEXlab members, past and present, I express my gratitude for creating a positive, supportive, and uplifting work environment. Your inventive work-related jokes have made everything more enjoyable.

Lastly, I would like to express my sincere thankfulness to my friends and family.

To Mateja, Nives, Tena, Irena, Katarina, and Manđo — thank you for your willingness to lend your lovely ears to my endless ranting and for your patience in dealing with my hermit phases. I owe you one!

To my parents, my grandparents, Lovro, and Sven -– your unwavering love and support have been my rock during the toughest parts of this process. I can honestly say that none of this would have been possible if it wasn't for you.

Abstract

The last decade marks a significant era for virtual reality (VR) technology. Ever since the initial prototype of Oculus Rift disrupted the industry by introducing VR to a wider audience, numerous consumer-grade VR systems have appeared on the market, boasting affordable prices with the promises of providing a highly immersive experience. However, despite initially greeting the revived technology with enthusiasm, to this day, consumers have remained reluctant to invest in personal VR devices.

The issue of inadequate user satisfaction with currently available immersive devices has prompted the research community to focus its efforts on investigating Quality of Experience (QoE) for immersive media. To better understand the ways in which QoE for immersive media, such as VR, differs from QoE for non-immersive platforms, it is necessary to consider the defining characteristics of the technology, from the ways in which it replaces the sights and sounds of the physical world with artificial stimuli, to the ways in which it utilizes six-degrees-of-freedom tracking capabilities of contemporary hardware.

Centered around VR gaming, this thesis proposes a three-tier set of models illustrating the possible relationships between QoE influence factors (IFs), QoE features, QoE constituents and the overall QoE for this type of service. Selected relationships proposed in the model are investigated over the course of multiple user studies. Recognizing the diversity in VR content, as well as the diversity in its consumer base, a novel taxonomy of VR interaction mechanics is proposed. Popular mechanics are evaluated using subjective measures of player experience, workload and VR-induced symptoms and effects, supported by objective measures of reaction time and task performance.

In addition to presenting the outcomes of user research, this thesis also provides insights into the process of designing and conducting user studies. It offers guidelines for the implementation of interaction mechanics, as well as the guidelines for evaluating their quality. Lastly, it addresses the challenges encountered when conducting multiplayer user studies and suggests potential directions for future research.

Keywords: Quality of Experience, virtual reality, VR gaming, game mechanics, VR-induced symptoms and effects

Procjena iskustvene kvalitete igara u virtualnoj stvarnosti

Dizajn prototipa sustava Oculus Rift 2011. godine obilježio je početak komercijalne ere tehnologije virtualne stvarnosti (engl. *virtual reality*, skraćeno VR). Tijekom više od desetljeća nakon njegovog izlaska na tržište, ovom su sustavu slijedili uređaji konkurentnih tvrtki poput HTC-a, Sonyja i Valvea. Međutim, unatoč početnom entuzijazmu šire javnosti, prihvaćenost ove tehnologije od strane korisnika još uvijek nije dosegnula očekivanu razinu, a čak se i najnoviji sustavi susreću s kritikama vezanim uz neudobnost i nepraktičnost uređaja te nedovoljnu kvalitetu sadržaja.

Kako bismo bolje razumjeli razlike u zahtjevima za postizanje odgovarajuće kvalitete doživljaja kod imerzivnih medija poput VR-a u usporedbi s doživljajem kod ne-imerzivnih medija poput stolnih računala i pametnih telefona, potrebno je sagledati jedinstvene mogućnosti i izazove imerzivnog hardvera i softvera. Suvremeni sustavi za virtualnu stvarnost korisnika odvajaju od stvarnog svijeta, blokirajući stvarne podražaje i zamjenjujući ih virtualnima, primjerice prikazivanjem sadržaja na zaslonima širokog polja pogleda (engl. *field-of-view*, skraćeno FoV). Imerzivna iskustva današnjice ostvaruju se kroz korištenje naprednih ulaznih modaliteta, poput upravljanja glasom ili gestama, velik je naglasak na praćenju pozicije i orijentacije glave te ruku u šest stupnjeva slobode (engl. *six degrees of freedom*, skraćeno 6DoF). Softverska rješenja prate ovaj napredak u hardverskim mogućnostima uređaja, omogućujući iznimnu interaktivnost imerzivnih iskustava.

Međutim, imerzivnost i interaktivnost ove vrste sustava i usluga dolaze nauštrb njihove udobnosti, pristupačnosti i sigurnosti. Suvremeni zasloni montirani na glavu (engl. *head-mounted display*, skraćeno HMD) teški su i loše balansirani, što može izazvati umor, bol i neispravno držanje korisnika. Također, imaju nedovoljnu razlučivost i nalaze se vrlo blizu očiju, što može izazvati zamor i naprezanje. Određene VR usluge mogu izazvati mučninu, dezorijentaciju i opću slabost, a u slučajevima kada se od korisnika zahtijeva značajna razina fizičke aktivnosti prilikom interakcije, korištenje VR-a može izazvati i umor i bol u mišićima, pa čak i ozlijede. Uz navedene negativne posljedice korištenja VR-a, simptomi i učinci izazvani VR-om (engl. *VR-induced symptoms and effects*, skraćeno VRISE) mogu uključivati i psihološke učinke, poput usporenih reakcija i emocionalne traume nakon izlaganja uznemirujućem sadržaju.

Razumijevanje ovih značajki VR sustava ključno je za istraživanje iskustvene kvalitete i komponenti koji doprinose njezinoj formaciji. U ove komponente ubrajamo čimbenike utjecaja iskustvene kvalitete (engl. *QoE influence factors*, skraćeno QoE IFs) vezane uz korisnika, sustav i kontekst korištenja, značajke iskustvene kvalitete (engl. *QoE features*) percipirane tijekom korištenja, te u konačnici, ključne sastavnice iskustvene kvalitete (engl. *QoE constituents*). Cilj istraživanja iskustvene kvalitete VR sustava jest ustanoviti odnose između različitih komponenata te u tom procesu identificirati rješenja za postizanje što većeg stupnja imerzivnosti i interaktivnosti uz što manju učestalost i intenzitet VRISE-a.

U fokusu ovog doktorskog rada su igre u virtualnoj stvarnosti. Ova složena vrsta usluge nastaje kroz integraciju različitih disciplina, od inženjerstva do umjetnosti, što utječe na želje i zahtjeve korisnika. Također, riječ je o izrazito interaktivnoj vrsti usluge koja kombinira različite ulazne i izlazne modalitete, te pred korisnika postavlja ciljeve čije ostvarenje zahtijeva mentalni i fizički napor (posebno kad su u pitanju VR igre). Za razliku od iskustvene kvalitete usluga koje koristimo za praktične primjene, iskustvena kvaliteta igara prvenstveno ovisi o njihovoj sposobnosti da zabave igrača. Budući da se radi o interaktivnim sustavima koji postavljaju ciljeve pred korisnika, u igrama se doživljaj zabave i užitka postiže kroz svladavanje izazova proporcionalnih vještini igrača. Međutim, napor koji proizlazi iz svladavanja izazova u VR-u istovremeno može potaknuti igrače na rani prekid igre ili ih u potpunosti odvratiti od ove usluge i platforme. Postizanje zadovoljavajuće ravnoteže između imerzije, interaktivnosti, VRISE-a, napora i zabave u ovakvim sustavima još je složeniji zadatak kad su u pitanju višekorisničke igre, kod kojih ishod ne ovisi samo o igri i pojedinom igraču, već o vještinama i rezultatima svih pojedinih igrača. Pri tome rezultirajuća iskustvena kvaliteta dodatno ovisi i o društvenim odnosima igrača i dodatnim tehničkim značajkama sustava, poput onih vezanih uz umreženost.

Iako već godinama postoje smjernice za ispitivanje iskustvene kvalitete igranja (npr. ITU-T preporuka P.809), a sve više raste broj smjernica i preporuka vezanih uz ispitivanje iskustvene kvalitete imerzivnih usluga, specifični slučaj igranja u VR-u (engl. *VR gaming*) zahtijeva zasebni pristup i smjernice za provođenje korisničkih ispitivanja i razvoj igara vođen korisničkim iskustvom. Dodatno postoji i potreba za sveobuhvatnim modelom iskustvene kvalitete igranja u VR-u koji kombinira postojeće spoznaje o iskustvenoj kvaliteti igranja s postojećim spoznajama o iskustvenoj kvaliteti korištenja VR usluga, te koji je utemeljen na korisničkim studijama koje uključuju široko dostupne VR uređaje i sadržaj. Pritom je osobito bitno obratiti pozornost na raznolikost dostupnog VR sadržaja. Budući da se igre u VR-u mogu razlikovati po pitanju podržanih modaliteta, estetskih karakteristika, fleksibilnosti i složenosti priče i zadataka te načinu kretanja i interakcije s virtualnim okolišem, ne postoji univerzalno rješenje za ostvarenje igre koja je zabavna, ugodna, sigurna i pristupačna. Šarolikost korisničkih iskustava s VR igrama rezultat je i raznolikosti u demografskim značajkama, ukusima, potrebama i vještinama njihovih postojećih ili potencijalnih igrača, ali i raznolikosti u kontekstu (vremenskom, prostornom ili društvenom) u sklopu kojeg se odvija epizoda igranja. Uzimajući u obzir navedene potrebe, u sklopu ovog doktorskog istraživanja provedene su sljedeće aktivnosti:

- Razvoj konceptualnih modela koji identificiraju i opisuju odnose između čimbenika utjecaja, značajki i sastavnica iskustvene kvalitete te ukupne iskustvene kvalitete.
- Klasifikacija interakcijskih mehanika VR igara.
- Korisničke studije koje istražuju utjecaj različitih mehanika igara i žanrova na mjere

radnog opterećenja, fizičke nelagode, fizioloških simptoma i kognitivnih učinaka.

- Korisničke studije koje istražuju utjecaj različitih parametarskih vrijednosti na izvršenje zadatka te iskustvenu kvalitetu i njezine značajke za različite interakcijske mehanike u virtualnoj stvarnosti.
- Korisničke studije koje istražuju utjecaj faktora vezanih uz društveni kontekst, kvalitetu mreže i razinu iskustva, na iskustvenu kvalitetu i njezine značajke u kontekstu igranja višekorisničkih igara u virtualnoj stvarnosti.
- Analiza prikupljenih podataka.
- Razvoj modela koji opisuju odnose između odabranih čimbenika utjecaja i značajki iskustvene kvalitete.
- Formulacija smjernica za dizajniranje i procjenu interakcijskih mehanika za igranje u virtualnoj stvarnosti.

U sklopu doktorskog rada opisani su rezultati osam studija, od kojih je sedam provedeno u sklopu doktorskog studija, a jedna provedena u sklopu ranijeg istraživanja te iznova analizirana u sklopu doktorskog istraživanja. Prilikom provođenja studija korištene su metode perceptualne procjene, odnosno metode koje uključuju ispitivanje ljudskih sudionika. Sudionicima je prezentiran jedan ili više testnih stimulusa te se od njih tražilo da koriste testni sustav sami ili uz drugog sudionika. Korištene su subjektivne (upitnici) te objektivne metode (mjere kognitivnih učinaka te uspješnosti izvršenja zadataka) procjene korisničkog iskustva.

Obzirom na raznovrsnost VR sustava i sadržaja, svi uređaji i testni materijali korišteni u sklopu istraživanja morali su zadovoljiti unaprijed određene kriterije. Korišteni su samo komercijalno dostupni HMD-ovi s pratećim kontrolerima koji podržavaju praćenje u 6DoF. Komercijalno dostupne igre korištene za ispitivanje podržavale su stajaći i/ili tzv. *roomscale* način rada, odnosno mapiranje 1:1 između kretanja po stvarnom prostoru i kretanja u virtualnom prostoru, bez korištenja potpomognutih metoda navigacije unutar virtualnog okoliša. U svim odabranim igrama, interakcija s virtualnim okolišem ostvarena je prvenstveno uporabom ruku, uz mapiranje 1:1 između pokreta kontrolera i virtualne ruke. Fokus istraživanja bio je na vizualnim elementima iskustva, kao i na elementima poze i pokreta. Sav korišteni sadržaj bio je sintetički (računalno generiran), uz lokalno iscrtavanje. Iako su komercijalne igre korištene kao testni sadržaj u većini provedenih studija, tri su studije koristile prototipnu testnu platformu za istraživanje iskustvene kvalitete interakcijskih mehanika u VR-u, što je omogućilo fleksibilnost pri zadavanju testnih scenarija te veći broj i preciznost prikupljenih objektivnih mjera.

U sklopu doktorskog rada predstavljen je pregled područja. Iznesene su definicije iskustvene kvalitete te su na temelju postojeće literature predstavljeni njezini čimbenici utjecaja, značajke i sastavnice. Predstavljeni su primjeri postojećih modela koji opisuju korisničko iskustvo s VR tehnologijom. Opisani su različiti pristupi metodološkom dizajnu korisničkih studija fokusiranih na interaktivni VR te postojeće smjernice i uvriježene prakse vezane uz odabir testnog materijala, subjektivnih i objektivnih metrika, vremenskog i prostornog konteksta studije, te samih sudionika.

Predstavljen je konceptualni troslojni skup modela koji prikazuju odnose između ukupne iskustvene kvalitete, njezinih sastavnica, značajki i čimbenika utjecaja. Model visoke razine predstavlja sastavnice iskustvene kvalitete zajedničke svim iskustvima igranja u VR-u (igraće iskustvo, radno opterećenje i VRISE) te dodatne sastavnice karakteristične za iskustva višekorisničkog igranja (percipirana kvaliteta umrežavanja, percipirana društvena interakcija, iskustvo međuigračke uključenosti) i opisuje kako one utječu na ukupnu iskustvenu kvalitetu. Model srednje razine razlaže svaku od navedenih sastavnica na skup mjerljivih značajki. Tako se primjerice percipirana kvaliteta umrežavanja sastoji od sljedećih značajki: percipirano kašnjenje igrača i suigrača, responzivnost i glatkoća ulaznih kontrola te percipirana degradacija performansi. Predstavljen je i skup modela niske razine koji predlažu pretpostavljene odnose između čimbenika utjecaja vezanih uz korisnika/igrača, sustav i kontekst. Zbog velikog broja potencijalnih čimbenika, svaki od modela niske razine predstavlja samo djelomični skup potencijalnih čimbenika i odnosa. Prvi model (model naziva VR_QOE_LLM_1) predstavlja odnose između žanra i mehanika igre te radnog opterećenja i VRISE-a. Drugi model (VR_QOE_LLM_2) predstavlja odnose između specifičnih konfiguracija interakcijskih mehanika te odabranih značajki igraćeg iskustva, radnog opterećenja i VRISE-a. Treći model (VR_QOE_LLM_3) predstavlja odnose između kašnjenja u mreži, odnosa između igrača i prethodnih iskustava s igranjem, VR tehnologijom i sportskim aktivnostima te igraćeg iskustva, percipirane kvalitete umrežavanja, percipirane društvene interakcije, iskustva međuigračke uključenosti.

Uzimajući u obzir raznovrsnost VR igara, predstavljeni su pojmovi mehanika igre te interakcijskih mehanika. Kombiniranjem uvriježenih kriterija za klasifikaciju interakcijskih tehnika (npr. simetrija, sinkronost, interakcijska vjerodostojnost) s novim kriterijima fokusiranim na korištenje alata, implementaciju meta i dimenzije interakcijskog prostora, osmišljena je i predstavljena nova taksonomija interakcijskih mehanika za igranje u VR-u. Korištenje ove taksonomije omogućava lakši opis i usporedbu testnih materijala korištenih u korisničkim studijama te interpretaciju dobivenih rezultata.

Studije 1 (N = 20) i 2 (N = 20) provedene su s ciljem određivanja utjecaja različitih žanrova i mehanika igara na pojavu, učestalost i intenzitet VRISE-a te na intenzitet radnog opterećenja. Primjeri VRISE-a istraženi u sklopu ove dvije studije uključuju bol i umor ruku, vrata i leđa, pojavu mučnine izazvane VR-om te različite vrste nelagode izazvane zaslonom montiranim na glavu (npr. nelagoda zbog kablova, težine, temperature, kvalitete zaslona), kao i negativne učinke VR-a na kognitivne procese, konkretnije na usporavanje vremena reakcije (učinak koji je u prethodnoj literaturi najčešće tumačen kao posljedica mučnine izazvane VR-om). Na temelju rezultata studije ustanovljeno je da čak i vrlo kratkotrajne epizode igranja VR igara (20 minuta) mogu uzrokovati raznovrsne, a ponekad i vrlo intenzivne simptome i učinke, od kojih se mnogi

razlikuju ovisno o značajkama odigrane igre. Rezultati studija potvrdili su neke od odnosa predstavljenih u inicijalnom modelu VR_QOE_LLM_1.

U svrhu detaljnijeg proučavanja utjecaja različitih konfiguracija popularnih interakcijskih mehanika igara u VR-u na korisničko iskustvo, osmišljen je radni okvir skraćenog naziva INTERACT koji predstavlja skup smjernica za njihovu evaluaciju putem specijalizirane platforme. Određen je skup implementacijskih parametara odabranih mehanika, te predloženi primjeri mjera koje se mogu koristiti za njihovu evaluaciju. Predstavljena metodologija korištena je u studijama 3 (N = 30), 4 (N = 30) i 5 (N = 30), s ciljem istraživanja utjecaja specifičnih konfiguracija interakcijskih mehanika na igraće iskustvo, radno opterećenje i VRISE te na objektivne mjere izvršenja zadataka. Na temelju analize dobivenih rezultata osmišljen je skup smjernica za dizajn interakcijskih mehanika uzimanja i postavljanja objekata (engl. *pick and place*), pucanja (engl. *shoot*) i rezanja/siječenja (engl. *slash*). Dodatno, rezultati navedenih studija potvrdili su neke od odnosa koji su predstavljeni u inicijalnom modelu VR_QOE_LLM_2.

Studije 1-5 fokusirane su na korisničko iskustvo prilikom igranja jednokorisničkih (engl. *singleplayer*) igara, odnosno igara namijenjenih jednom igraču. Za razliku od njih, studije 6-8 fokusirane su na korisničko iskustvom prilikom igranja višekorisničkih (engl. *multiplayer*) kolaborativnih (studije 6 i 7) i kompetitivnih (studija 8) igara. Preliminarni rezultati dobiveni u studijama 6 i 7, u kojima je istražen utjecaj mrežnih performansi na korisničko iskustvo, pokazali su da čimbenici poput društvenog konteksta, težine igre i implementacije interakcijskih mehanika potencijalno mogu djelovati na korisničku percepciju mrežnog kašnjenja i ukupnu iskustvenu kvalitetu. Stoga je studija 8 zamišljena kao sveobuhvatna studija koja, uz utjecaj mrežnih čimbenika, istražuje i utjecaj igre, društvenog konteksta te igračkih vještina i prethodnog iskustva na odabrane značajke igraćeg iskustva, percipirane kvalitete umrežavanja, percipirane društvene interakcije i iskustva međuigračke uključenosti. Rezultati studije 8 potvrdili su neke od odnosa koji su predstavljeni u inicijalnom modelu VR_QOE_LLM_3.

Istraživanjem opisanim u ovom doktorskom radu ostvaren je znanstveni doprinos koji se sastoji od sljedećih elemenata:

1. Specifikacija utjecaja mehanika igre u virtualnoj stvarnosti na objektivne i subjektivne mjere nelagode te odabrane fiziološke simptome, kognitivni učinak i značajke iskustvene kvalitete igranja.
2. Modeli koji opisuju odnos između relevantnih čimbenika utjecaja čovjeka, sustava i konteksta te odabranih značajki iskustvene kvalitete igranja u virtualnoj stvarnosti.
3. Smjernice za dizajn i evaluaciju mehanika igre u virtualnoj stvarnosti, temeljene na rezultatima studija za procjenu iskustvene kvalitete.

Keywords: iskustvena kvaliteta, virtualna stvarnost, igranje u VR-u, mehanike igara, simptomi i učinci izazvani VR-om

Contents

Chapter 1

Introduction

This chapter presents the background and motivation for the thesis (Section 1.1) and introduces the problem statement (Section 1.2). After discussing the chosen method of solution and the scope of the thesis (Section 1.3), the chapter provides a summary of the main thesis contributions (Section 1.4), before laying out its structure (Section 1.5).

1.1 Background and motivation

According to the QUALINET White Paper on Definitions of Immersive Media Experience [1], virtual reality (VR) is defined as a type of system that ‘‘*occludes physical space to provide interactive and non-interactive experiences of a fully computer-simulated “virtual” world or a photographically “captured” real world”*. While the term VR encompasses different types of devices, such as Cave Automatic Virtual Environment (CAVE) systems, today it is predominantly used to refer to experiences realized through the use of VR head-mounted displays (HMDs) and the accompanying input and output (I/O) devices. While the history of HMDs dates back to 1968 and Ivan Sutherland’s invention of The Sword of Damocles [2], HMD technology remained inaccessible to the broader public for several decades to come. However, a significant revolution for this immersive technology occurred in 2011 with the first prototype of Oculus Rift marking the beginning of a new era of virtual reality.

In the years following its conception, Oculus Rift paved the way for similar systems manufactured by companies such as HTC, Sony, and Valve. These contemporary takes on VR technology boasted affordable prices while providing a highly immersive experience with innovative methods of interaction. Unfortunately, even though studies addressing user experience demonstrated the advantages of VR compared to less immersive platforms [3, 4, 5], to this day the idea of actually investing in a personal VR system seems to resonate more with a devoted community of niche enthusiasts than with a mainstream audience of consumers. This sentiment is supported by relatively modest estimations and growth projections for augmented reality (AR) and VR technology presented in Statista Market Insights in June of 2023 [6].

The limited market infiltration of VR and AR devices to date brings forth the issue of inadequate user satisfaction with currently available products. As a way of approaching this challenge from an academic perspective, over the last several years, the research community has increasingly focused on assessing Quality of Experience (QoE) for immersive media, a term defined as:*"the degree of delight or annoyance of the user of an application or service which involves an immersive media experience. It results from the fulfillment of his or her expectations with respect to the utility and/or enjoyment of the application or service in the light of the user's personality and current state."* [1].

To better understand the ways in which QoE for immersive media differs from QoE for non-immersive platforms, one needs to consider the unique possibilities and challenges of immersive hardware, and the ways in which these possibilities and challenges are being utilized — or *overlooked* — by stakeholders responsible for the development of immersive platforms and their respective content. For the purpose of this thesis, the following distinguishing characteristics of the VR platform are identified as being significant contributors to the overall user experience: immersivity, multimodal interactivity, and obtrusiveness.

The term immersivity, as used in this thesis, refers to immersion as a system property, defined as *"the degree to which immersive media environments sub-merges the perceptual system of the user in computer-generated stimuli"* [1, 7]. In the case of contemporary VR, experienced via commercially available VR systems, immersive audio reproduced through headphones integrated into HMDs overrides real-world sounds, while computer-generated images presented on stereoscopic displays — also characterized by a wide field-of-view (FoV) — cover and replace the image of the physical environment around the user. While VR controllers packaged with HMDs of today provide rudimentary haptic feedback, as technology progresses, the importance of tactile design can only be expected to increase, further improving the immersivity of VR technology in comparison to non-immersive platforms.

While certain immersive VR experiences rely primarily on multiple output modalities (e.g., 360 degree video), input modalities are key for a truly interactive experience as they allow users to navigate and/or manipulate the virtual environment. Recent generations of VR hardware moved on from mouse and keyboard input toward tracking the movements of the user in six degrees of freedom (6DoF), which is often supplemented with button and touchpad input from the handheld controllers. Other input modalities — such as gaze tracking or voice control — may also be utilized, leading to the high level of multimodal interactivity that is characteristic of the VR platform. This deviation from the way in which input modalities have been utilized in non-immersive platforms presents a challenge to VR content developers as they focus on designing novel interaction techniques to be used with immersive platforms, whilst also looking to identify (and emulate) real world interactions that translate well to the virtual landscape.

Although increased immersivity and multimodal interactivity of the VR platform provide numerous opportunities over less immersive platforms, it is necessary to point out one of its biggest disadvantages — its obtrusiveness. While the industry is increasingly leaving cumbersome cables behind and focusing on device portability, as of the time of this writing, users are still being weighed down by sizeable headsets. Moreover, in addition to the positioning of bright displays causing significant eye strain, the resolution of current HMDs is not enough to avoid a visible screen door effect, bringing the users out of the virtual experience. These issues are further aggravated when hardware is used for VR apps that trigger cybersickness, or require a lot of movement, increasing the platform's obtrusiveness to the point of discomfort and possibly endangering the health and safety of its users.

Understanding and accounting for these characteristics of the VR platform is crucial for the investigation of QoE and QoE aspects that contribute to its formation: QoE influence factors (IFs) related to users, system, and context, QoE features that can be perceived during use, and QoE constituents that are formed as a result of a cognitive process that aggregates multiple features [8]. Even though the ultimate goal of such research may seem straightforward — aim to eliminate elements of the tested system that are found to be bothersome, such as those contributing to its obtrusiveness, while boosting the elements that increase its immersivity and interactivity — the issues at hand are complex and interrelated. By increasing elements of applications that foster immersivity and presence, applications may become more likely to trigger cybersickness [9]. Increasing the role of movement in an application, while contributing to its immersivity and interactivity, may lead to accessibility issues or physical exhaustion.

This balancing act is further complicated when considering the difference between systems as “toys” and systems as “tools” as discussed by Vlahovic et al. in [10] and based on [11]. “Tools" are described as systems that are utilized to accomplish an external goal. An ideal "tool" is designed for a specific, highly practical purpose, requires minimal effort on the user's part, and yet functions in a perfectly reliable, efficient way. On the contrary, the quality of digital games, which fall under the "toy" category, is not judged through the lens of utility, but by the level of enjoyment their use brings to the user. As highly interactive systems that provide players with goals and test their skills, digital games are usually meant to be at least somewhat challenging. In the context of VR, this challenge often pertains to gross motor movement. Therefore, even though it may contribute to obtrusiveness, completely minimizing the workload necessary for a given task may actually detract from the experience. Likewise, eliminating effects that impose a perceptual load on the player may decrease relevant hedonic qualities of the game, such as its aesthetics, or the sense of awe and excitement it aims to inspire. The aforementioned paradoxes highlight the importance of close examination of various factors that shape the experience of a VR gamer, as well as the importance of a thorough analysis of ways in which these factors relate to each other and contribute to the overall QoE score.

1.2 Problem statement

As discussed by Vlahovic et al. [12], even though general guidelines for the subjective evaluation of gaming Quality of Experience have existed for several years (e.g., ITU-T Recomm. P.809 [13]), and ongoing efforts conducted by the research and standardization communities are directed towards investigating user experiences with immersive services [14, 15, 16, 17], the specific case of VR gaming has not yet been addressed in depth.

Sitting at the intersection of QoE for VR and QoE for gaming, the challenge of exploring QoE for VR gaming requires researchers to combine different aspects of both fields, honoring both the unique characteristics of VR as a platform and the specific challenges of gaming as a service. This includes essentially cherry-picking the most relevant influence factors (IFs), QoE features, and QoE constituents that have already been identified separately by VR QoE researchers and gaming QoE researchers, and creating a coherent and comprehensive amalgamation of these aspects that serves the VR gaming use-case, whilst also addressing the ways in which these aspects relate to one another and contribute to the overall QoE. In other words, there is a need for a coherent and comprehensive model of QoE for VR gaming which builds off of existing knowledge of QoE for VR and QoE for gaming, and is grounded in user studies involving consumer-grade VR hardware and commercially available VR games.

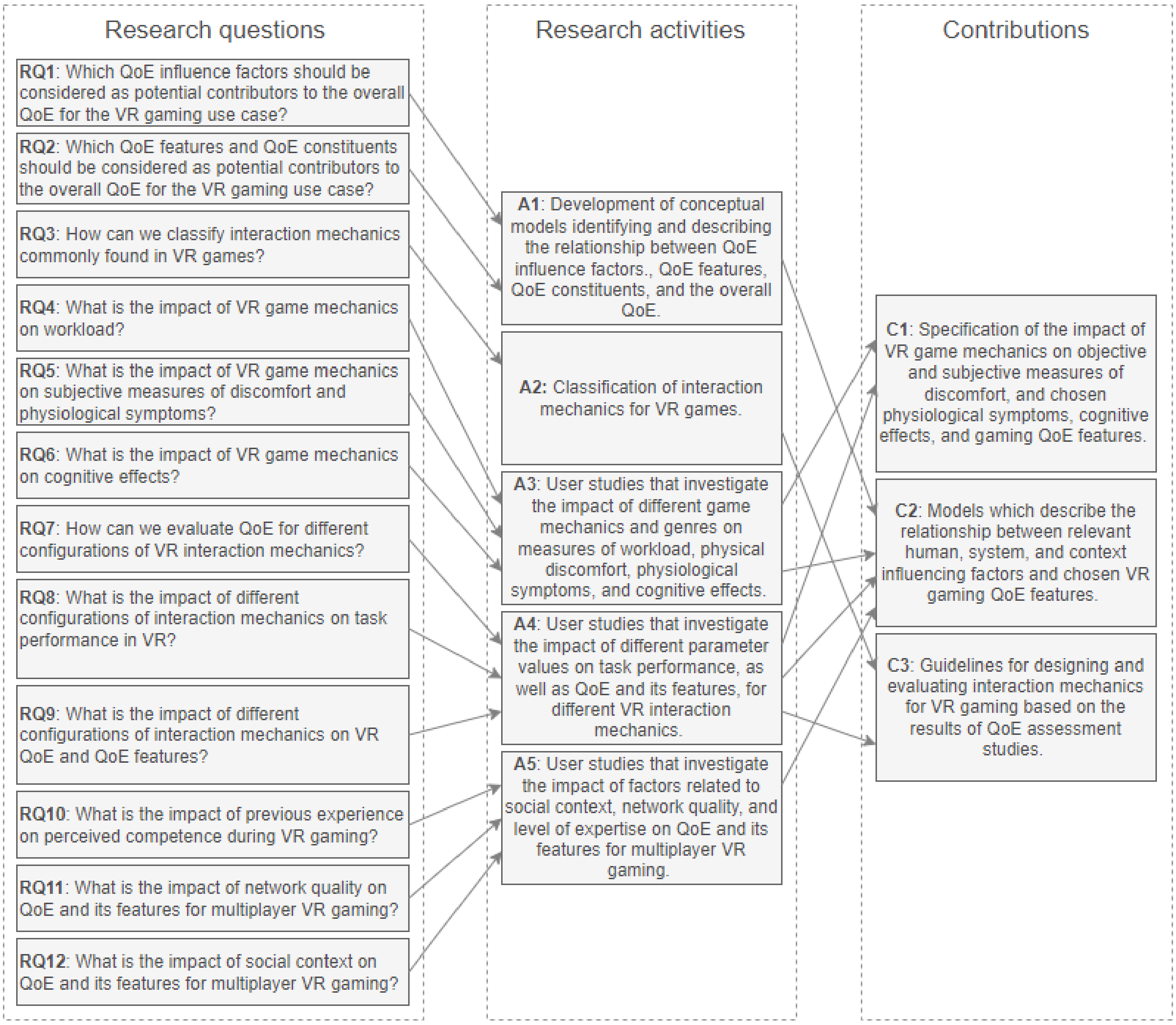
In order to perform user studies to support and/or extend the proposed models, it is necessary to employ the appropriate methodology. As with indentifying key QoE aspects, researchers performing VR gaming user studies usually consider and combine methods and measures often used in VR user research (e.g., cybersickness measures such as the Simulator Sickness Questionnaire [18]) with those used in gaming user research (e.g., questionnaires such as the Game Experience Questionnaire [19, 20]). However, when developing novel QoE models and/or guidelines for QoE assessment for VR gaming, it is also necessary to note that merely borrowing and combining established approaches from the two related — but separate — research fields may not be enough. Circling back to the themes of the previous section, it is important to consider the specific differences between VR games and non-immersive games, as well as between VR games and other types of VR applications, with regard to the aforementioned distinguished characteristics of the VR platform — immersivity, multimodal interactivity, and obtrusiveness — as a way to identify possible gaps in existing methodologies that require a more specialized approach.

Even though FoV width of immersive displays improves the sense of presence, it is likely to also trigger an increase in cybersickness [21]. Thus, although cybersickness may occur with non-immersive platforms, this issue is more pronounced for VR games, as seen in [22]. Moreover, although non-immersive games require some level of fine motor skills, due to different input modalities they generally do not have the same requirements in terms of gross motor skills that VR games often do. Unlike desktop, console and mobile games which rely on touchscreens and buttons, VR game mechanics often make use of movements and postures similar to those of real-life activities. While movement during VR gaming is generally welcomed by players [23], research analyzing the Steam game market [24] demonstrates that players have a tendency to rate VR games lower than non-VR games, with a preference toward VR games that are shorter in duration. According to the authors, this occurrence is likely due to obtrusiveness, as experiencing physical fatigue, eye strain, thermal discomfort and cybersickness may drive users away from VR gaming. Because these issues are less prominent for other platforms, the critical importance of workload and physical comfort for VR gaming is generally not sufficiently addressed in more general-purpose/cross-platform gaming QoE models or methods.

In terms of other types of VR services, although they tend to share the same immersive display technology and input modalities as VR games, the capabilities of those modalities may not be utilized to the same degree. For example, watching 360-degree videos does not necessarily require manual interaction and participating in a VR meeting does not require the level of physical activity, speed, or dexterity that playing a VR game does. Although certain aspects of VR-related discomfort (namely cybersickness) are often examined in user studies involving the use of VR, other ergonomic issues of VR use have not yet received adequate attention by the research community, as highlighted by [25, 26]. Furthermore, there is a need to take a closer look at particular physically demanding game mechanics and game interaction techniques (which we refer to as interaction mechanics, as explained further in Chapter 4 that repeatedly occur across many VR games, such as those discussed in the article by Vlahovic et al. [10] and presented as a part of this thesis.

Further challenges arise once research goals are extended to cover not only singleplayer, but also multiplayer experiences, which rely on satisfying networking solutions, as well as on collaboration, competition, and social interaction between players. Taking this into account, while this thesis focuses on the implementation of common VR game mechanics, the aim is also to address QoE IFs pertaining to players themselves and the social context in which the game is being played, along with network access and quality. Addressing both singleplayer and multiplayer experiences, the goal is to use findings obtained by performing user studies to inspire and support the proposed set of models describing QoE for VR gaming, both in terms of high level relationships between QoE and its constituents, as well as the low level investigation of the effects of different IFs on QoE features.

Aiming to address the aforementioned research challenges, specific research questions that are addressed in this thesis are presented in Figure 1.1, along with their mapping to research activities and the novel contributions of the thesis. Note: activities performed in the latter stages of research, namely those related to data analysis, modeling and formulation of guidelines, have been omitted from the figure for the sake of visual simplicity, but will be acknowledged (along with their links to other activities and contributions) in the visual overview of research activities in the following section.



Research questions

RQ1: Which QoE influence factors should be considered as potential contributors to the overall QoE for the VR gaming use case?

RQ2: Which QoE features and QoE constituents should be considered as potential contributors to the overall QoE for the VR gaming use case?

RQ3: How can we classify interaction mechanics commonly found in VR games?

RQ4: What is the impact of VR game mechanics on workload?

RQ5: What is the impact of VR game mechanics on subjective measures of discomfort and physiological symptoms?

RQ6: What is the impact of VR game mechanics on cognitive effects?

RQ7: How can we evaluate QoE for different configurations of VR interaction mechanics?

RQ8: What is the impact of different configurations of interaction mechanics on task performance in VR?

RQ9: What is the impact of different configurations of interaction mechanics on VR QoE and QoE features?

RQ10: What is the impact of previous experience on perceived competence during VR gaming?

RQ11: What is the impact of network quality on QoE and its features for multiplayer VR gaming?

RQ12: What is the impact of social context on QoE and its features for multiplayer VR gaming?

Research activities

A1: Development of conceptual models identifying and describing the relationship between QoE influence factors., QoE features, QoE constituents, and the overall QoE.

A2: Classification of interaction mechanics for VR games.

A3: User studies that investigate the impact of different game mechanics and genres on measures of workload, physical discomfort, physiological symptoms, and cognitive effects.

A4: User studies that investigate the impact of different parameter values on task performance, as well as QoE and its features, for different VR interaction mechanics.

A5: User studies that investigate the impact of factors related to social context, network quality, and level of expertise on QoE and its features for multiplayer VR gaming.

Contributions

C1: Specification of the impact of VR game mechanics on objective and subjective measures of discomfort, and chosen physiological symptoms, cognitive effects, and gaming QoE features.

C2: Models which describe the relationship between relevant human, system, and context influencing factors and chosen VR gaming QoE features.

C3: Guidelines for designing and evaluating interaction mechanics for VR gaming based on the results of QoE assessment studies.

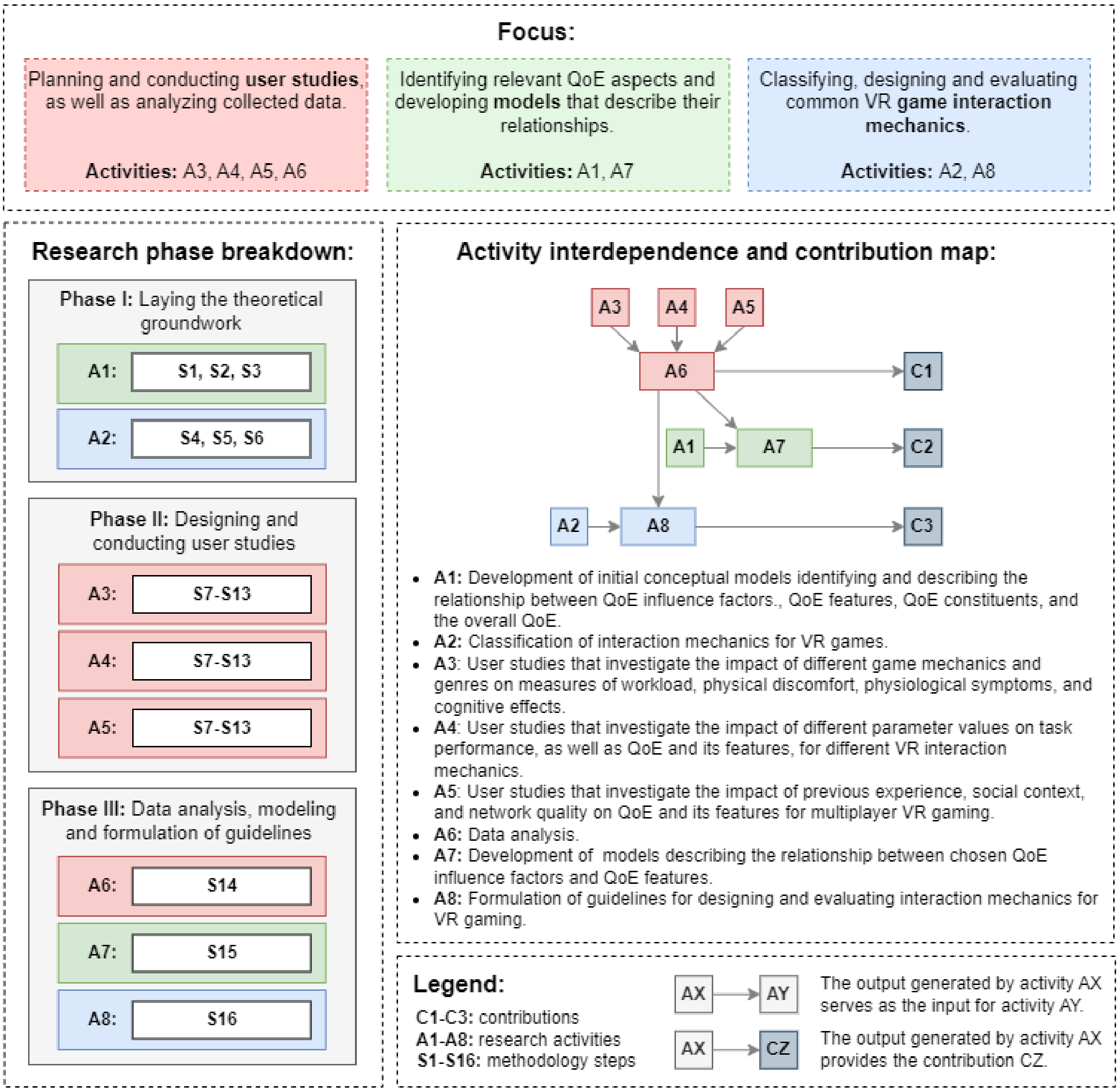
Figure 1.1: Mapping of addressed research questions, research activities, and contributions of the thesis

1.3 Method of solution and scope

Research conducted for the purpose of this thesis involved eight research activities over three research phases, as illustrated in Figure 1.2. Each of these activities encompassed one or more methodology steps, which are specified in Table 1.1. It is necessary to note that research activities, research phases and methodology steps presented herein were not necessarily performed in the order listed, as the described process was not linear but iterative, with the results of each performed user study informing the process of subsequent studies, as well as further inspiring the development of proposed models and guidelines.

For the purpose of this thesis, we performed perceptual assessment, i.e., we used methods that involve testing human evaluators, which we refer to as participants. As explained in [27, 28], during the course of perceptual assessment studies, participants may be presented with one or more test stimuli, asked to interact with the tested system, and/or use the tested system in interaction with another participant. Based on these experiences, users provide subjective evaluations which are either qualitative or quantitative. Moreover, researchers may also employ objective methods of evaluation, which may include different physiological, behavioral, and task performance measures.

The thesis is grounded in eight user studies, with their individual methodologies summarized in Table 1.1. The order in which the studies are listed and referenced follows the structure of the thesis; however, it is not chronological. All studies were conducted during the course of the author's doctoral research (2018-2023), except for Study 6, which was conducted as part of the author's master's thesis [29], with its methodology later revisited and extended in Study 7. The general approach combined various subjective measures administered via single- or multi-item questionnaires with objective measures of cognitive (reaction time) and task performance.


Focus:

Planning and conducting **user studies**, as well as analyzing collected data.
Activities: A3, A4, A5, A6

Identifying relevant QoE aspects and developing **models** that describe their relationships.
Activities: A1, A7

Classifying, designing and evaluating common VR **game interaction mechanics**.
Activities: A2, A8

Research phase breakdown:

Phase I: Laying the theoretical groundwork

Phase II: Designing and conducting user studies

Phase III: Data analysis, modeling and formulation of guidelines

Activity interdependence and contribution map:

- **A1:** Development of initial conceptual models identifying and describing the relationship between QoE influence factors., QoE features, QoE constituents, and the overall QoE.
- **A2:** Classification of interaction mechanics for VR games.
- **A3**: User studies that investigate the impact of different game mechanics and genres on measures of workload, physical discomfort, physiological symptoms, and cognitive effects.
- **A4**: User studies that investigate the impact of different parameter values on task performance, as well as QoE and its features, for different VR interaction mechanics.
- **A5**: User studies that investigate the impact of previous experience, social context, and network quality on QoE and its features for multiplayer VR gaming.
- **A6:** Data analysis.
- **A7:** Development of models describing the relationship between chosen QoE influence factors and QoE features.
- **A8:** Formulation of guidelines for designing and evaluating interaction mechanics for VR gaming.

Legend:

C1-C3: contributions
A1-A8: research activities
S1-S16: methodology steps

The output generated by activity AX serves as the input for activity AY.

The output generated by activity AX provides the contribution CZ.

Figure 1.2: A visual overview of research phases and research activities

Table 1.1: Methodology steps pertaining to different research phases (RP) and research activities (RA).

RP	RA	Label	Methodology step
I	A1	S1	Identifying the QoE IFs most likely to have a significant influence on the VR gaming experience.
		S2	Identifying features and constituents that comprise the overall VR gaming QoE.
		S3	Proposing conceptual models with hypothesized relationships between selected factors, features, constituents, and the overall QoE.
	A2	S4	Proposing a taxonomy of VR game interaction mechanics (IMs).
		S5	Identifying configurable IM parameters that serve as potential QoE IFs.
		S6	Proposing a set of guidelines for the evaluation of VR IM QoE.
II	A3, A4, A5	S7	Selecting QoE aspects (IFs, features, constituents) to be manipulated and examined.
		S8	Deciding on the methodology of each study, i.e., choosing appropriate methods and measures.
		S9	Deciding on the appropriate hardware and test material to be used in user studies with consideration of what is generally representative of the VR gaming use case and suitable for the particular selection of observed QoE aspects.
		S10	Deciding on the environment and setup of each study.
		S11	Deciding on the temporal context of each study, i.e., the number of study sessions and the overall duration of QoE exposure.
		S12	Defining the criteria for the participant sample demographic and performing participant recruitment.
		S13	Conducting each study and collecting data.
III	A6	S14	Choosing and performing appropriate statistical tests and interpreting data.
	A7	S15	Using the results to support and extend the proposed conceptual models of VR gaming.
	A8	S16	Using the results to formulate guidelines for designing and evaluating VR IMs.

As explained in Vlahovic et al. [10], while the medium of VR can be experienced using a wide range of input devices, locomotion, and interaction techniques, for the purpose of this thesis the focus was on widely-used interaction techniques that are manual, controller-based, and isomorphic, i.e., characterized by the one-to-one mapping between movements of the physical hand and the corresponding motion of the in-game virtual hand. In addition to observing game interaction mechanics (IMs) that use the virtual hand metaphor to interact with target objects in a direct manner, mechanics that support tool-mediated interaction with target objects were also considered. For the sake of this thesis, the focus was primarily on the visual aspect of the VR experience (as opposed to audio or haptic modalities), while also addressing the role of position and movement as detrimental aspects of controller-based VR. Moreover, considering that the issue of locomotion in VR is a complex topic with its own set of specific challenges, like cybersickness, the focus was placed on standing and Room-Scale games that do not employ specific in-game locomotion methods other than providing a direct mapping between user movement within the tracked physical space and their position within the virtual environment.

Test material used for all user studies was synthetic and locally rendered. The majority of studies were conducted with commercial games as test material, which provided a more realistic experience of VR gaming. As explained by Vlahovic et al. [30], based on research regarding user preferences in VR gaming [24], commercial games that belong to most frequently downloaded genres — action, shooter, and simulation — were chosen as test material, along with an example of a music/rhythm game in Study 1, as the authors of [24] note that this particular type of game (in addition to the action genre) tends to be especially well-received by VR gamers. The choice to include games that belong to different genres corresponded with the decision to choose games that differ in terms of interaction mechanics [10]. Considering a relatively short duration of each gaming session, all chosen games had to be simple and beginner-friendly, with easy-to-grasp rules and mechanics. Studies 6-8 were centered around multiplayer gaming, while singleplayer games were used in Studies 1 and 2.

Because commercial games are essentially black boxes from the perspective of a QoE researcher, when using them as test material it is not possible to access detailed information regarding task performance, nor is it possible to control and evaluate specific configurations of individual game mechanics. Because of this, Studies 3-5 were conducted using a prototype test platform for singleplayer use developed by with the help of students Monika Matokanović and Filip Nemec specifically to be used as test material for the purpose of this thesis.

While studies conducted in the field provide greater external validity, all eight studies were conducted in a laboratory because it provided a controlled environment. This was necessary for several reasons, such as ensuring that all participants used the same hardware setup for studies exploring VRISE and mechanics (Studies 1-5), enabling manipulation of network parameters in multiplayer gaming studies (Studies 6-8), and facilitating scenario switching and data collection when using a specialized test platform (Studies 3-5). Regarding temporal context of VR exposure in conducted studies, it is important to note that, for the purpose of this research, participants were immersed in VR for relatively short durations (i.e., an approximate total of 20 to 45 minutes per session). Studies 1 and 2 took place over three gaming sessions held on separate days, as a way of preventing carryover effects and excessive accumulation of physical symptoms. All other studies consisted of a single study session.

A total of 219 participants took part in the research presented in this thesis. Due to challenges with participant recruitment (especially for the duration of the COVID-19 pandemic), but also due to the dynamic and physically demanding nature of tested VR games, a decision was made to only recruit participants who do not experience common health (e.g., epilepsy) and mobility limitations of VR use. Considering that cybersickness is most pronounced in very young users (up to the age of 12) [31] and increases for users above 50 years old [32], with mobility also decreasing with age [33], special care was taken to include only young adults in studies that explored physical discomfort — Studies 1-2 (age range: 20-29) and Studies 3-5 (age range: 18-34). Moreover, recognizing the sex differences in experiencing VR [34, 35, 36], conscious effort was directed toward recruiting as many female participants as possible, with Studies 1,2, and 8 achieving an equal balance between sexes.

While this section was written to summarize and explain the common threads in methodology design that connect all eight studies, it is important to note that the design of each study will be further described in their respective chapters, along with their results. Due to small

Table 1.2: A summary of conducted user studies

		Study 1	Study 2	Study 3	Study 4
Year		2020/2021	2021	2022/2023	2022/2023
QoE aspects	Observed QoE IFs	system (game genre/mechanics)	system (game genre/mechanics)	system (interaction mechanics)	system (interaction mechanics)
	Observed QoE constituents (based on subjective measures)	workload, VRISE	workload, VRISE	workload, VRISE, player experience	workload, VRISE, player experience
	Collected objective measures	cognitive performance metrics (reaction time)	cognitive performance metrics (reaction time)	slashing task performance metrics	pick-and-place task performance metrics
Participants	Number	20	20	30	30
	Sex	10 female, 10 male	10 female, 10 male	13 female, 17 male	16 female, 14 male
	Age	20-29 (average: 24.5)	21-29 (average: 23.85)	18-33 (average: 23.4)	19-31 (average: 23.1)
Hardware	HMD	HTC Vive Pro Eye	HTC Vive Pro Eye	HTC Vive Pro Eye	HTC Vive Pro Eye
	Controllers	VIVE (2018)	VIVE (2018)	Valve Index	Valve Index
Test material	Game/platform	commercial games (BS, OU, SS)	commercial games (FN, DB, SPT)	specialized test platform for the evaluation of VR game mechanics	specialized test platform for the evaluation of VR game mechanics
	Game mechanics	slash (BS), pick-and-place (OU), shoot (SS)	slash (FN), pick-and-place (DB), shoot (SPT)	slash	pick-and-place
Test material	Game genre	action/rhythm (BS), cooking simulator (OU), action/shooter (SS)	action (FN), cooking simulator (DB), action/shooter (SPT)	action	puzzle
	Game mode	singleplayer	singleplayer	singleplayer	singleplayer

		Study 5	Study 6	Study 7	Study 8
Year		2022/2023	2018	2019	2021/2022
QoE aspects	Observed QoE IFs	system (interaction mechanics)	system (network)	system (network)	system (network, game genre/mechanics), player (level of expertise), context (social context)
	Observed QoE constituents (based on subjective measures)	workload, VRISE, player experience	perceived quality of networking	perceived quality of networking	perceived quality of networking, perceived social interaction, interplayer involvement experience
	Collected objective measures	shooting task performance metrics	in-game survival rate	in-game survival rate	/
Participants	Number	30	24	33	32
	Sex	13 female, 17 male	10 female, 14 male	12 female, 21 male	16 female, 16 male
	Age	18-34 (average: 22.1)	14-38 (average: 26.5)	15-51 (average: 25.6)	21-42 (average: 24)
Hardware	HMD	HTC Vive Pro Eye	Oculus Rift (examined participant) + HTC Vive (active player)	Oculus Rift (examined participant) + HTC Vive (passive player)	Meta (Oculus) Quest + Meta (Oculus) Quest
	Controllers	Valve Index	Oculus Touch (1st generation) + VIVE	Oculus Touch (1st generation) + VIVE	Oculus Touch (2nd generation) + Oculus Touch (2nd generation)
Test material	Game/platform	specialized test platform for the evaluation of VR game mechanics	commercial game (SS)	commercial game (SS)	commercial games (ETT, BO)
	Game mechanics	shoot	shoot	shoot	hit (ETT), shoot (BO)
Test material	Game genre	action/shooter	action/shooter	action/shooter	sport (ETT), action/shooter (BO)
	Game mode	singleplayer	collaborative multiplayer (2 active players)	collaborative multiplayer (active player + passive player)	multiplayer (2 active players)

sample sizes and non-normal distribution of data, results were analyzed using non-parametric methods of statistical analysis. Accounting for these limitations, models described in this thesis are not statistical models, but rather conceptual models of VR gaming QoE grounded in statistics. Because exploring every aspect of VR gaming QoE is a complex challenge that is beyond the scope of this thesis, the initial high-level conceptual model that describes the relationships between QoE features/constituents and overall QoE is purely theoretical. However, individual relationships between QoE IFs and QoE features described in low-level models are highlighted based on results that were found to be statistically significant.

1.4 Summary of contributions

The contributions of this thesis can be summarized as follows:

- **C1:** Specification of the impact of VR game mechanics on objective and subjective measures of discomfort, and chosen physiological symptoms, cognitive performance, and gaming QoE features.
- **C2:** Models which describe the relationship between relevant human, system, and context influencing factors and chosen VR gaming QoE features.
- **C3:** Guidelines for designing and evaluating VR game mechanics based on the results of QoE assessment studies.

1.5 Thesis structure

The structure of the thesis is as follows. Chapter 2 presents a state-of-the-art review of relevant literature and standards pertaining to QoE assessment for interactive VR applications. Further focusing on VR games as a subset of interactive VR, Chapter 3 is centered around identifying key QoE aspects, i.e., influence factors, features, and constituents that comprise the overall VR gaming QoE. Hypothetical relationships between identified QoE aspects are visualized in a set of proposed conceptual models describing the process of QoE formation.

Chapter 4 presents the definitions of game and interaction mechanics, followed by a taxonomy of common VR IMs. Chapter 5 presents the methodology and results of two user studies (Study 1 and Study 2) addressing the impact of experiencing different VR game genres with different game mechanics on negative effects of VR use. Further focusing on VR game mechanics, Chapter 6 lays out several considerations regarding the implementation of interaction mechanics and presents a framework of guidelines for the evaluation of VR IM quality. The concepts and guidelines presented in the introductory parts of the chapter are demonstrated in three user studies (Study 3, Study 4, Study 5) evaluating the impact of different configurations of slash, pick-and-place, and shoot IM implementations on QoE and its features. In addition to presenting the significant results of these studies, this chapter also presents useful guidelines for the design of VR IMs.

Shifting the focus from singleplayer experiences to multiplayer setups, Chapter 7 provides a concise overview of two preliminary studies (Study 6 and Study 7) focused on exploring the impact of network factors on QoE for a collaboration game, and the results of a more comprehensive user study (Study 8) exploring the impact of player, context, and system factors on QoE and its features for two competitive games. All of the chapters presenting the findings of user studies refer to related conceptual models from Chapter 3 and further confirm hypothesized links between QoE IFs and QoE features with statistically significant results. Thesis conclusions are are summarized in Chapter 8, along with a discussion on its limitations, and additional ideas for future research.

Chapter 2

An overview of Quality of Experience for interactive VR applications

2.1 Introduction

This chapter presents an overview of related research pertaining to the topic of QoE for interactive VR applications, such as VR games. The chapter is divided into two parts: the first one delves into the relevant aspects (influence factors, features, and constituents) impacting QoE for VR, along with an outline of existing models related to the topic. The second part of the chapter is focused on the practicalities of perceptual assessment of QoE for interactive VR, addressing various issues regarding methodology design, such as deciding on the materials and methods, and recruiting participants for the study. The material in this chapter has been published as a journal article [28]. As the published version offers a significantly more in-depth overview of the topics covered in this chapter, accompanied by numerous references, readers are encouraged to consult it for additional details and sources. Compared to its published version, however, the condensed version of the text, as presented in this thesis, was slightly updated with additional information and references.

2.2 Aspects of Quality of Experience for VR gaming

Quality of Experience is a complex construct formed as an aggregation of various elements, affected by a wide range of factors occurring both inside and outside of the user. This section presents an overview of relevant QoE influence factors before delving into a concise discussion on the definitions of QoE features and QoE constituents as higher-level aspects of QoE. The section is concluded with a short overview of existing models pertaining to user experience with VR technology.

2.2.1 QoE influence factors

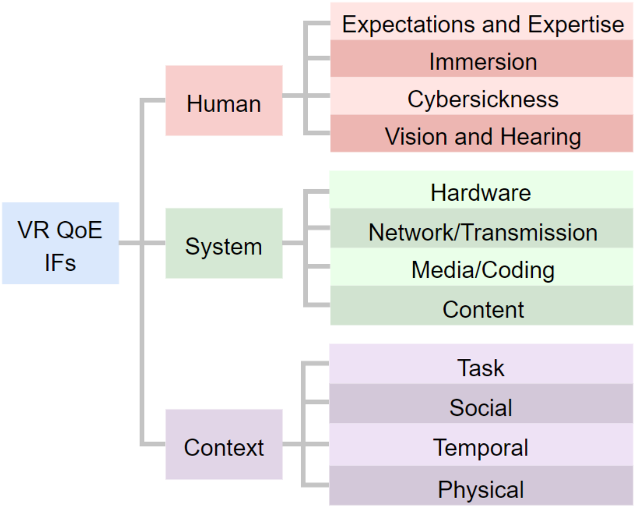


Figure 2.1: VR QoE influence factor categories (adapted from [37])

The Qualinet White Paper [38] defines QoE **influence factors** (IFs) as traits exhibited by the system, service, application, or even users themselves, that may potentially influence QoE of the users of an application or service. Our concise overview of influence factors affecting the interactive VR experience is based on — but not limited to — the classification of influence factors for VR as presented in ITU-T Recommendation G.1035 [37] (Fig. 2.1).

Human Influence Factors

In terms of human (also referred to as user) influencing factors, researchers often choose to examine *dynamic* human factors, such as the current affective state of the user, as well as *static* human factors, which refer to the fixed traits of the participant (e.g., age, sex, etc.). With the common occurrence of VR-related discomfort being an impetus for further research, a high importance is placed on human IFs such as history of illness (e.g., migraine, motion sickness), as well as relevant factors related to vision and hearing. Additionally, previous history of technology use may greatly influence task performance, level of discomfort, and overall satisfaction with the used system. To facilitate comparison of these aspects based on user expertise, participants can be classified according to their general experience with interactive applications (e.g., games) or immersive technology, experience with a particular type/genre of application or, even more specifically, previous experience using a particular application. While listed as influence factors in the ITU-T G.1035 recommendation, cybersickness and immersion may also be examined as QoE features or constituents, dependant on other human, system, and context factors, and are referred to as such in the remaining part of this thesis.

System Influence Factors

Hardware Influence Factors: In general, input (e.g., controllers. gesture control, movement tracking) and output modalities (e.g., headsets, haptic devices) play a significant role in user experience by greatly affecting different quality features. Unfortunately, current VR technology is riddled with ergonomic issues. For example, a greater size/weight of VR HMDs may be distracting and uncomfortable to some users andincrease the overall physical workload required to interact with the system [39]. As a result of their limitations in terms of adjustability, certain commercial headset designs are not adapted to suit the dimensions of a significant percentage of the population [40]. Individuals who use visual correction aids are even more likely to struggle with adjusting the headset to suit their needs [41]. Additionally, original versions of contemporary commercial VR headsets have been tethered to the PC and dependent on external sensors, which entails various issues with setup, tracking[42], and cumbersome cables[43]. However, as of late, standalone versions have been appearing on the market (e.g., Oculus Go, Oculus Quest), offering greater mobility and easier setup at the expense of computing power.

Network Influence Factors: Exploring the impact of networking factors (delay, jitter, bandwidth, packet loss) is currently especially crucial for VR applications centered around 360-degree video streaming(e.g., [44]), although networking issues may also cause significant issues forlocally-renderedinteractive networked VR applications (e.g., multiplayer games [45], teleoperation [46], or telepresence/collaboration applications). However, 5G and beyond networks are expected to be a disrupting force, revolutionizing the capabilities of immersive interactive VR as we know it. In addition to enabling split rendering, through significant improvements in network bandwidth, latency, and reliability, 5G and beyond networks provide the means for achieving hyper-realistic holographic telepresence.

Media/Coding Influence Factors: This group includes factors related to compression approaches used for encoding audio and video data, as well as other relevant types of information — e.g., point clouds.Aimed at facilitating efficient storage and network transmission, these factors are generally more relevant in the context of 360-degree video(e.g., [47])and cloud VR (e.g., [48]), compared to synthetic, locally rendered VR services, and are therefore mostly out of scope for this thesis.

Content Influence Factors: It is important to take into account different characteristics of the application used in a particular QoE study. In case of interactive applications, such as games, different genres/types can exhibit different levels of sensitivity to different kinds of impairment, such as latency, or produce different levels of immersion and discomfort. Notable examples of aspects that are of interest to VR researchers include different characteristics of the avatar (e.g., [49, 50]) and the visual environment(e.g., [51]), implementation of the locomotion method (e.g., [52, 53]), narrative(e.g., [54, 55]), UI design(e.g., [56, 57]), etc. Interaction techniques and, in case of gaming, game mechanics, are also one of the key content IFs, and will be addressed in depth in the latter chapters of this thesis.

Context Influence Factors

Following the discussion of content IFs, different 'tasks' performed by end users when evaluating QoE during VR use are relevant to consider, such as tasks involving different interaction or locomotion techniques. User experience may greatly differ depending on the duration and/or frequency of VR use, which impacts the formation of QoE. The physical environment may not be visible to the user immersed in a VE, but environmental variables may be distracting or facilitate the occurrence of discomfort, in addition to affecting internal and external validity. Further, social context is a relevant factor in case of multiplayer/collaboration applications. Arguably, it may be even more relevant for immersive applications compared to conventional platforms, considering that, in addition to an increase in perceived immersion [58, 59], VR multiplayer games may result in higher levels of empathy in users when compared to non-VR [59]. The role of temporal and spatial factors will be further discussed in the following section, and social context in Chapter 7.

2.2.2 QoE features and constituents

A *quality feature* is defined as *"a perceivable, recognized and nameable characteristic of the individual's experience of a service which contributes to its quality"* [60].Generally speaking, as described in [38, 61], quality features can be classified on several levels: level of direct perception (e.g., brightness, contrast, flicker, color perception, loudness, sound localization), level of action (e.g., immersion, perception of space, perception of one's own movements/motion within that space), level of interaction (e.g., responsiveness, naturalness of interaction), level of the usage instance of the system (e.g., learnability, intuitivity, ease of use, aesthetics), and level of service beyond the particular usage instance (e.g., appeal, usefulness, utility, acceptability). In the context of VR as an interactive, immersive, multi-modal medium, all examples mentioned above can be considered relevant features, but the extent of their individual contributions towards the overall QoE may vary depending on the particular type of VR service.

For example,Figure 2.2 displays a taxonomy of gaming QoE features, as presented in ITU-T Recomm. P.809 [62], and based on Möller et al. [63]. However, while certainly transferable to VR games, it does not include one of the most distinguishing characteristics of the platform — outside of depicted aspects, evaluating VR QoE often includes examining aspects such as discomfort and cybersickness, which happen as a result of the more physically intrusive nature

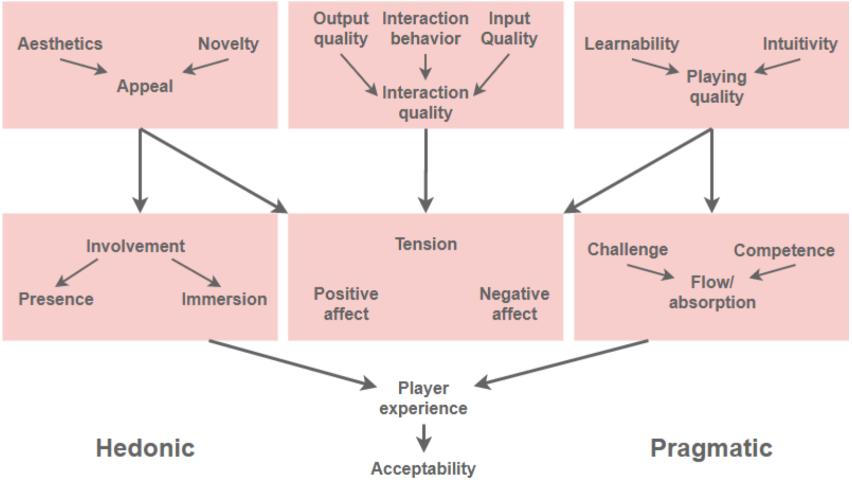


Figure 2.2: Quality features of gaming QoE (taken from [62]; based on [63])

of the platform, and may significantly degrade user experience.Indeed, aspects such as fatigue and discomfort have previously been recognized as some of the main features of QoE for certain media (i.e., 3D-TV [64]). This issue highlights the need for a general high-level taxonomy (or multiple service-specific taxonomies) of QoE features pertaining specifically to interactive VR. However, relevant features may be extracted from existing models of user experience with VR technology, such as those presented in the following subsection, as well as the one proposed in Chapter 3.

The recent ITU-T Recommendation on Quality of Experience assessment of extended reality meetings [8] introduces the term QoE constituent, an aspect formed as a result of a cognitive process which combines multiple quality features (Figure 2.3), such as those pertaining to perceptions of audiovisual stimuli and physiological responses. Based on this source alone, it is difficult to fully distinguish whether specific aspects of the VR experience should be categorized as features or constituents. However, the introduction of this terminology provides a welcome opportunity to observe QoE as a multilayered construct, which is explored further in Chapter 3. In our interpretation of this concept, this term is utilized to signify higher-level constructs that serve as umbrella terms for multiple features that are semantically connected and, in a significant percentage of cases, measured alongside each other in questionnaires that are commonly used in the field.

Because interactive VR is a complex multimedia service, and also covers a diverse range of applications, there is a large number of dimensions to identify and investigate, especially when this exploration is performed across different layers of the QoE formation process, as it is when we choose to distinguish QoE features from QoE constituents. Because of this, instead of dedicating a chapter to their definitions, some of them will be mentioned as elements of existing models discussed in the following subsection, while the ones that are most relevant to this thesis will be listed in Chapter 3 and further discussed in the related work sections of latter

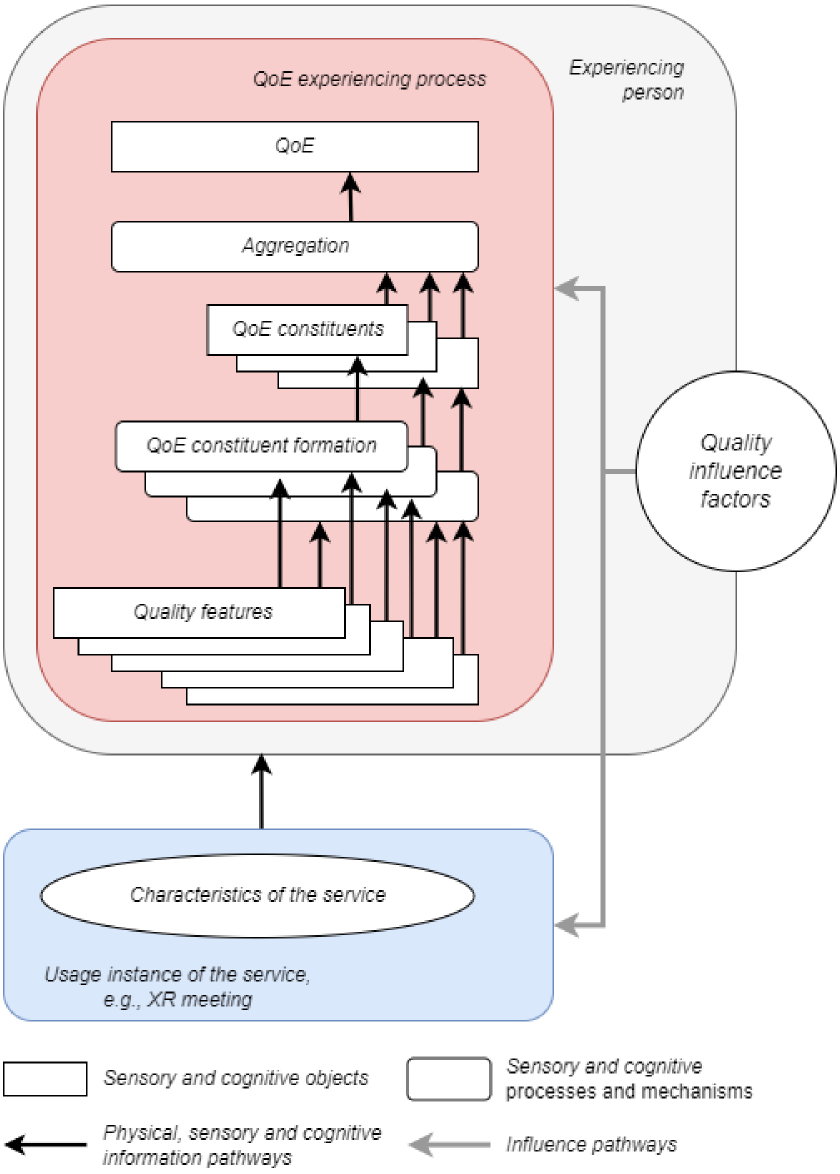


Figure 2.3: Visualization of the relationships between QoE, QoE constituents, quality features and QIFs (taken from [8])

chapters. The published version of this chapter [28], however, provides a more detailed review of selected features such as immersion, presence, and cybersickness, along with references to relevant sources. Because the distinction of "features" and "constituents" is not yet a fully established part of the QoE community vernacular, in the following subsection, which summarizes existing models developed by other researchers, all QoE aspects which could be classified as either features or constituents will be referred to as QoE or UX dimensions.

2.3 Modeling user experience with VR technology

To systematise the factors that influence the user to consider using or purchasing VR technology, researchers have developed appropriate technology acceptance models. Sagnier et al. [65] present a VR-adapted extension to the Technology Acceptance Model (TAM) [66]. The model describes the impact of different dimensions of user experience on perceived ease of use and perceived usefulness. Perceived ease of use was found to be significantly influenced by pragmatic quality, i.e., the usability and the utility of the product [67]. Perceived usefulness was found to be significantly influenced by stimulation (a hedonic quality that refers to *"the individual's pursuit of novelty and challenge*" [67]) and personal innovativeness. Participants' intention to use VR appears to be significantly increased by perceived usefulness, and significantly decreased by the severity of cybersickness symptoms, while a significant direct effect of presence has not been found, although it may pose an indirect influence by affecting other variables.

A similar TAM-based model, focusing on VR hardware acceptance, is presented by Manis and Choi [68]. The model distinguishes between intention toward using VR hardware, and intention toward purchasing VR hardware. Unlike the model by Sagnier et al. [65], this model does not examine the influence of presence and cybersickness, but it does account for user-related factors such as age, previous experience, and the price they were willing to pay for the product. The authors discuss curiosity, perceived usefulness, and perceived ease of use.

Kari and Kosa [69] presented an extension of the hedonic-motivation system acceptance model (HMSAM) developed by Lowry et al. [70]. With the goal of assessing the impact of different dimensions on immersion and intention to use, dimensions present in the original model (perceived ease of use, perceived usefulness, curiosity, joy, control) were joined by aspects pertaining to perceptions of physical and mental health and well-being. The study, conducted on 473 VR gamers, found that VR gaming was affected by hedonic gaming aspects more than the utilitarian and inconvenience-related factors. Although participants reported notable levels of physical exertion during VR gaming, it appears that enjoyment of the activity is a more significant motivator for VR use than its utalitarian role as an excercise aid. Contrary to results of Sagnier et al. [65], VR-induced discomfort and cybersickness did not impact gaming intention nor immersion, nor were they associated with physical health and well-being.

Schafer et al. [71] investigated the impact of VR gaming on game enjyoment. While the resulting enjoyment path model also includes dimensions such as perceived interactivity, spatial presence, and perceived reality, as supported by prior research on gaming (but not VR gaming) experience [72, 73, 74], Schafer et al. place their focus on the role of cybersickness as a significant negative effect of VR use. Their study (N=160) was conducted using several games that differed based on the extent of sensory conflict, displayed on two generations of Oculus headsets. While hardware innovations did not seem to mitigate the symptoms of cybersickness, different games had a more noticeable effect. The presented model also confirms previous research [72, 73, 74] by showing that perceived interactivity and realism predict spatial presence, which is a significant predictor of enjoyment.

Tcha-Tokey et al. [75] presented the UXIVE (User Experience in Immersive Virtual Environment) model based on the study involving 152 participants. The UXIVE model is comprised of 10 dimensions: presence, engagement, immersion, flow, usability, skill, emotion, experience consequence (i.e., an assortment of VR-induced symptoms such as cybersickness and fatigue), judgement, and technology adoption. Contrary to Schafer et al. [71], the authors found that negative effects of VR use influence technology adoption in a negative way. They further found that presence was influenced by flow and engagement, while flow was influenced by experience consequence and skill. Experience consequence, along with flow and presence, also influenced emotion. Technology adoption was influenced by flow, engagement, usability, and experience consequence.

Large sample sizes used for singular studies explored in papers described in this subsection allowed for the use of statistical methods such as structural equation modeling. The models described in the following chapter and revisited in subsequent chapters of this thesis, however, were developed as conceptual rather than statistical models, following an exploratory, rather than confirmatory approach. With the open-ended goal of gaining a more holistic understanding of QoE for VR, proposed models are based on several smaller studies, conducted in a controlled laboratory environment, and focused on exploring different constituents and features, as opposed to a single large study or online survey. Furthermore, studies and models presented in this thesis focus primarily on the relationship between specific influence factors and QoE features, rather than the interplay of higher-level QoE aspects such as features and constituents. While different from the other models described in this section, the models described in this thesis are in line with previous studies in that they recognize VR discomfort as a significant dimension affecting VR user experience, while further extending to include relevant aspects of multiplayer gaming.

2.4 Perceptual assessment of QoE for interactive VR applications

This section elaborates on practical considerations of performing perceptual assessment. Starting from the selection of test material, common practices in VR user research are listed, pertaining to the choice of subjective and objective assessment methods, the common flow of study procedure, temporal and environmental considerations, and recruitment of participants.

2.4.1 Test material

The test material used for conducting VR user studies depends on the aim of the study, and can range from applications with a practical purpose, such as those intended for therapeutic use (e.g., physical therapy, cognitive therapy, phobia treatment), educational applications or scientific visualisations (e.g., medical applications, military training), to applications intended for entertainment purposes (e.g., games, drawing in VR). Test material can be developed specifically for the study, or it may be a short sample of an existing application. The latter option is especially appropriate for VR gaming studies. As suggested by the ITU-T Recomm. P.809 [62], researchers should carefully select a sample that displays a mechanic that is typical for the game (or another type of application). If using a fixed level of difficulty, researchers should aim to select a sample that is appropriate for participants with various levels of experience. Otherwise, they may choose to keep it adjustable, so that it can be adapted to fit the skill level of each participant. Prior to conducting the actual study, test material should be thoroughly examined to ensure that the application runs smoothly. As discussed in [76], the frustration caused by encountering bugs and crashes during a test session is likely to degrade reported user experience.

Schatz et al. [77] highlighted the deficit of standardized VR content as well as a lack of standardized test tasks that would enable the reproduction of user studies across different laboratories and research groups. An example of such a test task can be found in [42], where the authors use a simple pick-and-place task to compare the performance of different VR systems. While design, development, and distribution of standardized test content remains an open challenge, researchers can facilitate comparison between studies by describing the used application, as well as chosen methods of interaction and locomotion.

Spending a longer period of time in VR may cause issues with discerning between the virtual world and the physical reality, as seen in [78]. While short-term effects, such as experiencing so-called *Game Transfer Phenomena* (GTP) [79] shortly after exposure to a non-stressful VR application, may not pose a significant threat to psychological and emotional well-being of the participant, the impact of immersion may be increased or prolonged in case of exposure to stressful, scary or otherwise disturbing content. Despite obviously not being real, disturbing media content (e.g., a horror movie) can leave a long-lasting, even lifelong, negative impact on the consumer, resulting in *media-induced trauma* [80]. However, a study done by Lin [81] showed that lingering effects of a horror game in VR may not be as common or as intense as one might expect, considering only a small number of participants reported experiencing them the day after the study. Despite these findings, it is advisable to avoid exposing users to uncomfortable content unless it is highly relevant for the specific study. To avoid inflicting physical harm while researching the condition or conducting VR user studies in general, researchers should choose or develop test material based on the state-of-the-art knowledge of design factors that might impact the occurence of cybersickness and other types of discomfort. A compilation of guidelines and useful findings is presented in Table 2.1. Respecting accessibility guidelines for VR applications, such as the ones presented in [82], may also be of great benefit to participants.

2.4.2 QoE assessment methods

At the time of this writing, there is no standardised methodology for assessing the QoE of VR applications (although efforts are underway in the scope of ITU-T Study Group 12 [102]). However, there are a number of instruments that have been used accross various studies addressing the assessment of QoE-related features, such as immersion and presence, as well as side-effects such as cybersickness. Methods/measures used in QoE research can generally be classified as either subjective or objective.

Subjective methods

The use of questionnaires is the most common subjective method used in QoE studies, although it may be supplemented with other methods, such as interviews and diary entries. In most cases, individuals are asked to fill out questionnaires directly related to tested scenarios either during or immediately after testing. Most commonly, users are required to mark their answer on a rating scale.Users may be asked to provide their rating of the overall QoE or its individual dimensions. Instead of using individual questions, researchers often choose to use more established multi-item questionnaires designed to measure a certain aspect (or multiple aspects) of quality.

Certain questionnaires used in QoE research cover a diverse range of features and are intended to be used as a single tool for the evaluation of the overall quality, such as the Game Experience Questionnaire (GEQ) [20, 103] or the Player Experience Inventory (PXI) [104, 105] which are designed for the gaming use-case. Unfortunately, due to the specific characteristics of interactive VR, questionnaires that were initially developed with non-immersive platforms in mind can not be used on their own (i.e., they need to be combined with other measures, which can sometimes be fatiguing for participants and complicates subsequent analysis of results) as they do not include certain aspects that are especially relevant to the VR platform, such as discomfort and cybersickness. This highlights the importance of developing questionnaires that can be used for the evaluation of QoE/UX based on specific features that are relevant for interactive VR. An example of a VR questionnaire that evaluates multiple different features (i.e., general user experience, game mechanics, in-game assistance, symptoms and effects induced by VR) is the Virtual Reality Neuroscience Questionnaire (VRNQ) [106], but its use is limited to VR gaming, rather than VR in general. Tcha-Tokey et al. [107] developed a more general-use VR UX questionnaire comprised of nine subscales: presence, engagement, immersion, flow, emotion, skill, judgement, experience consequence (which measures symptoms of fatigue and cybersickness), and technology adoption. Other questionnaires may be focused on exploring a singular aspect of user experience, as researchers often choose to use questionnaires focused on determining the extent of users' perceived presence (such as the Presence Questionnaire [108, 109], the Igroup Presence Questionnaire — IPQ [110], Slater-Usoh-Steed Questionnaire — SUS [111, 112], or Immersive Experience Questionnaire — IEQ [113]) and/or experienced levels of cybersickness (e.g., Simulator Sickness Questionnaire — SSQ[114], Cybersickness Questionnaire — CSQ [115], Virtual Reality Sickness Questionnaire — VRSQ [116]).

The problem with subjective measures is that they are self-reported and therefore cognitively mediated, which leads to distortions and undermines their validity. E.g., participants often tend to avoid either extreme of the scale (*central tendency bias*), or respond in an excessively positive/agreeable manner (*acquiescence bias*), while further issues stem from the improper or unclear wording of questions themselves. Lastly, since participants' view of the real world is obscured by the VR HMD, their answers are often noted by an administrator, which may influence the participant [117]. Therefore, if possible, subjective assessment questionnaires should be integrated into VEs used for testing [118].

Objective methods

In addition to subjective methods, objective methods (physiological, behavioral, and task performance measures) are often used to assess user experience in a less biased way. Physiological methods are based on measuring different physiological signals such as electrocardiography (ECG), electroencephalography (EEG) and galvanic skin response (GSR). Due to their design, certain medical instruments used to collect this data may hinder user experience and degrade QoE scores, so less intrusive devices, such as fitness bands and smart watches, can also be used for collecting physiological signals [119]. As discussed in [120], the use of psychophysiological measurements in assessing user experience improves existing QoE models, especially in terms of user-related factors, and mitigates issues stemming from the use of self-reported assessments [121, 122]. However, it should be noted that it can be challenging to adequately recognize the affective state of the user based on physiological measures only, as different states may be indicated by very similar physiological symptoms [123, 124] — for example, both excitement and stress tend to increase the heart rate of the user. Furthermore, certain methods for measuring physiological signals appear to be sensitive to noise introduced by head movement (e.g., EEG [125]), while others, such as functional magnetic resonance imaging (fMRI), require complete stillness. Therefore, the results of such methods may not be accurate unless the study happens to be consciously designed in a way that aims to keep the user as stationary as possible. Since head movement in VR is not only extremely common, but also highly encouraged through VR application design, the degree to which the results acquired in stationary conditions can be considered representative of realistic VR use is yet to be determined.

Behavioral methods refer to methods that are based on observing and tracking user behaviors, such as physical movement (e.g., "ducking" to dodge an approaching virtual object [126]) and social interaction (e.g., moving away from an avatar or an embodied agent [127]). To assess user preferences or adaptation mechanisms in the context of VR application use, researchers may decide to track and categorize different actions that the user chooses to perform inside of the interactive VE. In addition to larger bodily movements and conscious actions, researchers may choose to observe more subtle behaviours by incorporating methods such as gaze tracking and emotion recognition, made possible by the growing inclusion of eye tracking and facial recognition technology in more recent headsets.

In general, user performance in multimodal interactive systems, such as VR, encompasses three components [128]: perceptual effort, cognitive workload, and physical response effort. Task performance measures (e.g., time to complete task, measures pertaining to spatial and temporal accuracy) aid in quantifying the effort produced to accomplish a task, and may serve as an objective indicator of the impact of different factors on the users ability to interact with the service in a successful and efficient way. However, to increase the chances of obtaining conclusive and valid results, it is important to choose tasks and measures that are relevant to the observed system/environment.

2.4.3 Test procedure

In accordance with the principle of respect for persons [129], the autonomy of each study participant has to be respected, which means that researchers have the responsibility to provide relevant information about the study and ask for consent prior to actual data collection. After the consent form is signed, a pre-test questionnaire is given as a way to collect personal information about the participant. Similarly to questionnaires used in gaming research [62], pre-test questionnaires used in VR studies usually encompass questions about the basic demographic data (age, sex/gender, profession, ethnicity), as well as inquiries about the skill level and prior experience.

Participants may be inquired about their history of illness, or asked to fill out specialized questionnaires as a way to assess their personality traits, or psychological and/or physiological sensitivities. Along with questionnaires, researchers may choose to include standardized vision acuity tests in their pre-testing process. Acquired information aids in later analysis and interpretation of study results, but it can also serve as exclusion criteria.

Participants should be made aware that they are allowed to pause or terminate the experiment at any time. Instructions regarding equipment, material, and assessment methods should be carefully worded, easy to understand, and presented to each participant in an identical way, which helps mitigate instruction bias [76]. Following the instruction phase, participants are equipped with VR and measurement devices, the positioning of which may require some assistance from the administrator. It is highly advisable to sanitize the equipment (headset, handheld controllers, and any other devices that come into contact with the participant) before each session, which is especially relevant in light of the recent COVID-19 pandemic. If possible, study administrators should provide each user with a disposable mask that provides a barrier between their skin and the headset.

It is advisable to warn participants against operating a vehicle following the exposure to VR content. Although there are no official guidelines at the time of this writing (to the best of the author's knowledge), the duration of the recommended waiting period will likely depend on the intensity of the application and the duration of the VR exposure [130]: several minutes of exposure to a commercial VR application may require only a short 30–45 minute waiting period, while a longer exposure to a flying simulator may require a waiting period of 12 to 24 hrs. The last step before the actual testing phase is a short tutorial session which facilitates adaptation to the application and the technology.

2.4.4 Temporal aspects of QoE assessment

Figure 2.4 depicts time spans of user experience, based on models presented in [131] and [132]. Before the user even starts interacting with the system, they form a set of expectations about the experience (an *internal reference* [133]). E.g., these expectations may form as a consequence of the user's previous experience with a similar system, or they may be a result of the *halo effect*. As the user begins to interact with the system, they perform a series of momentary evaluations of the experience (comparing actual experience to their *internal reference*), based on which they are able to form a reflective evaluation of an episode of use. Repeated use of the system allows the user to make judgements over the span of multiple episodes, and impacts their summative evaluation of the system as a whole. An in-depth analysis of temporal development of QoE is given in [133].

Subjective ratings of reflective QoE are usually collected post-episode via single- or multiple-item questionnaires. During use, the perceived QoE is continuously changing based on the current (momentary) level of quality, and may even increase or decrease drastically in case of

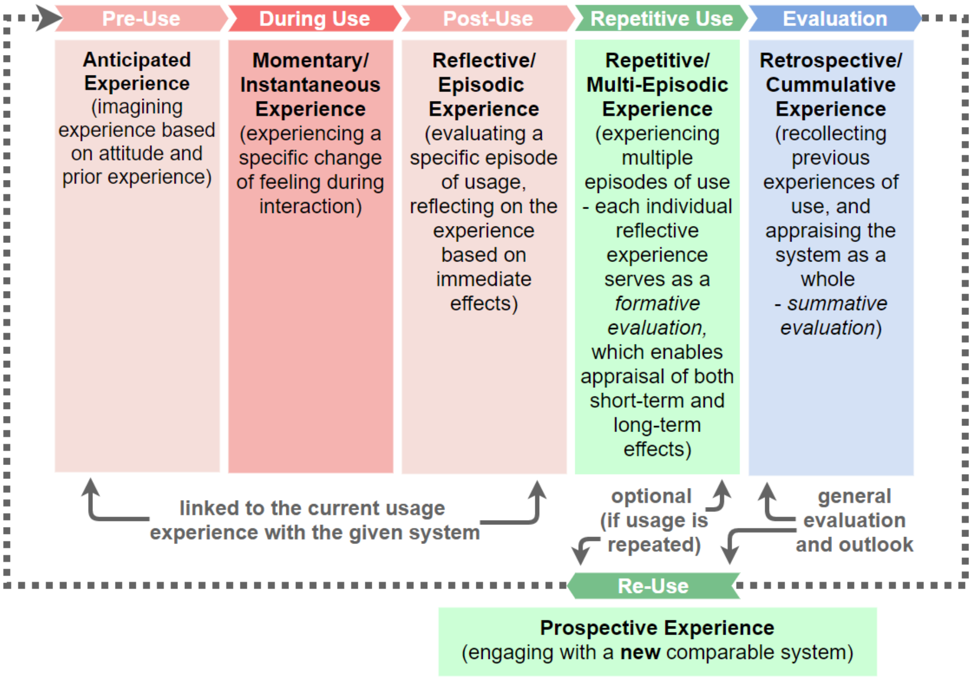


Figure 2.4: Time spans of user experience (adapted from [131, 132])

sudden changes. Researchers may ask the participants to evaluate subjective momentary QoE by assessing the quality of a series of very short (i.e., several seconds) samples which comprise a longer test stimulus, or by continuously reporting the quality of a longer stimulus using a slider or some other type of mechanism that allows for continuous collection of momentary ratings [133]. In the context of user experience with a medium such as VR, which strongly relies on the sense of “being" in the virtual world, divided attention and/or constant interruptions are likely to diminish the level of presence/immersion experienced by the user [134] and thus significantly affect the overall VR QoE. A less obtrusive approach relies on the use of physiological measures with a high sampling rate [135], such as EEG, GSR, and heart rate, while subjective measures are usually reserved for evaluating reflective QoE.

An important issue with measuring reflective QoE is determining the optimal duration of a test scenario. ITU-T Recomm. P.809 [62], which focuses on subjective evaluation of gaming QoE, describes two testing approaches depending on the aim of the user study. A short interactive test, lasting between 90 and 120 seconds, should be adequate for assessing more straightforward QoE features (i.e., quality of interaction). Long interactive tests, usually lasting between 10 and 15 mins, are more suitable for measuring affective states and evaluating complex features such as immersion, presence, or flow. Researchers should also consider VR-specific issues and health risks when determining the duration of VR exposure for user studies. In the context of VR QoE studies, carryover effects may happen with factors/features such as physical symptoms (e.g., cybersickness, eye-strain, fatigue, pain), ease of use, affective states, as well as task performance measures. To a degree, this effect may be counteracted by randomizing test scenarios, while using test tasks that are designed with user comfort in mind (if appropriate for the study) may prevent or reduce physical symptoms. Recommendations for the overall duration of VR exposure differ, with Stanney et al. [136] recommending 15 minutes, and Drachen et al. recommending a 30 minute limit [76].

Karapanos et al. [137] discuss different approaches to collecting samples of user data in the context of repeated use. The *pre-post* approach refers to collecting and comparing participant data twice (i.e., at a point in time which is close to the beginning of the study, and again after a certain time period). The *longitudinal* approach is based on collecting a greater number of measurements. Wilson and McGill [138] highlight the deficit of longitudinal user studies evaluating the use of VR and its consequences. Considering that commercial VR is still in its early stages, there is a lack of knowledge regarding long-term usage and the way it reflects on one's psychological and physiological health. Aside from health related issues, examining VR use over a longer period of time is vitally important for gaining a deeper level of insight about user behaviour and preferences, and the way they change over time.

2.4.5 Environmental aspects of QoE assessment

Even though the goal of every VR experience is to immerse the user into the virtual world, the physical environment of the study remains a relevant aspect of study design. In many interactive VR applications, the user's physical movement translates to movement in a virtual environment (i.e., by moving within the tracked space, the user controls the movement of their avatar). This proves to be a safety issue as VR headsets obscure the user's view of the real world, which can potentially lead to injury and material damage. In order to counteract these threats to participant safety, participants should be supervised at all times, and studies should be conducted in a spacious, uncluttered environment. Certain environmental conditions, such as hot temperatures or high humidity, may increase the likelihood of cybersickness [139]. Thus, it is advisable to keep the space well-ventilated, provide water and snacks[32], and a comfortable place for participants to sit or lie down in case they experience the onset of cybersickness symptoms.

Stepping away from the issue of participant comfort and safety, the location and the overall context of the experiment pose a significant influence on decisions regarding methodology, as well as on the overall outcome of the study and its internal/external validity. Conducting a user study in a laboratory is a very common practice in VR research. Designated laboratories adapted for VR testing are usually spacious, and supplied with advanced VR equipment, which can often be problematic in terms of transportation and setup. Laboratories may also provide access to specialized equipment used to manipulate network conditions or measure physiological symptoms. Furthermore, conducting the study in a specialized enclosed space gives researchers more control over factors such as temperature, humidity, and the allowed number of people, which creates a higher level of comfort (both physical and psychological) compared to a public setting, while the presence of an administrator serves as an additional safety measure compared to non-supervised studies, such as those conducted in participants' living spaces. The most obvious benefit of a laboratory environment, however, is the increased internal validity of the study, which is a result of controlled environmental variables.

Choosing to conduct the study outside of a laboratory requires changes in methodology and duration. These changes can go both ways - compared to laboratory studies, methodology may be more limited in case of public walk-in studies, or more extensive in case subjects are able to participate from the comfort of their homes. Likewise, study duration of field studies varies greatly - for example, a study conducted at a public place/event may have to be shortened to only a few minutes (e.g., [140]), while moving the study to a home setting may even allow for longitudinal research (e.g., [141]).

Public walk-in studies allow for efficient recruitment of participants, but are limited with regard to measures and study duration. Furthermore, participants may feel uncomfortable and exposed in public conditions. Mai et al. [142] elaborate on multiple issues with the public use of VR, such as unwanted touches and the increased likelihood of injury in case of collision with a bystander. Based on these observations, the authors present valuable findings and suggestions on the use of spatial, visual and auditory separation between the person immersed in a VR experience and other people, the inclusion of a supervising person to watch over the user and help them feel more comfortable, and scenario/methodology design that allows the user to slowly ease into the VR experience without feeling too self-conscious. Additional guidelines on how to provide a more comfortable experience for participants using VR in public are presented in [143].

A large percentage of VR applications is intended for personal use in a private space, such as educational institutions and workplaces. In the context of VR gaming, users' homes tend to provide the most realistic study setting. While evaluating the use of VR in home conditions is slowly becoming more achievable, as the number of casual VR users has started to increase over the last several years [138], VR owners are still a definite minority. When conducting from-home studies, the majority of participants will have to be provided with the necessary equipment, which tends to be highly impractical and/or financially straining for research institutions conducting the experiment, especially in terms of more advanced VR systems. Depending on the goal of the study, a more achievable solution may be to focus on mobile VR, which is less expensive, standalone (i.e., does not rely on a VR-ready computer), and easy to set up. However, a more promising solution for this issue may be found in the use of crowdsourcing for QoE assessment [144], leveraging Internet platforms for the recruitment of VR owners for participation in online studies.

2.4.6 Participants

In addition to the issue of recruiting a sufficient number of participants for their study, QoE researchers are faced with the challenge of determining whether chosen participants are a right fit for the study, e.g., based on their demographic or level of expertise. While this holds true for user research in general, due to its obtrusiveness, the VR platform requires a particularly careful consideration of the following factors.

Ethics, health and safety

Madary and Metzinger [145] highlight the importance of pre-screening as a way to remain in compliance with the principle of *non-maleficence*[*], which instructs researchers to construct their experiments in a way that ensures no significant or long-term harm would come to subjects as a consequence of participating in the study. The authors especially warn about the well-being of participants with psychiatric disorders (whether diagnosed or undiagnosed). Depending on the aim of the study, researchers should define appropriate *exclusion criteria* by using specialized questionnaires to assess whether the user has previously exhibited or currently exhibits signs and symptoms of certain disorders that may get aggravated by the experience. Behr et al. [146] suggest screening participants for space-related phobias (e.g., claustrophobia, agoraphobia), as well as other phobias specifically related to the test material.

Lewis and Griffin [147] offer suggestions for screening participants prior to the clinical use of VR. They advise against including participants who are ill with diseases such as influenza, ear infections or ocular defects, suffering from balance disorders and/or taking medication that affects visual or vestibular fuction, currently under the influence of alcohol, or prone to motion sickness or cybersickness. In terms of visual impairment and ocular symptoms, participants may be excluded based on their scores on visual acuity, color-blindness, or stereopsis tests. In general, it is often advised that people who show high levels of sensitivity to cybersickness should not be exposed to VR [146] even in a research setting. However, from the perspective of product developers, including more vulnerable participants allows for a deeper level of insight, which can then be utilized to improve the application or system.

Researchers exploring interactive VR may benefit from examples and guidelines regarding the inclusion of users with disabilities in gaming user research, presented in [76]. In order to adapt the process to the specific needs of each participant, researchers may need to consult them personally, along with medical experts, therapists, and/or caretakers, if necessary. In general, it is very important to keep the whole process of testing as flexible and adaptable as possible.

---

[*] https://www.apa.org/ethics/code/

Diversity and inclusion

At the time of writing, VR is still widely considered to be a *niche* type of technology, owned by a small number of early adopters (e.g., gaming enthusiasts). Due to the relative scarcity of VR systems, samples in studies that recruit locally available participants (as opposed to crowdsourced studies) may be skewed toward novice VR users. While Fairchild et al. [148] noted that novice users may experience VR in a negative way, Hupont et al. [58] attribute positive affective states experienced by test subjects to the novelty of the VR platform. Whether positive or negative, the potential impact of platform novelty on user experience should be considered when choosing test subjects and/or interpreting study results.

VR systems and applications (especially games) are mostly geared toward a younger, tech-savvy audience. On top of that, college students are over-represented in user studies regarding human psychology and behaviour [149]. While recruiting young users for interactive VR studies is likely in line with app/game developers' choices of target audience, this approach may prevent researchers from sufficiently assessing the impact of VR use on other demographics. For example, differences in VR user experience between users of various age/age groups have already been noted by researchers (e.g., [150, 151]). Moreover, the perceived ease of use with VR technology [68] may differ based on age/age group, and age may play a certain role in the susceptibility of cybersickness, illusion of body ownership [152], as well as immersion and presence [153].

Researchers (as well as VR system manufacturers and VR application developers) warn about the unknown impact of VR use on children and young adults with regard to their psychological and neurophysiological development [145]. On the opposite end of the age spectrum, elderly users may find the hardware and software difficult to navigate, especially with regard to specific physical movements, leading to frustration and lowered confidence [76], in addition to being at a higher risk of falling during immersion in VR [154]. But age is not the only factor to keep in mind — perception of presence [155] and susceptibility to cybersickness [35] have been shown to vary between sexes, with research suggesting that VR technology tends to be more adapted to male users. Furthermore, many VR devices and applications fail to accommodate a diverse range of needs, preferences, and abilities of its broad audience, often excluding users with disabilities through inaccessible design.

It is advisable to consider this issue when recruiting participants, either as motivation for extending the sample with underrepresented demographics to better reflect the diverse range of users, or for being deliberate with choosing to focus on homogeneous samples to improve generalizability. Whichever path they choose, researchers should disclose and discuss the implications of sample demographics on study results.

2.5 Chapter summary

Providing the context for the rest of the thesis, this chapter presents an overview of different aspects of QoE for immersive VR applications, listing relevant human, system, and context influence factors, explaining the distinction between QoE features and constituents, and summarizing existing models describing the relationship between various dimensions of the user experience. This chapter also delves into common practices in VR user research, from preparing test material, to choosing QoE assessment methods, outlining the test procedure, deciding on temporal and environmental aspects of conducting user studies, and recruiting participants.

Table 2.1: Guidelines for VR application design to mitigate cybersickness, discomfort, and health risks (taken from [28])

Category	Guideline	Reference
Temporal Aspects	Maintain a framerate of 90 fps or higher.	[76]
	Keep motion-to-photon latency constant at 20 ms or less.	[76, 83]
Camera	Link camera movement directly to HMD movement.	[76]
	Allow users to control camera movement at all times.	[76, 84, 85, 86]
	Avoid incorporating zooming options, as they may worsen cybersickness.	[76]
	Be cautious when changing camera elevation; lifting it high above ground or placing it close to the ground may trigger vertigo or cybersickness.	[76, 83]
Movement/Locomotion	The most effective way to prevent discomfort is to omit movement/locomotion from the application.	[83, 87]
	Avoid rolling motion/rotational optic flow, as it can worsen cybersickness.	[76, 88, 89]
	Identify that acceleration, including deceleration, is the primary cause of discomfort in VR.	[76, 83, 85, 90]
	Implement blinks and snap turns (up to 8 times per second) to disrupt the perception of motion and reduce discomfort.	[83, 91]
	Prefer discrete motion (teleportation) to minimize cybersickness, as continuous motion tends to exacerbate it.	[76, 83]
	Reduce discomfort by avoiding visual motion cues in peripheral vision through techniques like vignetting or virtual enclosed vehicles during continuous movement.	[76, 83, 85, 92]
	Beware of emulating experiences likely to cause motion sickness in real environments, as they can induce cybersickness in a virtual environment.	[76, 93]
	Note that head bobbing and navigating uneven terrain (e.g., stairs, hills) can trigger cybersickness.	[86, 94]
	Understand that forward movement is more natural and comfortable than lateral movement.	[83]
	Enable users to stay in control of their movement, which minimizes cybersickness.	[83, 85, 86, 95]
Embodiment	Ensure the position and orientation of virtual hands align with the user's hands to avoid discomfort due to proprioceptive mismatch.	[83]
	Avoid hand animations that conflict with the user's real movements.	[83, 96]
	Reduce proprioceptive mismatch by omitting virtual representations of arms above the wrist.	[83, 97]
Visual Aspects	Avoid motion blur, blurriness, flickering, and flashing lights/elements.	[76, 83, 85]
	Use darker colors as brighter ones can cause display flicker and visual fatigue.	[83, 98]
	Render visual effects (e.g., particle effects, bloom) to be visible to both eyes with respect to binocular disparity.	[83]
	Be cautious with repeated patterns, textures, and optical illusions of depth, as they may lead to misperception of depth and discomfort.	[83]
	Avoid high-spatial-frequency textures, which can trigger cybersickness and photosensitive seizures.	[83, 99]
	Note that higher levels of detail and realism in visual design can increase cybersickness.	[93, 99, 100]
	Counteract optical distortion introduced by HMD lenses through appropriate distortion correction in the SDK.	[83]
Focus	Incorporate dynamic focus to reduce discomfort, allowing users to focus on objects regardless of their distance or position in the field of view.	[85]
	Maintain a transition focus speed of at least 500 milliseconds in applications with dynamic focus.	[85, 101]
UI	Position UI elements to avoid large, sudden head and/or eye movements by the user.	[76]
	Incorporate UI elements into the scene to prevent them from occluding the scene unnaturally and leading to misperception of depth cues.	[83]

Chapter 3

Proposed conceptual models of QoE for VR gaming

This chapter lays the theoretical groundwork for the QoE models presented in subsequent chapters. After identifying different influence factors that may impact QoE during VR gaming (Section 3.1), conceptual high-level, mid-level, and low-level models of VR gaming QoE are presented (Section 3.2), building a hypothetical map of potential relationships between QoE aspects — QoE IFs, QoE features, and QoE constituents — and the overall QoE.

3.1 QoE influence factors for VR gaming

While key influence factors affecting QoE for immersive platforms have already been identified and classified in ITU-T Recommendation G.1035 [156], this thesis focuses on influence factors for VR gaming. Thus, the existing classification of influence factors, as presented in [156], was further refined and summarized by incorporating information from sources focused on investigating the gaming experience [157, 158] and extending the list of player IFs to include more specific categories pertaining to history of illness, injury, mobility limitations, and current state of the player. Note: the term *player IFs* is used here to further emphasize that proposed models pertain specifically to the gaming use-case, but the term can be used interchangeably with *human IFs* or *user IFs*.

The resulting schematic diagram is presented in Figure 3.1. Compared to [156], the diagram provides additional information regarding whether the influence comes from within the player (internal influence) or outside of them (external influence), while also addressing different levels of abstraction of the observed factors. Recognizing the complex interplay between factors contributing to QoE, IF categories are presented as actors interacting with and impacting one another based on the work by Nacke and Draachen [157]. For instance, players are likely to **perceive** the experience differently based on the context they are playing in, but they may also

influence the said context, e.g., through social interaction with their co-players, altering their environment, etc. Likewise, while system IFs exert an obvious influence on player IFs through their **function**, players also influence the system through providing **input** that affects system behavior, e.g., by making particular choices within the game, providing data that needs to be encoded and transmitted, or adjusting wearable equipment.

Because this thesis pertains to the field of computing, the focus is predominantly on exploring the impact of system IFs. The impact of playing games of different genres with different game mechanics on QoE constituents is presented in Chapter 5, while a more detailed observation of various parameters crucial for the implementation of common VR interaction mechanics is presented in Chapter 6. Furthermore, the impact of network IFs in the context of multiplayer locally rendered gaming is presented in Chapter 7.

However, more abstract influence factors will also be addressed to a lesser extent, as the multiplayer gaming study presented in Chapter 7 includes the examination of past experiences and expertise (player IF) and level of familiarity between co-players (context IF).

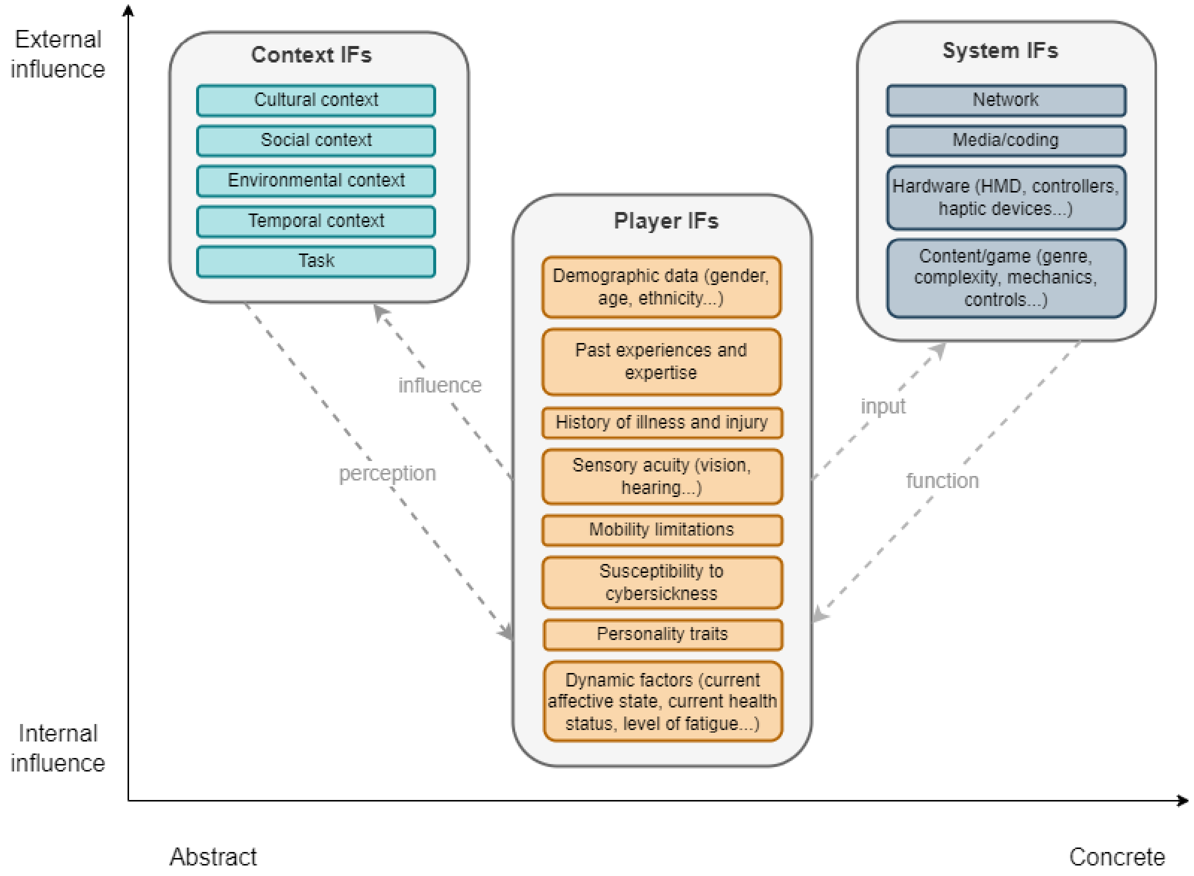


Figure 3.1: A schematic diagram illustrating the classification of QoE influence factors for VR gaming (based on [156, 157, 158])

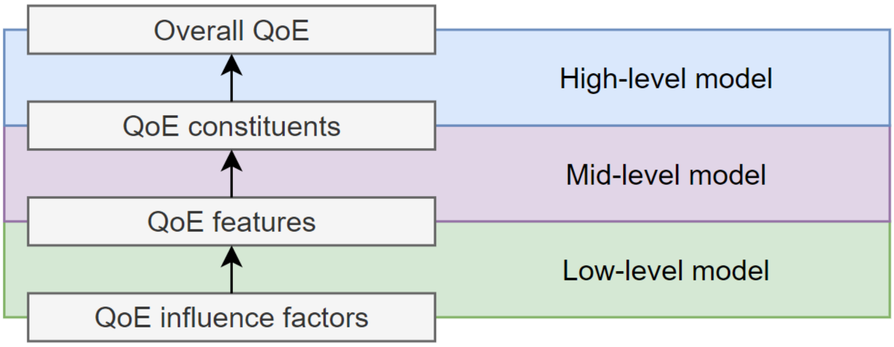


Figure 3.2: Three-tier schema of QoE models

3.2 Conceptual models of QoE for VR gaming

Proposed conceptual models describe the relationships between QoE influence factors, QoE features, QoE constituents, and the overall QoE. Each model is conceived as belonging to a three-tier schema, as illustrated in Figure 3.2. Based on definitions from [8, 61], the relationships between QoE and QoE aspects outlined in proposed models may be described as the following:

- **high-level model:** QoE is comprised of different QoE constituents;
- **mid-level model:** QoE constituents are formed as a result of a cognitive process which aggregates multiple QoE features;
- **low-level model:** QoE features are perceivable, recognized and nameable characteristics of the player's experience occurring as a result of QoE IFs influencing the QoE formation process from the outside of the process.

Proposed models are presented in a top-down order. It is important to note that, due to a very broad range of possible features and IFs contributing to the QoE formation process, presented models are not intended to be all-encompassing. Rather, they are hypothetical compositions of chosen QoE aspects explored in this thesis, with the focus further narrowing with the lower tiers. The latter chapters of this thesis will address and evaluate the relationships hypothesized in the low-level models.

3.2.1 The high-level model of VR gaming QoE

The proposed high-level model (VR_QOE_HLM) of VR gaming QoE describes the relationships between QoE and its constituents, and consists of two sub-models, as pictured in Figure 3.3. Constituents positioned in the central sub-model (referred to as the singleplayer sub-model to streamline the naming convention) are the ones that emerge in all experiences of locally-rendered VR gaming. The singleplayer sub-model is enveloped in a multiplayer sub-model which includes all constituents of the singleplayer sub-model, extending it with additional constituents relevant to a networked multiplayer experience of VR gaming. While the multiplayer

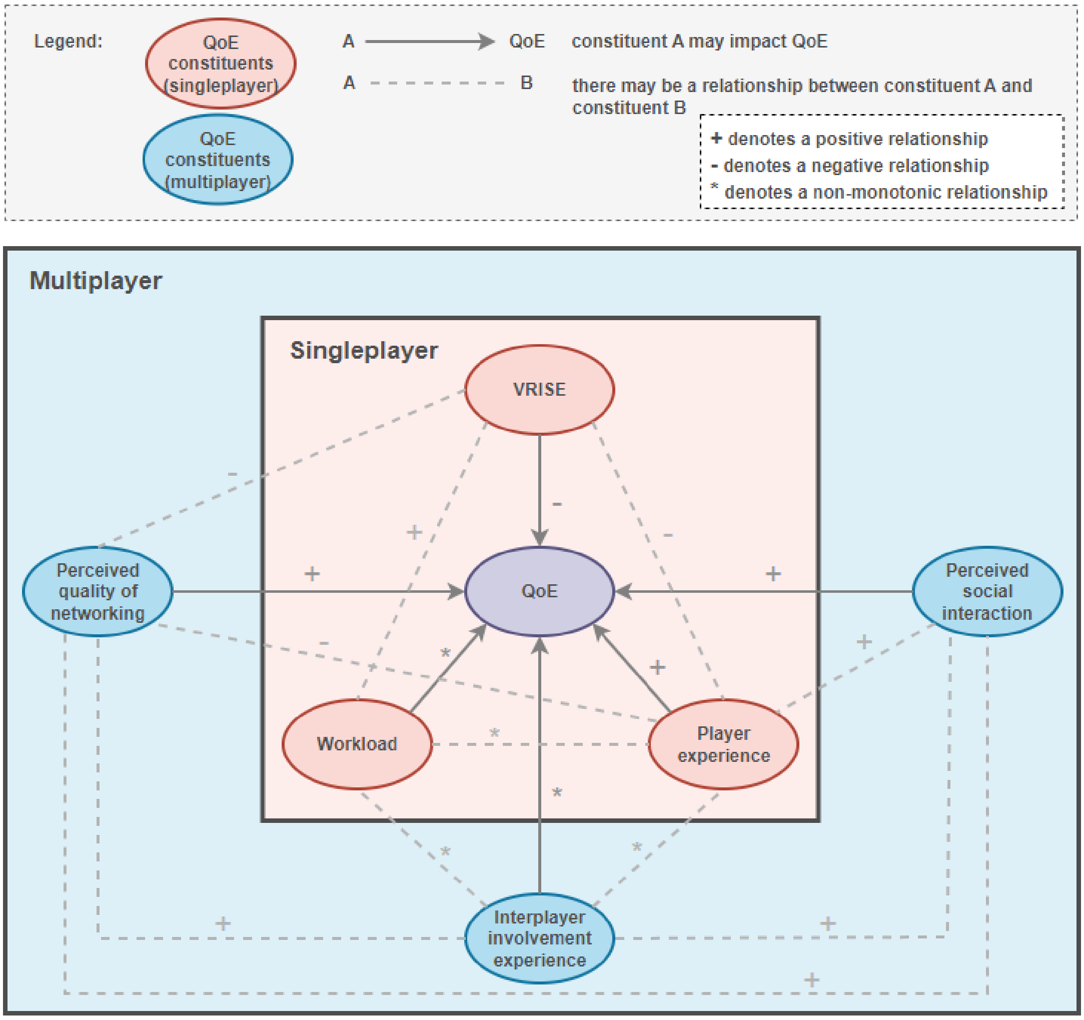


Figure 3.3: High-level model (VR_QOE_HLM) of QoE for VR gaming with hypothesised relationships between QoE constituents and overall QoE

sub-model encompasses all QoE constituents presented in the high-level model of VR gaming QoE, it only pertains to multiplayer experiences. On the contrary, the restricted central model is ubiquitous to all gaming experiences (singleplayer or not).

The singleplayer sub-model consists of three QoE constituents, which can be described as follows:

- **player experience:** based on [19, 63], this constituent encompasses various features, some of which are pertaining to clearly positive (e.g., positive affect, immersion) or clearly negative (e.g., negative affect) effects of VR gaming; common to all platforms, this constituent represents the very core of the gaming experience by encompassing features whose fulfillment is the ultimate goal of every player as they set out to play a game of any kind: to be immersed, entertained, and challenged, experiencing a flow state void of frustration and boredom.
- **workload:** based on [159, 160], this constituent encompasses different types of workload (mental, physical, temporal), problems with task complexity, controls, distractions and

perceptual load, as well as the degree of stress and frustration experienced by the player; in this way, this constituent aggregates the degrees of exertion, fatigue, frustration, and overwhelm arising from a lack of adequate balance between the game's difficulty level and the player's skillset.

- **VR-induced symptoms and effects (VRISE):** named after a term originating in 1999 [161], this constituent pertains to symptoms and effects that occur when the obtrusiveness of the VR platform develops beyond a tolerable threshold; with many diverse manifestations ranging from perceptual and cognitive effects (e.g., slowing down of reaction time) to changes in physiological functions (e.g., excessive sweating or increased heart rate), VRISE can also manifest as postural instability, muscle pain and fatigue, eye strain, discomfort caused by HMD design, even media-induced trauma and physical injury.

The multiplayer sub-model extends the singleplayer sub-model by three additional QoE constituents:

- **interplayer involvement experience:** a new term coined for the purpose of this thesis, this constituent references players' perceptions, feelings, and opinions pertaining to the interaction between players as directly dictated by the rules and goals of the game; it encompasses feelings that arise during collaboration and/or competition with another player or players, such as enjoyment of competition, enjoyment of collaboration, willingness to aid another player, desire to win, malicious delight at another's loss, vengefulness, jealousy at the skills of another, perception that co-players or opponents are of a satisfactory skill level, etc.
- **perceived social interaction:** this constituent describes players' perceptions, feelings and opinions pertaining to their interaction with the other player or players during gaming, but in a way that is not fully dictated by the rules and goals of the game, although they may pose a certain influence by providing a context in which social interaction is experienced; encompassing features such as perceived quality and quantity of communication (e.g, via voice chat), feelings of co-presence, togetherness, and social bonding, this constituent transcends features addressed in the interplayer involvement experience constituent to focus on the concept of gaming as a social activity.
- **perceived quality of networking:** this constituent pertains to players' awareness of any network impairments, encompassing features such as input smoothness and responsiveness, perception that actions of the self and/or other(s) are delayed as a result of suboptimal network performance, and perception that gaming performance suffers as a result of network impairment (e.g., shooting precision plummets as a result of latency); it is important to note that this constituent is significantly more detrimental to cloud gaming experiences, as opposed to locally-rendered gaming experiences explored in this thesis.

In line with the explanation of QoE constituents in [8], the purpose of this model is not to present QoE constituents or features as orthogonal dimensions. Instead, a degree of overlap between constituents is expected. Constituents may therefore have a positive relationship because they represent similar aspects of the experience or they may even be connected by a somewhat causal relationship (e.g., improving network quality facilitates social interaction). Likewise, others may have a negative relationship (e.g., experiencing an increase in VRISE is likely going to negatively affect player experience, by impairing its positive QoE features such as immersion or positive affect). Some constituents may even be described as having a non-monotonic relationship with each other or the QoE, following the so-called ideal-point model [61, 162]. For example, a slight increase in workload may result in a more positive experience, but increasing the workload too much may detract from it by increasing feelings of frustration and annoyance. Similarly, desire to win as part of the interplayer involvement experience may also increase player experience, but if it goes too far the element of fun may suffer as a consequence. Hypothetical relationships (positive, negative, or non-monotonic) between constructs, as well as between constructs and QoE, are denoted in Figure 3.3.

3.2.2 The mid-level model of VR gaming QoE

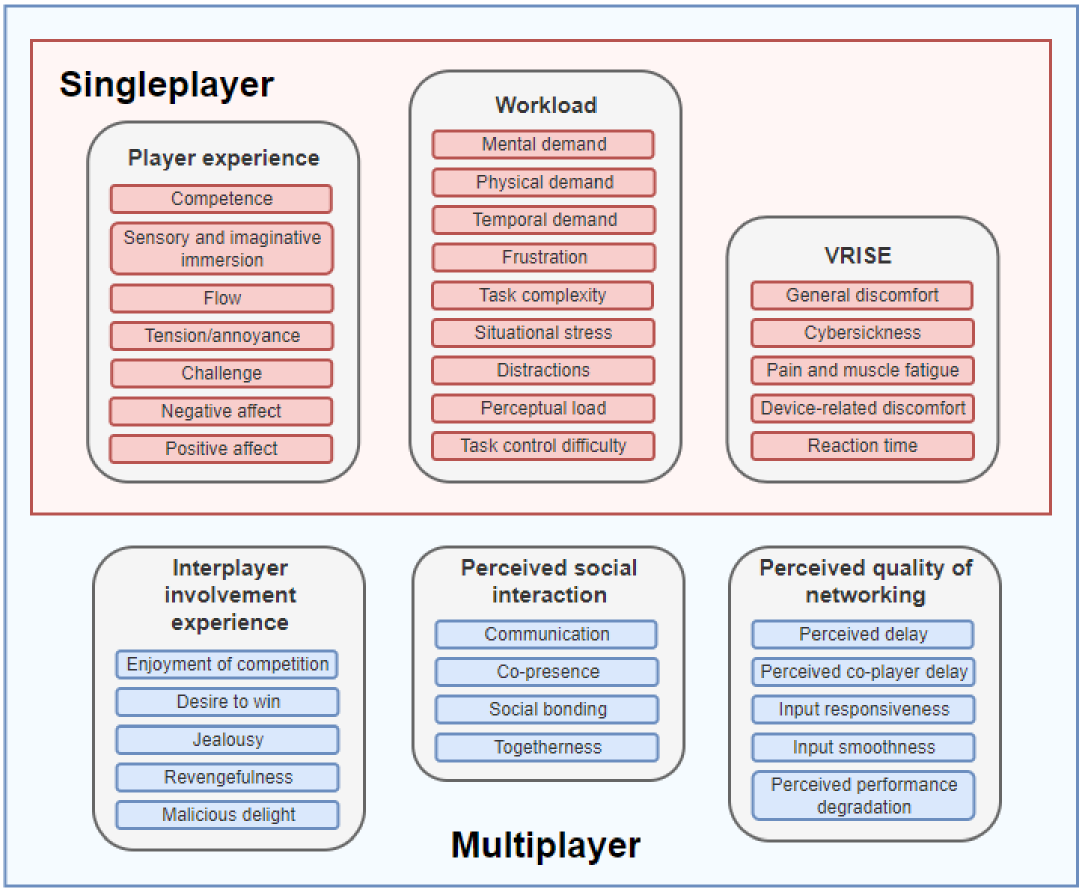


Figure 3.4: Mid-level model (VR_QOE_MLM) of QoE for VR gaming depicting the aggregation of QoE features into QoE constituents for a competitive multiplayer scenario

Figure 3.4 summarizes the QoE features that are aggregated in each QoE constituent for the particular example of a competitive multiplayer game. It is important to note that, while each constituent is an amalgamation of different features, the contributions of each feature are not expected to be the same. For example, some features may be more relevant for a particular gaming setup than others. Moreover, some features are positive contributors (as the feature increases, so does the constituent), while others are negative contributors (as the feature increases, the constituent decreases), or even non-monotonic contributors, in case the relationship between feature and constituent is not unidirectional.

The hypothetical contributions of presented features on constituents presented in the mid-level model (VR_QOE_MLM) are the following:

- **player experience:** tension/annoyance and negative affect are negative contributors, competence and challenge are non-monotonic contributors, sensory and imaginative immersion (not to be confused with immersivity as a system property), flow and positive affect are positive contributors;
- **workload:** all features are positive contributors;
- **VRISE:** all features are positive contributors;
- **interplayer involvement experience:** enjoyment of competition and malicious delight are positive contributors, other features are non-monotonic contributors;
- **perceived social interaction:** all features are positive contributors;
- **perceived quality of networking:** perceived delay, co-player delay and performance degradation are negative contributors, input responsiveness and smoothness are positive contributors.

3.2.3 The low-level models of VR gaming QoE

Moving down to the lowest tier and narrowing the topic toward the focus of this thesis, this subsection presents three partial conceptual models, each presenting hypothetical links between chosen IFs and QoE features. By seeking statistically significant differences among various scenarios and identifying statistically significant correlations between IFs and QoE features through several user studies, we have supported some of these hypothetical relationships using statistical methods. The resulting models are presented in the latter chapters of this thesis.

The first low-level model (VR_QOE_LLM_1), illustrated in Figure 3.5 and explored in Chapter 5, presents hypothetical links between game genres with different game mechanics and QoE features contributing to two QoE consituents: workload and VRISE. The second low-level model (VR_QOE_LLM_2), illustrated in Figure 3.6 and explored in Chapter 6, presents hypothetical links between specific implementations of three types of common VR interaction mechanics (slash, pick-and-place, and shoot) and QoE features contributing to three QoE constituents: player experience, workload, and VRISE.

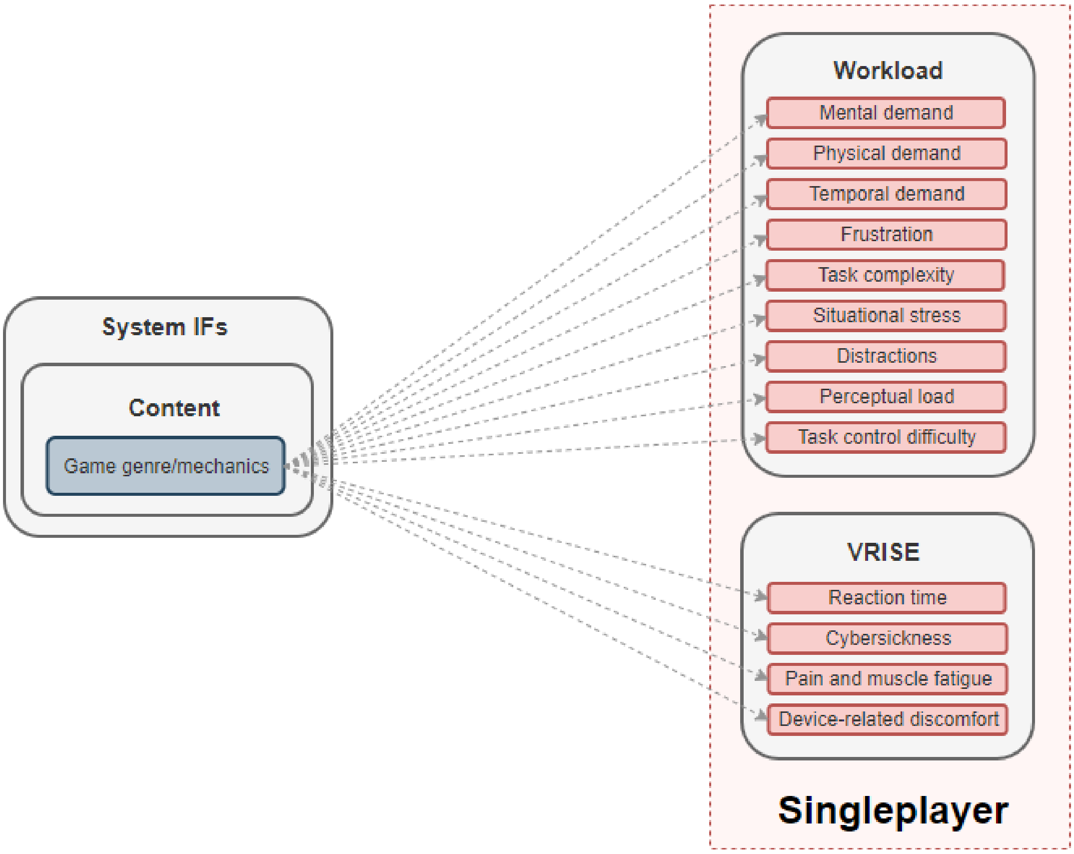


Figure 3.5: Partial low-level model (VR_QOE_LLM_1) of QoE for VR gaming depicting the hypothesized relationships between game genres/mechanics and chosen QoE features for a singleplayer scenario (analysis of hypothesized relationships is provided in Chapter 5)

The third low-level conceptual model (VR_QOE_LLM_3), illustrated in Figure 3.7 and explored in Chapter 7, presents hypothetical links between network factors, social context, and past experiences and expertise of the player, and different QoE features contributing to different constituents that shape the competitive multiplayer experience. While all of these relationships are considered for two separate games of different genres and mechanics, the inclusion of game genre and mechanics as an explicitly listed influence factor was omitted from this illustration. This decision was made to ensure clarity and maintain a stronger emphasis on other important factors.

3.3 Chapter summary

This chapter presents different categories of QoE influence factors relevant for the use-case of VR gaming, namely influence factors pertaining to the context, the system, and the player. The chapter also presents a three-tier schema of models which investigates relationships between different QoE aspects, from QoE influence factors on the lowest level and the overall QoE on the highest level. Starting from the high-level model, the chapter defines six different QoE constituents, from those ubiquitous to all gaming experiences to those characteristic of multiplayer gaming. Further deconstructing these constituents, different QoE features are presented and illustrated with conceptual models exploring their relationship with QoE IFs. This lowest tier of models serves as a hypothetical foundation to be further supported with experimental data over the course of this thesis.

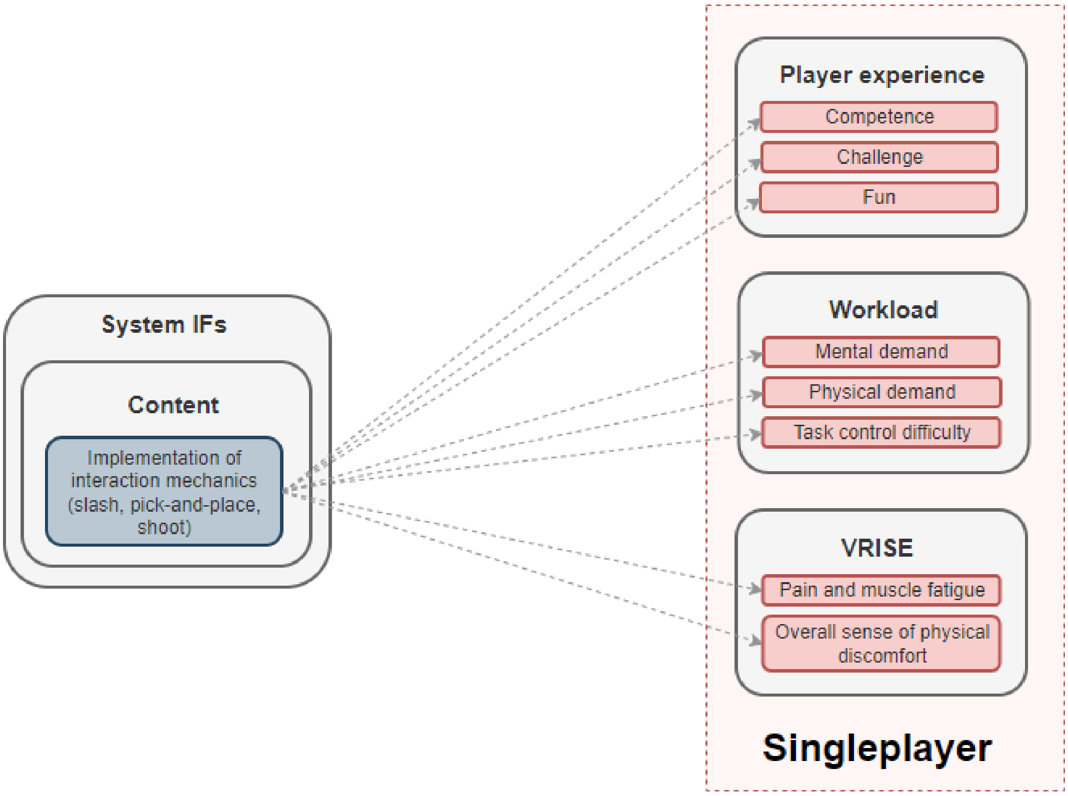


Figure 3.6: Partial low-level model (VR_QOE_LLM_2) of QoE for VR gaming depicting the hypothesized relationships between the implementation of interaction mechanics and chosen QoE features for a singleplayer scenario (analysis of hypothesized relationships is provided in Chapter 6)

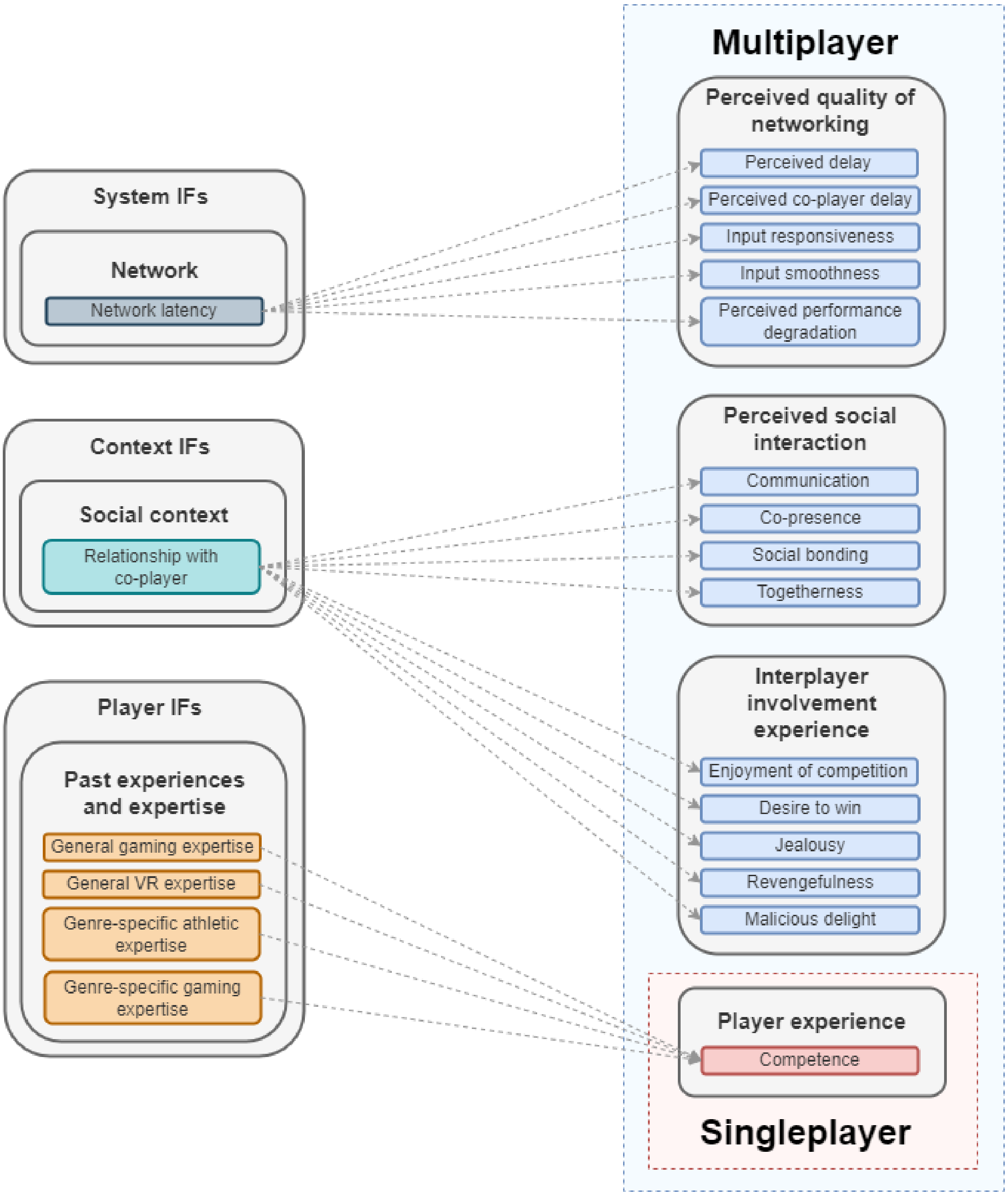


Figure 3.7: Partial low-level model (VR_QOE_LLM_3) of QoE for VR gaming depicting the hypothesized relationships between network factors, social context, and past experiences/expertise, and chosen QoE features for a competitive multiplayer scenario (analysis of hypothesized relationships is provided in Chapter 7)

Chapter 4

Defining and classifying interaction mechanics for VR games

4.1 Introduction

Despite stepping away from the topics of QoE assessment and modeling to focus on game design and 3D user interfaces, this chapter continues to lay further theoretical groundwork for the design of conducted studies and the interpretation of their results. Starting from existing definitions of game mechanics, the chapter proceeds to introduce a new term, *interaction mechanics*, used to describe the way users (players) interact with the game system. This is followed by a proposed novel taxonomy of interaction mechanics commonly used in commercial VR games. The text and figures laid out in this chapter have been published as part of a journal article presenting the framework for the classification and evaluation of game mechanics for VR games [10].

4.2 Game mechanics and interaction mechanics

Multiple definitions of the term *game mechanics* can be found across literature. For example, Järvinen [163] describes game mechanics as *means* afforded by the game, to be used for interaction with game elements as the player attempts to reach the goal of the game, which is determined by its rules. Sicart [164] defines game mechanics as *"methods invoked by agents, designed for interaction with the game state"*. Both Järvinen and Sicart argue for the formalization of game mechanics as verbs (e.g., *aiming*, *shooting*, etc.)

On the other hand, Fabricatore [165] defines game mechanics as interactive subsystems that are based on rules, but only presented to the player as a black box that generates a certain output upon receiving a certain input. According to this definition, game mechanics are referred to as tools used to perform gameplay activities and described as state machines. As the player triggers different interactions with the mechanics, the state of the machine will change. The author gives an example of the door mechanics. Triggered by the appropriate interaction (*unlock door*) the door mechanics will change from the locked state to the unlocked state. In the work of Fabricatore, the verbs used in relation to game mechanics are referred to as either interactions (actions that trigger state change) or gameplay activities (activities carried out through the use of game mechanics), as opposed to game mechanics themselves, which highlights the difference between this approach and the approaches taken by Sicart [164] and Järvinen [163].

Regardless of their differing definitions, all of the aforementioned authors make it a point to distinguish between different mechanics depending on their relevance and frequency of use within the observed game. Most notably, each of the source papers presents some version of a definition for the term *core mechanics*. Fabricatore [165] defines them as mechanics that are used to execute gameplay activities that most frequently occur within a certain game, and are essential for its successful completion. Järvinen [163] defines core game mechanics as those that are available throughout the entire game and notes that a game can have multiple core mechanics, but they have to be performed *"one at a time"* (i.e., during a turn or a certain game state), while Sicart [164] describes them as mechanics utilized by agents to reach a *"systemically rewarded end-game state"*.

In this chapter, verb-based names are used to refer to game mechanics, but considering the existence of well-established monikers for certain mechanics that are colloquially expressed in root/imperative verb form (e.g., *hack and slash*, *pick-and-place*, etc.), unlike Järvinen [163] and Sicart [164] the gerund form (e.g., *shoot**ing***, *slic**ing***) is not used. Bearing in mind the extent to which games in virtual reality (more so than any other platform) are based on immersing the player in the midst of the action, the use of the term *game mechanics* in this thesis predominantly addresses methods invoked by the player, which is consistent with the anthropocentric approach taken by Järvinen [163]. However, as methods invoked by other *agents* (as highlighted by Sicart [164]) largely influence the way in which the player is able to utilize core game mechanics within a certain game or even a particular moment within that game, there is a need to further distinguish player-invoked mechanics from other mechanics available in the game.

As explained by LaViola Jr. et al. [166], *interaction techniques* are methods that allow the user to perform a certain task. The use of this term primarily refers to the user's input and the mapping between that input and the system, but may also extend to those elements that provide feedback to the user as they interact with the system. While not exclusively related to the gaming use-case, the concept of interaction techniques as a mapping between user input and the system resembles the description of game mechanics as a link between behavioral and systemic elements of the game, as explained by Järvinen [163]. Borrowing from this definition, a new term — *interaction mechanics* — is introduced in this thesis to refer specifically to game mechanics that are directly and solely controlled by player input (e.g., for a VR shooter game, the *primary* interaction mechanics can be described as expelling a bullet from a gun — an action that provides immediate visual, auditory, and haptic feedback to the player — in a specific direction determined by controller position in a particular moment determined by a trigger press), as opposed to other methods present in the game but invoked by agents other than the player (e.g., enemy behaviour controlled by artificial intelligence).

Furthermore, it is important to distinguish between different types of core interaction mechanics. Both Järvinen [163] and Sicart [164] elaborate on the concept of *primary* mechanics as core game mechanics that remain consistent throughout the game and are directly utilized to overcome challenges that contribute toward the end-game state, as opposed to other types of core mechanics, referred to as *secondary* mechanics [163] or *submechanics* [164], which play a supporting role to the primary mechanics, generally only indirectly contributing toward the end-game state. Järvinen [163] also introduces *modifier* mechanics, only available under certain conditions, giving the example of *"applying strength"* to a tennis hit as a modifier mechanic to the primary *hit* mechanics, considering that it has to be performed in a specific moment and a specific place. Järvinen also lists moving into appropriate position and aiming as examples of secondary mechanics to the primary *shoot* mechanics. This level of granularity is not used in the analysis of VR game mechanics conducted in this thesis for the following reason: focusing solely on methods that fit this narrow definition of *primary* mechanics (e.g., to *shoot* or *hit* an object) is unlikely to provide information that is necessary to draw relevant conclusions regarding user experience. From the perspective of the player, the specific action of shooting or hitting is immediately preceded by more complex perceptual, cognitive, and motor operations that strongly contribute toward various aspects of user experience. For example, the perception of aspects such as challenge, competence, or exertion in the context of the *shoot* mechanic is formed based on the complexity of actions taken by the player in preparation for the eventual *shoot*. The player's ability to notice and recognize the target, the complexity and speed of movements necessary to align the weapon with the target and press the trigger at precisely the right moment in time, followed by immediate feedback that signals that the intended action has been recognized by the system — these elements, along with their ultimate outcome, arguably contribute toward the formation of user experience far more than the isolated action of shooting. Therefore, to better fit the context of user experience evaluation, the classification provided by Järvinen [163] is slightly modified for this purpose. Instead of considering such actions as secondary or modifier mechanics, and therefore separate from the primary mechanics, this thesis extends the definition of primary mechanics to implicitly include preparatory mechanics such as aiming, positioning, and applying force (if required) as its integral elements.

Interaction mechanics may also be categorized as proactive (i.e., instigated by the player; the moment of instigation is thus decided by the player) or reactive (i.e., performed by the player as a direct reaction to mechanics instigated by game entities other than the player; the player is thus expected to react in the moment determined by the actions of those entities). For example, in a combative action game, interaction mechanics that are usually performed as an attack (e.g, swinging a sword to hit an enemy) can be classified as proactive mechanics, while interaction mechanics generally performed in defense (e.g., raising a shield to protect against enemy attacks) can be classified as reactive mechanics. Another example of the distinction between proactive and reactive mechanics can be found in sports games, such as throwing (proactive) and catching (reactive) a ball. In this thesis, the focus is placed mostly on proactive mechanics, with the goal of examining the interaction between the player and a particular target object, rather than addressing the complex interplay between the player and other entities of the game, which is considered to be out of scope for this thesis.

It is important to note that throughout this thesis both terms (*game mechanics* and *interaction mechanics*) will be used as a way of distinguishing mechanics controlled by the player (interaction mechanics) from the amalgamation of all game mechanics that influence the player in commercial games (game mechanics). For example, in Chapter 5, commercial games serving as test material for Studies 1 and 2 were chosen based on their primary interaction mechanics. However, it is impossible to fully distinguish the exact impact of player-instigated actions on user experience from the impact of actions instigated by the game system. The comparatively simpler test material described in Chapter 6 lacks the complex game system responses to participant actions. For example, in this case there are no AI-controlled entities to interact with, no counter-attacks, no changes in game behavior depending on the success of the participant, no visible progress feedback. Thus, the low-level model explored in Chapter 5 lists game genre/mechanics as the observed influence factor, while the low-level model explored in Chapter 6 lists the implementation of interaction mechanics as the observed influence factor.

4.3 Taxonomy of interaction mechanics used in commercial VR games

This section highlights several aspects to consider when attempting to classify primary interaction mechanics for virtual reality games. The resulting taxonomy is based on different characteristics of the virtual hand metaphor implementation, features of used tools and target objects, target object placement, as well as specific features of the task itself.

4.3.1 Interaction fidelity

VR games can vary in terms of *interaction fidelity*, or the extent to which they emulate the actions from the physical world. As discussed in [167, 168], interactions in 3D environments are often designed as either magical or literal. The literal approach to interaction design aims for a highly convincing reproduction of real-world interactions. On the contrary, the magical approach aims to improve the functionality of the interaction mechanics by providing the player with abilities that transcend the constraints of the physical world. Between the two extremes lies hypernaturalism, which combines the intuitive gestures of literalism with magical enhancements or *“intelligent guidance”* [169].

4.3.2 Symmetry and synchrony

As discussed in [166, 170], different tasks may require the use of one (i.e., unimanual task) or both (i.e., bimanual task) hands. In case of bimanual tasks, both hands may perform the same motions (i.e., bimanual symmetric task) or they can move in different ways (i.e., bimanual asymmetric task). Furthermore, tasks can be classified based on whether movements of both hands (symmetric or asymmetric) occur at the same time (i.e., bimanual synchronous task) or at different times (i.e., bimanual asynchronous task).

4.3.3 Targets

In general, when the user interacts with the virtual environment using hand-held controllers tracked in 6DoF with a one-on-one mapping between the controller and the virtual hand, this interaction can be considered as an interplay (either direct or achieved through the use of proxy objects) between the virtual hand of the user and the current target. Depending on the game, the *target* can take different forms. Moreover, the idea of what can be defined as a target may change from one moment to the next as the player works their way through different mechanics within a single game.

To illustrate the fluidity of this idea of an in-game target, it is useful to call upon the concept of *components*, as introduced by Järvinen [163], who defines them as objects within the game that can be manipulated and owned by the player. According to the author, the concept of ownership is highly relevant for the classification of components — depending on the current owner (the player, i.e., the *self*, other players, or the system itself), a component may be considered a component-of-self, a component-of-others, or a component-of-system. The process of obtaining ownership of a component is often instigated by some type of player interaction, e.g., by invoking the *pick up* interaction mechanics. This mechanics requires the player to first perceive the component, and subsequently perform cognitive and motor operations that are necessary for the successful acquirement of the component — during this process, the component is perceived as a *target*. Once obtained, it is controlled by the player and becomes the component-of-self — and the player’s perception of what they consider to be a current target begins to shift. For example, in case of a cooking simulator, the player may be required to pick up a sandwich and place it on the plate. In the context of the *pick up* mechanics, the *sandwich* will be perceived by the player as a target object. Once obtained, the sandwich becomes a component-of-self. As the player progresses to the *place* mechanics, the plate becomes the target. The concept of components does not pertain only to static objects, but to in-game characters as well. For example, an action game may require the user to operate a weapon (the component-of-self), directing it in a way that causes damage to enemies controlled by other players (characters-of-others) or the system itself (characters-of-system). In doing so, the player considers the other characters to be targets.

Static in-game objects and dynamic in-game characters are typical examples of what is referred to as *explicit targets* in the scope of this thesis. Their saliency with regard to the surrounding environment serves as an indication of their role in some sort of in-game interaction and their boundaries are clearly defined by sensory cues given to the player (predominantly visual, but commonly aided by other modalities). While static objects and characters are generally embodied — they serve as independent entities and are often subject to physical forces — an explicit target does not necessarily need to take the form of a three-dimensional entity with physical properties, as long as the interactable area/volume is circumscribed and its boundaries are clearly indicated to the player. For example, to score points within a basketball game, a ball needs to make its way to an explicit target. By itself, a basketball hoop, a salient physical object, is not actually a target. Instead, it serves as an indication of the position as well as the limits of the actual target — an empty space encircled by it. However, depending on the game, certain interactions with the surrounding virtual environment are performed without an explicit target, although cognitive and motor operations necessary for aiming and positioning are still being performed. For games that rely on *implicit targets*, having to first determine the position/direction of an implicit target does not only serve as a necessary precursor to further action, but is usually considered as being one of the fundamental parts of the challenge. An example is given in tennis, where the strategy of each player relies on attempting to direct the ball away from the other player. Although the existence of boundaries (e.g., the height of the net, the dimensions of the field, the position of the other player) is common — determining what is allowed by the game, as well as what constitutes a successful attempt — what is considered as an implicit target ultimately comes down to the player and their strategy. In other words, the game may provide explicitly defined boundaries determining the *"pool of possibilities"* afforded to the player, but the exact placement of the implicitly defined target is familiar only to the player performing the action.

4.3.4 Mediation and interaction space

Depending on the type of task and the position of the target with regard to the user, performed interactions may be direct (i.e., non-mediated) or mediated through the use of some type of hand-held tool, an object that, once obtained, becomes a component-of-self, and can be utilized by the player to perform (or facilitate the performance of) a certain task. Considering that manual interactions in VR rely on the freedom of physical movement in the real world, the type of task and tool usage within the game will define the possible placement of target objects within the virtual environment, as to accommodate the limitations imposed on the player based on the characteristics of their body and the boundaries of the tracked physical space. In VR games relying on the virtual hand metaphor with a one-to-one mapping — just like in the "real world" — the player can only interact with objects positioned within the immediate space that surrounds them, or the so-called peripersonal space. Everything beyond the player's peripersonal space is therefore, quite literally, outside of their reach, belonging to what we call their extrapersonal space. However, there is a certain flexibility to the boundaries between peripersonal and extrapersonal space, as the representation of peripersonal space may extend to a bigger area, for example in case of tool use [171]. Since the focus of this thesis lies elsewhere, the definition of peripersonal space will not be explored in more depth. The term, as used in this thesis, along with the concept of extending the singular peripersonal space beyond its initial boundaries via tool use — which may be considered an oversimplification [172, 173]— only serves as an aid for the broad classification of different VR interaction mechanics.

Further focusing on the issue of non-mediated vs. tool-mediated interaction mechanics, some types of handheld tools, commonly used in VR games (e.g., swords, clubs, bats), may be considered significant extensions of peripersonal space, enabling the user to interact with objects that are positioned in what can be described as near extrapersonal space. Thus, while non-mediated interactions only allow the player to act within the physical boundaries defined by the dimensions of the player's body and the limitations of tracked physical space, tool-mediated interactions increase the interactable area by a small margin, based on the length of the tool in question. However, in cases in which the virtual environment covers a broader area compared to the tracked physical space, target objects positioned further from the player may still remain physically unreachable. Obviously, this is not necessarily an issue — in fact, positioning target objects somewhere in the distant extrapersonal space can almost be considered a prerequisite for the implementation of certain game mechanics. For example, the ever popular shooter genre is inherently defined by tool-mediated interactions with distant targets. Furthermore, a significant percentage of sports — which are commonly emulated in VR games — are based on repeated and rule-governed actions that can be described as throwing, shooting, or hitting objects toward a distant target or targets. Therefore, we consider tool use and target positioning to be interconnected, both serving as key features based on which we choose to categorize different game mechanics. For the sake of simplicity, in this thesis *peripersonal space (PPS)* will be referred to as the space in which the user is able to directly interact with objects without the use of external tools. The space in which the user is able to physically interact with objects using handheld tools will be referred to as *extended peripersonal space (EPPS)*. The remaining area, which spreads beyond the boundaries of extended peripersonal space and is therefore not directly reachable, will be referred to as *distant extrapersonal space (DEPS)*. A simplified illustration of these concepts is provided in Fig. 4.1).

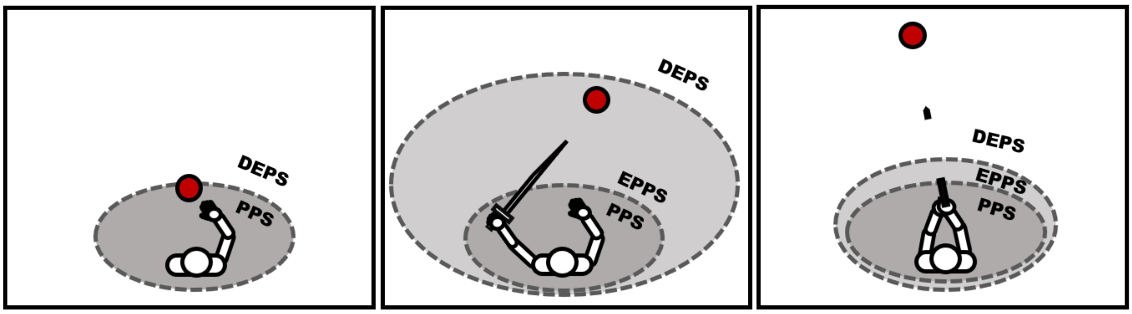


Figure 4.1: In-game interaction with a target object placed within the peripersonal space (PPS; left), extended peripersonal space (EPPS; middle) and distant extrapersonal space (DEPS; right); taken from [10]

To better describe the notion of tools as extensions of peripersonal space, it is useful to refer to the concept of the body as a kinetic chain, comprised of rigid segments connected together with joints. Physical movements at the core of various sports activities are usually performed with the objective of optimizing acceleration and speed at the terminal segments of the kinetic chain [174, 175] (e.g., foot in case of soccer, hand in case of volleyball). Hand-held controllers used with commercial VR equipment provide the system with an approximation of relevant terminal segments (i.e., hands) at any given moment. By providing the player with a hand-held tool, the virtual representation of terminal segments of the kinetic chain is being extended, either by modifying the size and shape of the original terminal segment, or by adding a new terminal segment, connected to the original terminal segment with a new joint. A sword attached at the end of a controller, fixed at a certain angle, but moving together with the controller as a single, rigid entity, connected to the rest of the chain by the wrist, serves as an example of the former approach. The latter approach is taken by Fletcher [176], who argues for the inclusion of an additional spring joint to separate the VR sword or other type of tool from the virtual representation of the controller, with the goal of presenting the player with a more convincing implementation of force feedback. While the implementation of interaction mechanics mediated by the use of swords, bats, and other types of rigid tools may include a single added joint, including a flexible tool (e.g., a chain, a flail) may even add multiple links to the original terminal segment of the kinetic chain. In addition to the fact that a hand-held tool extends the interactable space around the player to the extent afforded by its length, its dimensional and other physical properties determine its inertia and velocity profile in the context of the in-game physics system, contributing to the overall result and potentially affecting the overall perceived realism, utility and enjoyability of the game mechanics. Examples of different interaction mechanics and tools as extensions of peripersonal space are presented in Figure 4.2.

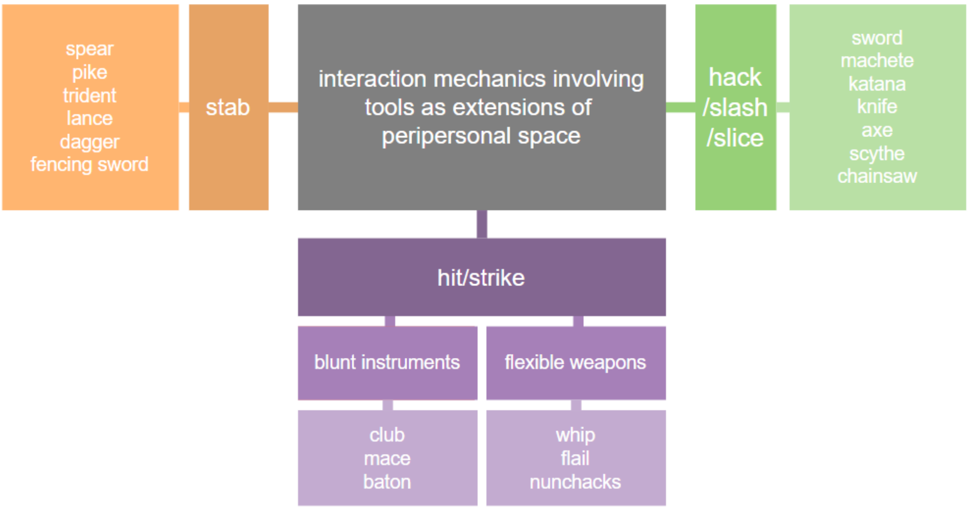


Figure 4.2: Specific examples of interaction mechanics based on tools as extensions of peripersonal space (taken from [10])

Whether talking about sports equipment or weapons, there is a common thread among tool-based interaction mechanics that involve targets primarily positioned in the distant extrapersonal space — their design usually relies on the implementation of projectile motion. A projectile is an object that has initially been propelled (launched) by a certain force, and is subsequently continuing to move along a certain path (i.e., the ballistic trajectory) under the influence of gravity and other external forces. While projectiles are very common in games, the gaming use case does not call for a high level of realism, meaning that projectile motion may be calculated based on simplified ballistic flight equations that often disregard the influence of forces such as air resistance and cross wind. A projectile does not necessarily need to be launched using a tool such as a gun or a bat — it may also be hand-thrown. Whether to choose to label this category of mechanics as tool-mediated is up for debate, as a projectile may be considered a tool in itself and often serves as an intermediary object that provides the means for the player to impact the target, but for the sake of simplicity such cases will be referred to as projectile-based non-mediated interaction mechanics, as to separate this type of mechanics from those that rely on a hand-held tool to propel the projectile.

Depending on the game's genre and its core mechanics, as well as the particular tool (or lack thereof) used for propelling the projectile, there may be significant differences in the extent to which the actions performed by the user can impact its trajectory, as mechanisms responsible for projectile propulsion vary significantly. With firearms, the angle of propulsion will depend on the position of the user's controller, but the magnitude of the muzzle velocity (i.e., the speed of a projectile as it exits the barrel) will remain constant, as determined by the implementation of the particular weapon of choice. With mechanisms based on elastic propulsion (e.g., a bow) the initial velocity of the projectile depends on the draw weight — the amount of force exerted on the bowstring (or the elastic band in case of a slingshot) as it gets pulled back in preparation for

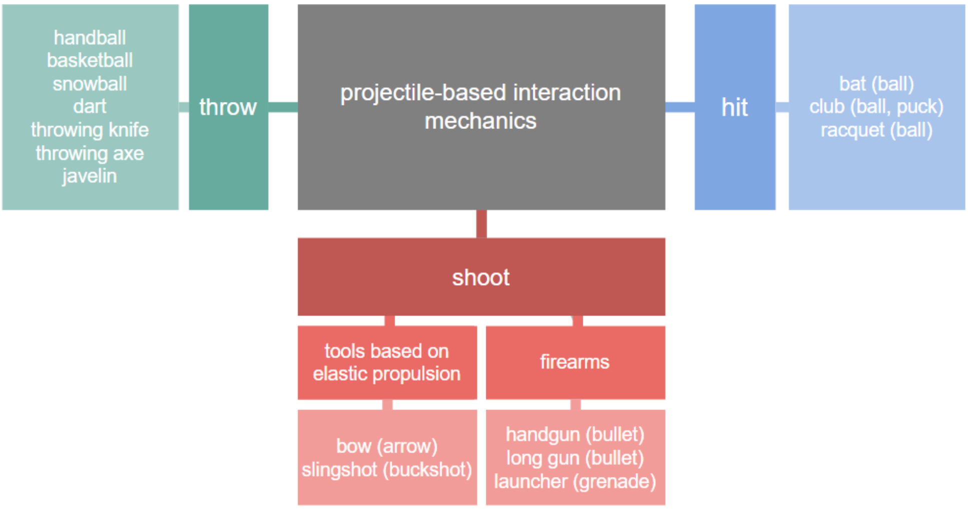


Figure 4.3: Specific examples of projectile-based interaction mechanics (taken from [10])

the projectile release. In the context of elastic propulsion weapons as implemented in controller-based virtual reality applications, at the moment of projectile propulsion, one of the virtual hands will likely be positioned on the handle of the bow or slingshot, while the other sits at the furthest point of the drawn bowstring or elastic band. The distance between the controllers determines the draw length, which is proportional to draw weight, and therefore serves as the basis for the calculation of the initial speed of the projectile following release. In cases where the projectile is being propelled by a manual throw (e.g., throwing a ball or a hand-thrown weapon) or as a result of a collision with a hand-held object (e.g., tennis racquet, baseball bat), its initial velocity will be determined by the velocity of the controller. It is important to note that determining the velocity of such projectiles depends on their distance from the rotational anchor (e.g., the wrist of the player) at the moment of propulsion. With hand-thrown objects, it is advisable to make sure that the object that is about to get propelled is snapped to the controller's center of gravity, as opposed to an arbitrary point on the controller [177], which provides the right radius necessary for an intuitive throw. In case the projectile is being propelled by a collision with a hand-held tool, the tool's length at the point of impact will be included in the calculation of the overall tangential velocity with respect to controller velocity.

When it comes to mechanics involving tools as extensions of peripersonal space versus projectile-based interaction mechanics, certain types of tools cannot be easily categorized as belonging to one or the other. For example, some types of firearms may be equipped with a bayonet, allowing for different types of strategies depending on the distance between the player and the enemy. A tennis racquet serves a dual purpose — it is used to interact with an explicit target within the extended peripersonal space (i.e., to hit a tennis ball once it draws near), but also to launch a projectile — the ball — toward a distant implicit target. The goal of the presented taxonomy was not to divide interaction mechanics in separate categories — each mechanics will likely belong to several — but to aid in understanding various aspects of each mechanics.

4.3.5 Comprehensive taxonomy of VR interaction mechanics

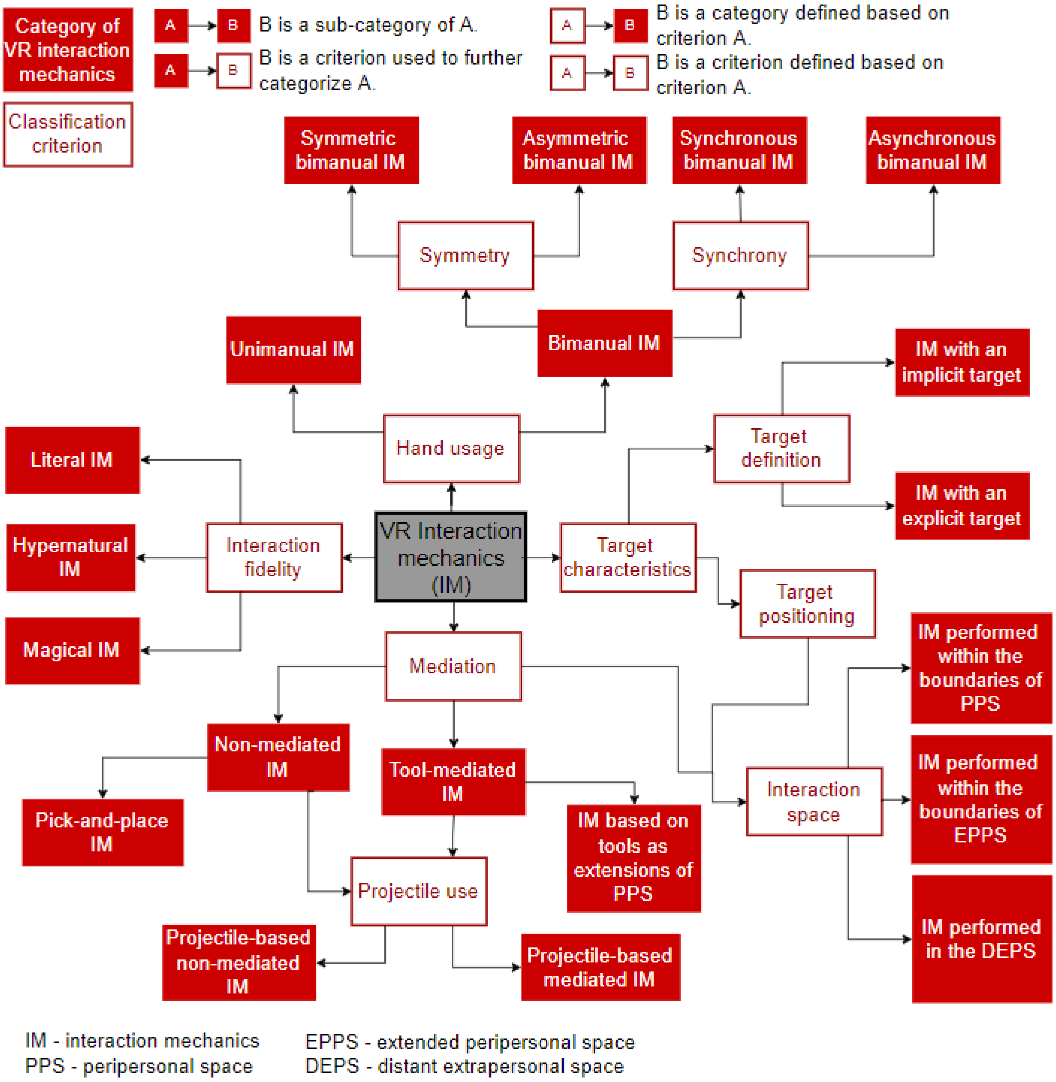


Figure 4.4: A taxonomy of interaction mechanics for VR gaming based on multiple criteria (taken from [10])

A diagram presenting a comprehensive overview of the interaction mechanics categorization discussed throughout this section is presented in Fig. 4.4, which highlights different categories of VR interaction mechanics, as well as criteria used for their classification. Following this taxonomy, the impact of playing VR games with popular primary IMs on VRISE is presented in the next chapter. Further dissection of each IM with the goal of defining the parameters that contribute to user experience is presented in Chapter 6, along with the results of three studies exploring their effects in practice.

4.4 Chapter summary

This chapter provides a discussion on the existing definitions of game mechanics and introduces the term *interaction mechanics* for referencing interaction techniques specific to VR gaming. Moreover, it combines existing criteria for the classification of interaction techniques (e.g., symmetry, synchrony, and interaction fidelity) with novel classification criteria centered around tool mediation, target definition, and interaction space. The result is a **proposed novel taxonomy of interaction mechanics for VR gaming**, which provides the means for further description and comparison of test materials, as well as for the interpretation of study results.

Chapter 5

Effects of popular VR game mechanics on VRISE and workload

5.1 Introduction

Uncomfortable sensations that arise during VR use have always been among the industry's biggest challenges. While certain VR-induced effects, such as cybersickness, have garnered a lot of interest from academia and industry over the years, others have been overlooked and underresearched. This chapter presents the results of two studies (Study 1 and Study 2) focused on exploring the prevalence and intensity of multiple features of workload and VRISE during VR gaming. Under the hypothesis that different game genres and mechanics may serve as relevant factors influencing the selected QoE features, commercial VR games that were chosen as test materials were selected to showcase different primary interaction mechanics commonly encountered in VR gaming. The initial results of Study 1 have been published as a conference paper [178]. Extending the initial conference paper, the text and figures in this chapter have been submitted as a journal article [30] (currently under review).

5.2 Background and related work

In previous work, Hirzle et al. [25] conducted an online user study with 352 participants who reported on symptoms of VR-induced discomfort after a 30 minute period of participating in a VR experience of their choice. The study was conducted on fairly experienced VR users, the majority of whom reported using VR technology on a weekly basis. Based on the results of the study, the authors' extracted factor model of VR-induced discomfort was divided into six factors — one for digital eye-strain, one for simulator sickness, and four relating to ergonomic symptoms, of which the most pronounced related to sweating and "perceived change" that occurs as the user is wearing the device (i.e., feeling physically different and perceiving differences in their movement, which becomes hindered as a result of wearing the HMD). Overall, digital eye-strain symptoms were less pronounced than ergonomic symptoms, with simulator sickness symptoms being the least pronounced. When asked to rate the perceived relevance of the symptom groups, participants rated simulator sickness symptoms as the least relevant, with digital eye-strain and ergonomic symptoms receiving similar ratings. Results of this study reaffirm the need for embracing more comprehensive methodological approaches for the evaluation of VR-induced discomfort.

5.2.1 Simulator sickness and cybersickness

Ranging from perceptual and cognitive effects to changes in affective state, physiological functions, postural instability, and ergonomic symptoms, there is a multitude of ways in which VRISE can manifest in a user, but the one that has arguably garnered the most attention is the state referred to as cybersickness. Usually characterized by symptoms such as disorientation, nausea, and oculomotor difficulties, cybersickness is most often thought to arise in response to a sensory conflict between the visual and the vestibular sense [179], although other theories are also discussed across literature [130, 180].

Over the years of researching cybersickness, different application-related factors were considered as potential contributors to its occurrence and intensity. As discussed in [181], cybersickness often increases with a pronounced sensation of vection — the so-called illusion of self-motion — arising as a result of optical flow as the user traverses through the virtual scene. Thus, the likelihood that a particular VR application will provoke cybersickness in its users depends on different aspects of the virtual environment, e.g., its level of visual detail, or the implementation of the used locomotion method.

Simulator sickness — the term used by [25] as it refers to symptoms measured by the Simulator Sickness Questionnaire (SSQ) [114] — is often used interchangibly with the term cybersickness, which we chose to use in this article. However, other researchers make sure to separate the two, as simulator sickness occurs during simulator use, and cybersickness comes as a result of experiencing virtual environments. Furthermore, according to [182], the two distinct types of sickness differ in terms of symptomology. Cybersickness was shown to cause symptoms that are higher in intensity compared to simulator sickness (as measured by the SSQ). Moreover, in case of cybersickness, scores for nausea (SSQ-N) and disorientation (SSQ-D) symptom groups were higher compared to oculomotor symptoms (SSQ-O), while the opposite was true for simulator sickness.

While the work presented in Stanney et al. [182] was conducted over 25 years ago, the topic of using simulator sickness measures, such as the SSQ, in the context of evaluating cybersickness is still of interest to researchers. Virtual Reality Sickness Questionnaire (VRSQ) [116] and Cybersickness Questionnaire (CSQ) [115] are fairly recent variations on the SSQ aimed specifically for measuring cybersickness rather than simulator sickness. Both questionnaires include only a subset of symptoms measured in the original questionnaire. Perhaps contrary to the aforementioned findings by Stanney et al. [182], symptoms that are excluded from both VRSQ and CSQ (sweating, increased salivation, stomach awareness, and burping) belong primarily to the nausea category of symptoms. Both VRSQ and CSQ include fullness of head, eye-strain, difficulty focusing and blurred vision, vertigo, and dizziness with eyes closed. We note the following differences between the two questionnaires:

- there are certain symptoms which are included in one of the SSQ variants, but missing from the other,
- general discomfort and fatigue are included only in the VRSQ, and
- nausea and dizziness with eyes open are only included in the CSQ.

The two questionnaires also differ with regard to the scoring system and highlighted dimensions of cybersickness, with VRSQ including oculomotor symptoms and disorientation as separate dimensions (note, we later refer to these as VRSQ-O and VRSQ-D respectively), and CSQ including difficulty in focusing and dizziness (we refer to these as CSQ-DF and CSQ-D respectively). Unlike SSQ and VRSQ, CSQ does not provide a way to calculate the overall (total) sicknesss score (referred to as SSQ-T and VRSQ-T).

Sevinc and Berkman [183] found that both VRSQ and CSQ outperformed the SSQ in terms of psychometric qualities for the evaluation of headset-based VR applications. Both VRSQ and CSQ were shown to provide a valid, reliable measure of cybersickness, although it remains unclear which of the two performs better. It is also worth noting that there were some limitations regarding the development and evaluation of both which calls for further research.

5.2.2 Ergonomics and device-related discomfort

Considering that VR experiences which utilize either hand tracking technology or handheld controllers tracked in six degrees of freedom (6DoF) usually rely on mid-air interactions, it is vitally important to consider ergonomics in application design. The so-called *gorilla arm syndrome* refers to muscle fatigue and perceived heaviness in the arms that occur following prolonged mid-air interaction [184]. Souchet et al. [26] list examples of user-related (age, body mass index), hardware-related (headset weight and fastening straps, used interaction devices, errors in position tracking, display resolution), and software-related factors (duration of VR use, required head and body rotation, required general posture, amplitude of gestures, task repetition, body parts representation and feedback) that contribute to VR-induced muscular fatigue.

Certain movements and postures are more likely to cause significant muscle fatigue. For example, having to interact with the system with the arm fully extended at shoulder height is more fatiguing than lowering the arm with a bent elbow [185, 186]. Previous work has shown that placing targets in VR 15 or more degrees above eye level or 30 or more degrees below eye level leads to greater discomfort in neck and shoulder muscles [187]. Thus, VR application designers can reduce the user's physical workload by mindful placement of target objects, avoiding positions that would impede prolonged use.

While the aforementioned findings can serve as a useful rule of thumb for applications that focus on efficiency rather than entertainment, there is a gap in research regarding the design of *interaction mechanics* for VR games [10], as decreasing physical workload in VR gaming may actually backfire. For example, Yoo et al. [23] have shown that players prefer VR games that require light physical exertion as opposed to those that require no exertion. Similarly, Evans et al. [188] suggest that players prefer games that require light physical exertion with mechanics that predominantly involve arm movements. Conversely, there is such a thing as too much exertion, as players wish to avoid sweating and excessive physical demands [23]. While temporary physical exhaustion and discomfort detract from the user experience, also possibly affecting player retention, this topic has not yet garnered a lot of interest (or raised serious concern) among the research community, at least compared with efforts invested toward investigating and mitigating application design factors that aggravate cybersickness symptoms.

Unfortunately, considering that wearing a VR headset was shown to impair user posture, stressing the musculoskeletal system [189], wearing the headset while performing energetic, repetitive movements may lead to serious consequences, even if the game of choice does not seem too demanding with regard to cardiovascular load. Baur et al. [190] describe a recent case of a healthy 31-year-old who fractured a vertebra as a result of playing *"a VR video game involving combinations of shoulder, arm and head movements to rhythmic visual and musical triggers"* [190, p. 2]. Stressing that energetic movements performed during VR gaming should not be underrated as a potential cause of injury, the authors predict further occurrences of such traumatic injuries as VR technology continues to grow in popularity.

While manufacturers are attempting to fix ergonomic issues of VR hardware, eliminating cables and reducing the size and weight of HMDs, other conditions of HMD use may be a strong determining factor in terms of whether or not a particular HMD is considered safe and comfortable. In addition to the time spent using the HMD, both speed and range of movement, as well as target placement, may impact the physical workload imposed on the musculoskeletal system, increasing energy expenditure and potential for injury. For example, Chihara and Seo [39] examined the effects of HMD mass and center of mass position on physical workload in different body postures. They stress the necessity of considering physical workload in the neck when designing HMDs, further noting that different neck postures (dependent on target placement) would benefit from different center of mass positions. The authors called for further research into HMD ergonomics using different test conditions.

5.2.3 Cognitive impact

Even with display improvements in refresh rate and resolution, visual design choices made during software development may produce oculomotor symptoms, while strobing lights used in special effects could be a trigger for migraine or seizures. But potential negative outcomes of VR use involve more than physical discomfort and fatigue, as researchers raise the issue of the effects of VR on cognitive performance (to be discussed in the remaining part of this section). Furthermore, it is important to consider that distinct VR aftereffects are not completely separate from each other, as they often share a common cause, or show a significant level of correlation.

In their article on the effects of VR-induced cybersickness on heart rate, cutaneous vascular tone, and reaction time (RT), Nalivaiko et al. [191] found that simple reaction time (SRT) increased by 20 to 50 ms following exposure to VR content (rollercoaster simulators). This increase in SRT was primarily attributed to nausea; however, the authors also noted that the intensity of such effects was greatly influenced by the specific characteristics of chosen VR content, such as visual motion cues. In light of this discovery, the authors expressed their concern about the safety of VR technology. The use of rollercoaster simulators in the exploration of the impact of VR use on reaction time and cybersickness can also be found in the work of Nesbitt et al. [192]. Similar to [191], their results indicated that immersion in VR may produce a significant increase in SRT (M = 26.25 ms, SD = 39.92), which correlates with the increase in cybersickness-related symptoms.

The effects of VR usage on different measures of cognitive performance were investigated by Mittelstaedt et al. [193], along with their relationship to cybersickness. While VR exposure did not impair performance in certain cognitive tasks — e.g., Mental Rotation Task and Corsi Block Task — the authors noted a significant change in reaction time, with SRT increasing by 17-29 ms after being exposed to VR. The authors offered a number of alternative explanations for this effect, rejecting the notion that cybersickness is the only cause behind the increase in SRT (even though a level of correlation between increased SRT and cybersickness was found). Possible explanations included visuomotor adaptation (based on the work by [194]) and cognitive adaptation to the slight latencies stemming from the use of I/O devices.

Compared to [191, 192] the VR content used by Mittelstaedt et al. [193] granted a higher level of control to the user, but all of the above examples display visual motion cues that are likely to produce a strong sensation of vection. In contrast, *VR table tennis* used in [195] and *Beat Saber* used in [196] are both based on mechanics that are strongly reliant on physical movement (namely gross motor movement of the arms), but without artificial self-motion. Using the CANTAB five-choice RT task instead of the commonly used SRT tasks, the authors of [195, 196] were able to distinguish between decision time and motor movement time. No significant differences between post-VR and pre-VR decision or motor movement times were found in [196]. VR use did not produce a significant change in overall RT in [195] — although there was a general increase in decision time, but not movement time — and there was no correlation between RT and cybersickness.

5.3 Methodology

5.3.1 Materials

Figure 5.1: Snapshots of the chosen games for *Study 1* (a-c) and *Study 2* (d-f): a) Beat Saber; b) Order Up VR; c) Serious Sam VR: The Last Hope, d) Fruit Ninja VR, e) Dungeon Brewmaster, f) Space Pirate Trainer (taken from [30])

Although VR experiences that produce significant levels of optical flow are a common choice for studies addressing VRISE, for the purpose of this thesis a choice was made to narrow the focus on commercial VR games with a first-person perspective, as well as a standing or Room-Scale gameplay (i.e., no in-game locomotion methods). Because of this absence of vection-producing elements and according to the industry comfort ratings available at the time of this writing [197, 198], all chosen games fall into the most comfortable *green* category of VR games.

As explained in Chapter 1, along with being simple in terms of gameplay, these games were chosen as representations of popular VR genres and showcased common types of VR IMs. In both studies, participants were exposed to an example of pick-and-place mechanics, as well as

Table 5.1: Distinguishing characteristics of chosen games (IM - interaction mechanics; PA - play angle); (taken from [30])

	Game	Genre	Core IM	Considerations regarding PA	Aesthetics, visual effects, and motion cues
Study 1	BS	Rhythm, action	Slash	Targets in front of player; game occasionally requires players to lower the body stance to avoid an obstacle	Nocturnal setting, sleek minimalist aesthetics, dark background with high-contrast emissive elements, no high-frequency textures, no characters, cubes and walls slide toward the player
	OU	Cooking simulation	Pick-and-place	Game occasionally requires full body rotation; game occasionally requires players to lower the body stance to acquire an object	Diurnal interior setting; bright, colorful low-poly aesthetics; no emissive materials, no high-frequency textures; cartoonish characters slide toward the player
	SS	Action, shooter	Shoot	Targets in front and on both sides of player; targets placed above eye-level — players occasionally need to look up	Diurnal semi-realistic exterior setting; dynamic special effects, semi-realistic monster characters with high-frequency textures and complex geometry move and fly toward the player, who is targeted by various projectiles
Study 2	FN	Action	Slash	Targets in front and slightly to both sides of the player	Switches between diurnal and light nocturnal exterior setting; colorful cartoonish aesthetics; dynamic special effects; simple objects shoot up in front of the player and subsequently fall down
	DB	Cooking simulation	Pick-and-place	Target objects in front and on both sides of player; game occasionally requires full body rotation; game occasionally requires players to lower the body stance to acquire an object	Nocturnal interior setting; low-poly cartoonish aesthetics; emissive materials and particle effects, no high-frequency textures; cartoonish monster characters with simple animations approach the player
	SPT	Action, shooter	Shoot	Targets in front and on both sides of player; targets placed above eye-level — players consistently need to look up; game occasionally requires players to lower the body stance to avoid damage	Nocturnal exterior setting; sleek minimalist aesthetics; dark background with high-contrast emissive elements and dynamic special effects; no high-frequency textures; minimalistic drones fly around and shoot projectiles at player; enemy attacks occasionally trigger a slow-motion anaglyph-like effect

two different types of tool-mediated mechanics — a swordplay-based one with targets closer to the player, and a projectile-based one with mostly distant targets.

When choosing games for *Study 2*, the focus was placed on games that were similar in genre and/or core interaction mechanics to the games chosen for *Study 1*. However, the focus was also on choosing games that differed in terms of aesthetics, considering whether each game had a colorful, diurnal setting, or a darker, nocturnal setting contrasted with emmissive materials and/or gleaming particle effects. Screenshots of the games chosen for *Study 1* (*Beat Saber* – BS, *Order Up VR* – OU, *Serious Sam VR: The Last Hope* – SS) and *Study 2* (*Fruit Ninja VR* – FN, *Dungeon Brewmaster* – DB, *Space Pirate Trainer* – SPT) can be seen in Figure 5.1, while Table 5.1 presents their characteristics.

Ergonomic effects of VR use are obviously highly dependent on headset design, although the impact of hardware design factors is beyond the scope of this thesis. However, the intensity of those effects can be significantly compounded by movements and postures required by the chosen application. Thus, focusing on the impact of such software-related factors, an identical hardware setup was used across all gaming sessions: the HTC VIVE Pro Eye with the accompanying VIVE (2018) controllers.

5.3.2 Procedure

To prevent symptoms from building up over time, each participant was asked to partake in three separate gaming sessions (one per game, in randomized order), each session being scheduled for a different day. The entire process is presented in Figure 5.2. All sessions were held in a laboratory and supervised by an administrator. Participants completed a pre-study questionnaire at the start of the first session, providing the required personal information (age, sex, level of education, etc.), as well as a self-evaluation of their level of experience with VR and gaming in general. Following this step, participants were given instructions on the use of the HTC Vive Pro system, from adjusting the headset straps and the inter-pupillary distance, to handling the Vive controllers.

Prior to each of the three VR sessions, necessary baseline measurements were collected. This included participants filling out the SSQ [114], as well as evaluating the current level of

Table 5.2: The Borg CR-10 rating scale [199], as used in this study — taken from [178]

Borg CR-10 rating scale	
0 - Nothing at all	
1 - Very weak	6
2 - Weak	7 - Very strong
3 - Moderate	8
4 - Somewhat Strong	9
5 - Strong	10 - Extremely strong

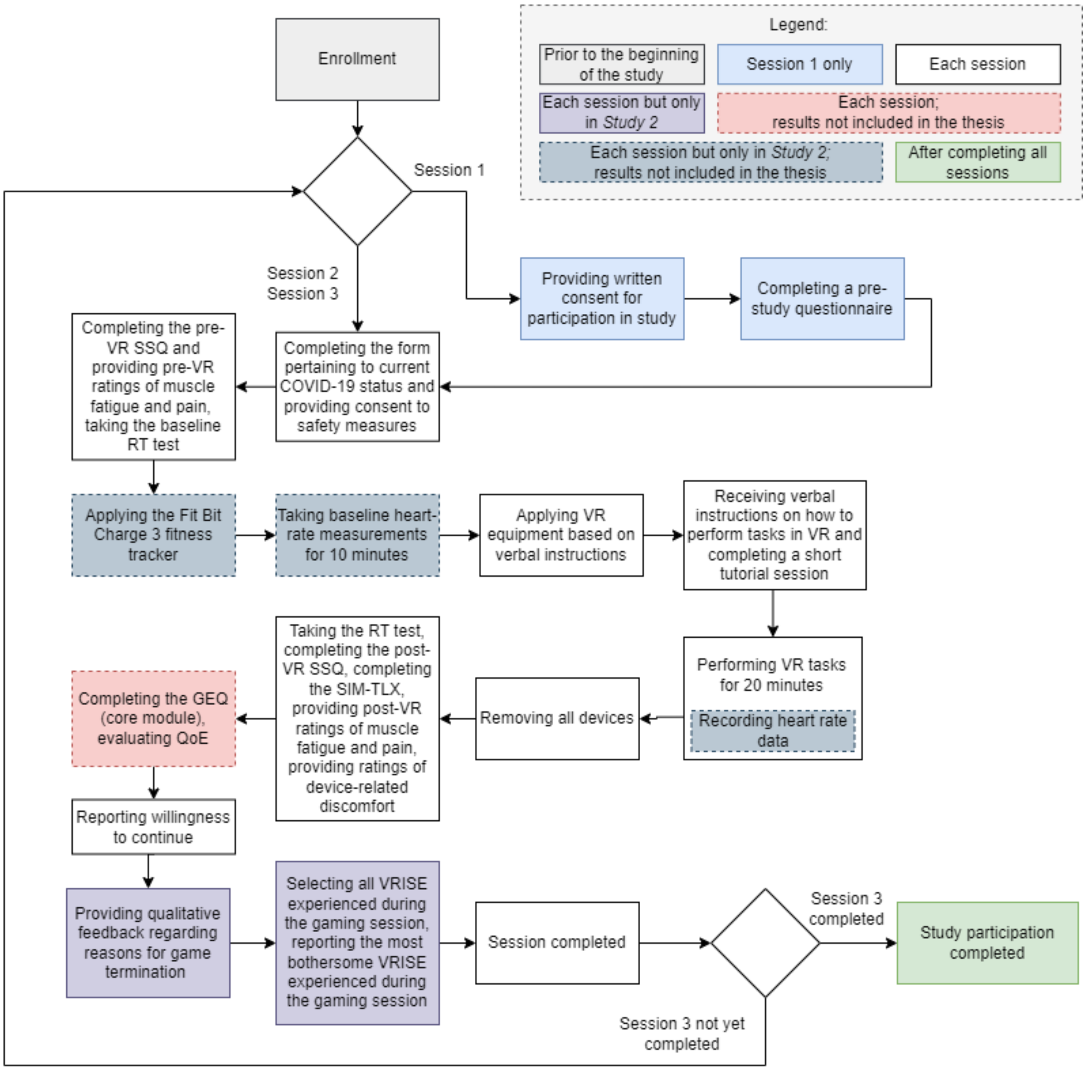


Figure 5.2: Flow diagram depicting the complete study procedure for each participant — the same core methodology was followed for both studies, but several additional steps were used only in *Study 2* (adapted from [30])

pain and muscular fatigue in the arms (i.e., shoulders to fingertips) and different regions of the back, as depicted on a simplified anatomical diagram given to each participant as a reference. While corresponding to the areas surrounding cervical, thoracic and lumbar spine, these regions were colloquially described as neck, upper back, and lower back with the aim of facilitating user comprehension.

The Borg CR-10 scale [199], which was used to report the sensations of pain and muscular fatigue, is presented in Table 5.2. Baseline values of objective measures were also collected, as both studies included the use of the Deary-Liewald RT task (DLRT) [200], administered using a desktop computer. This tool was used to collect pre-VR measurements of simple (SRT) and four-choice (CRT) visual reaction times. Additionally, for *Study 2*, heart rate data was also collected using the Fitbit Charge 3 fitness tracker device. To obtain the baseline measurement, participants were asked to sit still for 10 minutes. Given the large number of collected measures, this thesis will not contain all collected results (as indicated in Fig. 5.2), but only those results considered to be most relevant to the thesis focus.

Following the process of obtaining baseline measurements, participants were warned about the possibility of cybersickness and other uncomfortable symptoms that may appear as a consequence of partaking in a VR session, and were encouraged to pause or completely terminate the experiment in case they experience significant discomfort. They were then familiarized with the rules and control mappings of the particular game they were about to play and entered VR to take part in a short tutorial session. After grasping the mechanics of the game, participants started the 20-minute test session. Participants' heart rate was being recorded throughout the entire test session, while the remaining measures were collected as soon as the session ended. After exiting the session, participants were first instructed to retake both SRT and CRT tests. Afterward, they were asked to fill out the SSQ and report their post-VR evaluations of pain and muscular fatigue in the arms and back using the Borg CR-10 scale, as previously described. Again using the Borg CR-10 scale, participants were asked to report the level of discomfort attributed to different aspects of the headset device, with separate items pertaining to discomfort resulting from headset fit, weight, temperature, display quality, and annoyance with the headset cable.

The SIM-TLX questionnaire [160] was used to obtain evaluations of different dimensions of workload required to perform the given task. The SIM-TLX was based on the NASA-TLX [159] but includes items that are relevant for the specific use of simulated environments (such as VR). Although the full version of this questionnaire provides a way to obtain the overall workload score using pair-wise comparisons between different dimensions of workload, this weighing process was ommited to focus on individual dimensions of workoad, as is often done for NASA-TLX [201]. Considering multiple questionnaires were used over the course of both studies, this SIM-TLX modification reduced the overall number of evaluations the participants would have to perform.

Participants in both studies were asked to complete the core module of the Game Experience Questionnaire (GEQ) [19, 20] and additionally provide their rating of the overall Quality of Experience (QoE) on a 5-point Absolute Category Rating scale. They were also asked to report whether they would be inclined to terminate or continue the experience. For *Study 2* those inclined to terminate were asked to disclose a qualitative explanation for their choice. Finally, in *Study 2* participants were provided with a list of different VRISE (presented in the Results section) and asked to check the ones they experienced during the session, as well as report the single effect they found most bothersome.

It is important to note that both studies were conducted in the midst of the COVID-19 pandemic. To assure the safety of participants during this process, rigorous hygiene and safety practices were enforced. Prior to each session, participants were asked to provide written consent in which they expressed compliance with the enforced safety measures. Both participant and administrator were required to wear masks and maintain appropriate interpersonal distance at all times. To reduce physical discomfort, participants were allowed to take their mask off during the gaming session which was performed in a spacious laboratory set up in a way that enables safety during physical activity required by the games. During this time, the administrator would exit the room and continue to supervise the player through see-through glass walls of the laboratory. All experiments were carried out in a well-ventilated space, all surfaces and devices were thoroughly cleaned and disinfected between sessions, and participants were provided with single-use headset covers.

5.3.3 Participants

Taking into account the overall diversity of the broad population, because of the limited scale of both studies, the homogeneous sampling approach was selected as it would provide clearer generalizability [202]. Deliberately choosing participants who were alike in terms of age, health, and disability status across both studies, facilitated the comparison between their results. The participants in both studies were young adults, healthy, and without disabilities, which means they belong to a demographic which is expected to be the least vulnerable to the negative effects of VR use. It is likely that any negative impacts of VR gaming would be even more pronounced in virtually any other demographics. Therefore, because the participants of both studies do not have to deal with obstacles that many others are facing when it comes to VR device design and motor requirements of VR gaming, the extent of VR gaming-related health and safety issues as described in this chapter can arguably be considered *the "best" case scenario*. The characteristics of tested participant samples across both studies are presented in Table 5.3.

5.4 Results

5.4.1 Workload

Findings

The SIM-TLX questionnaire was used to assess the levels of Mental demand (MENT), Physical demand (PHYS), Temporal demand (TEMP), Frustration (FRUS), Task complexity (CMPX), Situational stress (STRS), Distractions (DIST), Perceptual load (PERC), Task control (CONT), and Total workload (SUM) after each game. Assessments from all participants were collected in both studies. However, due to data corruption issues we were only able to analyze workload data reported by 12 out of 20 participants in *Study 2*. Note, no corruption occurred for *Study 1* (i.e., data from all 20 participants was included in the calculations) and all other data (aside

Table 5.3: Characteristics of participant samples across both studies

		Study 1	Study 2
Number of participants		20	20
Age*	Average	24.45 (SD = 2.58)	23.85 (SD = 2.57)
	Min	20	21
	Max	29	29
Sex	Female	10	10
	Male	10	10
VR experience	Inexperienced	5	7
	Beginner	10	9
	Intermediate	4	2
	Expert	1	2
Gaming experience	Inexperienced	0	0
	Beginner	4	7
	Intermediate	8	10
	Expert	8	3
Proneness to motion sickness	None	7	3
	Mild	9	13
	Medium	2	2
	Strong	2	2
Proneness to cybersickness	None	11	9
	Mild	8	5
	Medium	0	4
	Strong	1	4

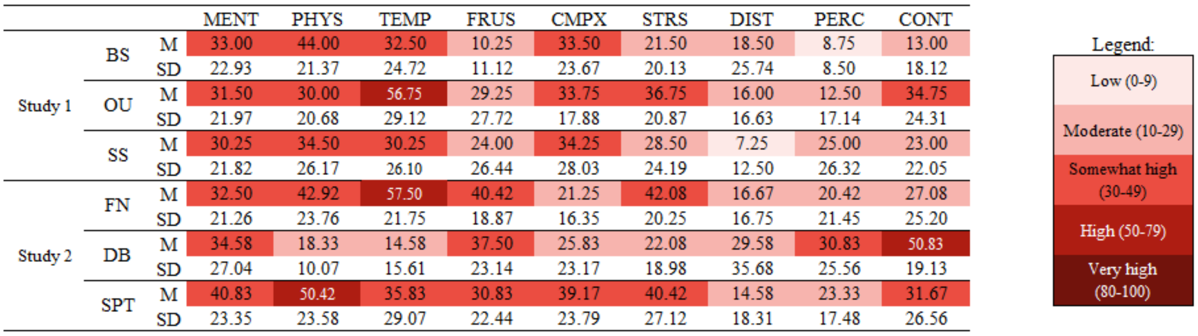
Age reported in years, all other items reported as the number of participants who fit each criteria (taken from [30])

from corrupted SIM-TLX scores) was collected and calculated for all 20 participants in *Study 2*.

As explained in the previous section, to simplify the scoring process, the weighting method explained in [160] was ommited. Instead, individual dimension ratings reported on the 21-point Likert scale were transformed to a 0-100 range with 5-point steps, as is often done with NASA-TLX [201]. To calculate the SUM score the number of points on the Likert scale (i.e., points in the 0-21 range as opposed to the 0-100 range) was aggregated for all individual dimensions of workload, resulting in the overall score of up to 189 points.

The mean SUM score of games in *Study 2* — FN (M = 68.17, SD = 23.06), DB (M = 61.83, SD = 23.27), and SPT (M = 70.42, SD = 31.30) — was higher compared to the SUM score of games in *Study 1* — BS (M = 52.00, SD = 21.57), OU (M = 65.25, SD = 23.32), and SS (M = 56.40, 25.24). In terms of the overall SUM score, Friedman test did not show a significant difference between games for either study, as notable differences were found only for individual dimensions of workload.

Mean SIM-TLX scores for both studies are presented in Figure 5.3. In line with the NASA-TLX interpretation scale reported by [203], we categorized the mean scores as either *Low* (0-9), *Medium* (10-29), *Somewhat high* (30-49), *High* (50-79), or *Very High* (80-100).



			MENT	PHYS	TEMP	FRUS	CMPX	STRS	DIST	PERC	CONT
Study 1	BS	M	33.00	44.00	32.50	10.25	33.50	21.50	18.50	8.75	13.00
		SD	22.93	21.37	24.72	11.12	23.67	20.13	25.74	8.50	18.12
	OU	M	31.50	30.00	56.75	29.25	33.75	36.75	16.00	12.50	34.75
		SD	21.97	20.68	29.12	27.72	17.88	20.87	16.63	17.14	24.31
	SS	M	30.25	34.50	30.25	24.00	34.25	28.50	7.25	25.00	23.00
		SD	21.82	26.17	26.10	26.44	28.03	24.19	12.50	26.32	22.05
Study 2	FN	M	32.50	42.92	57.50	40.42	21.25	42.08	16.67	20.42	27.08
		SD	21.26	23.76	21.75	18.87	16.35	20.25	16.75	21.45	25.20
	DB	M	34.58	18.33	14.58	37.50	25.83	22.08	29.58	30.83	50.83
		SD	27.04	10.07	15.61	23.14	23.17	18.98	35.68	25.56	19.13
	SPT	M	40.83	50.42	35.83	30.83	39.17	40.42	14.58	23.33	31.67
		SD	23.35	23.58	29.07	22.44	23.79	27.12	18.31	17.48	26.56

Legend:
Low (0-9)
Moderate (10-29)
Somewhat high (30-49)
High (50-79)
Very high (80-100)

Figure 5.3: Heat map depicting the mean SIM-TLX scores (and their standard deviations) obtained for individual dimensions of workload (different background shades represent mean score categorization; taken from [30])

For Study 1, Friedman tests followed by a post-hoc analysis with Wilcoxon signed-rank tests and Bonferroni correction found statistically significant differences between games for all dimensions of workload except for MENT, CMPX, STRS, and PERC. For PHYS ($\chi^2(2)$ = 10.50, p = 0.005), there was a significant difference between BS and OU (Z = -2.50, p = 0.01). For TEMP ($\chi^2(2)$ = 12.67, p = 0.002), the significant differences were found between BS and OU (Z = -2.72, p = 0.006), as well as OU and SS (Z = -2.41, p = 0.02). For FRUS ($\chi^2(2)$ = 10.64, p = 0.005), OU and BS were significantly different (Z = -3.15, p = 0.002). For DIST ($\chi^2(2)$ = 8.19, p = 0.02), differences were found between OU and SS (Z = -2.45, p = 0.01). Lastly, for CONT, significant differences ($\chi^2(2)$ = 12.97, p = 0.002) were found between BS and OU (Z = -3.16, p = 0.002), and BS and SS (Z = -2.40, p = 0.02).

For Study 1, Friedman tests followed by a post-hoc analysis with Wilcoxon signed-rank tests and Bonferroni correction found statistically significant differences between games for PHYS, TEMP, CMPX, and STRS. For PHYS ($\chi^2(2)$ = 14.44, $p < 0.001$), there was a significant difference between FN and DB (Z = -2.55, p = 0.01), as well as between DB and SPT (Z = -2.98, p = 0.003). For TEMP ($\chi^2(2)$ = 13.32, p = 0.001), a significant difference was found between FN and DB (Z = -2.94, p = 0.003). For CMPX ($\chi^2(2)$ = 8.97, p = 0.01), FN and SPT were significantly different (Z = -2.68, p = 0.007). For STRS, a significant difference ($\chi^2(2)$ = 9.52, p =0.009) was found between FN and DB (Z = -2.80, p = 0.005).

Discussion

The games in *Study 2* received higher SUM scores on average compared to games in *Study 1*. It is worth noting that games that shared similar genres/mechanics did not necessarily receive similar mean scores of total workload. Statistically significant differences in SUM scores were not found between games in either study.

When observing the results for separate dimensions of workload, the mean ratings did not exceed a *High* level of workload for either game. Overall, the majority of mean ratings could be described as either *Somewhat high* or *Moderate* with regard to intensity of workload.

All games except for BS and SS (games that did not receive any *High* ratings) obtained a *High* score in a single dimension: PHYS for SPT, TEMP for OU and FN, and CONT for DB. While both OU and FN were rated as *High* in TEMP, based on their gameplay we can conclude that they received those scores for different reasons — whereas FN required a series of quick, precise, discrete movements, OU was demanding a number of more complex actions (i.e., cooking and assembling food) to be performed in a relatively short period of time. MENT was the only dimension for which all games received a similar mean score, suggesting that manual, isomorphic, controller-based games tend to produce a *Somewhat high* level of mental workload regardless of their core mechanics. While both swordplay-based games achieved a very similar PHYS score, noted differences between OU (*High*) and DB (*Moderate*) scores, as well as SS (*Somewhat high*) and SPT (*High*), indicate that games with similar core mechanics should not necessarily be expected to require similar levels of physical exertion, as the imposed workload likely differs based on target placement, temporal constraints, or secondary mechanics. Interestingly, all games in *Study 1* were deemed similar in terms of CMPX, while SPT was rated as being more complex compared to the other games in *Study 2*, likely because of the frequency of offensive and defensive actions required from the player and the accompanying level of physical activity. The aforementioned reasons are likely also responsible for SPT receiving a highest CONT score of all games. However, it is also worth noting that OU and DB received a *Somewhat High* mean CONT score despite their relatively restrained physical requirements (especially in case of DB).

While games with pick-and-place mechanics may be less demanding in terms of gross motor movements — compared to more action-packed games — they are more demanding in terms of fine motor movements, specifically wrist movements necessary for orienting and aligning virtual objects. Furthermore, these mechanics are non-mediated [10]. With tool-mediated mechanics VR controllers serve as a physical, tangible representation of virtual tools. With both sword-like weapons and guns, both orientation and girth of handheld controllers generally match the orientation and girth of virtual tool handles to a sufficiently realistic degree, meaning that player's physical hands correspond very well with their virtual hands. The positioning and behaviour of virtual hands in case of non-mediated mechanics is not as straightforward, requiring substantial adaptation to different angles on the player's part, which is further complicated by the properties of virtual objects that need to be grabbed, repositioned, and interacted with.

Table 5.4: Wilcoxon Z-score comparing post-VR and pre-VR pain and muscle fatigue scores; * $p < 0.05$, ** $p < 0.005$, *** $p < 0.001$ (taken from [30])

	Game	Arm fatigue	Arm pain	Neck fatigue	Upper back fatigue	Upper back pain	Lower back fatigue	Lower back pain
Study 1	**BS**	**-3.45*** **	**-2.55***	-1.39	-1.35	**-2.16***	-1.14	-0.88
	OU	-0.77	-0.32	**-2.13***	-0.28	-0.30	-0.58	-1.41
	SS	**-2.30***	**-2.39***	-1.20	-1.54	-0.94	-1.64	-0.37
Study 2	**FN**	**-3.43*** **	**-2.78***	-1.31	-2.27	-1.93	**-2.07***	0.04
	DB	**-2.80***	**2.97****	-1.06	-0.79	-0.85	-0.64	-0.17
	SPT	**-2.88****	**-2.96****	-1.03	**-2.46***	-1.19	**-2.36***	**-2.00***

5.4.2 Pain and muscle fatigue

Findings

We performed Wilcoxon signed-rank tests to see whether post-VR results significantly differ from the baseline measurements. Neck pain was the only symptom that did not increase significantly for either game. All other results are shown in Table 5.4, with arm fatigue and arm pain scores reaching statistically significant increases after every game except for OU. While there were significant differences comparing pre-VR and post-VR scores of each symptom, when comparing the post-pre differences in pain and muscle fatigue between games, Friedman tests followed by Wilcoxon signed-rank tests with Bonferroni correction failed to identify significant differences for either study.

On average, playing each of the VR games resulted in an increase in pain and muscle fatigue. However, the mean post-pre difference was less than one point on the Borg-CR10 scale for all body parts except for arms. Overall, the mean increase in arm fatigue was 1.5 (SD = 1.40) for BS, 0.25 (SD = 1.48) for OU, and 1.2 (SD = 2.16) for SS for *Study 1*. For *Study 2* it was 1.95 (SD = 1.56) for FN, 0.85 (SD = 1.15) for DB, and 1.75 (SD = 2.19) for SPT. The mean increase in arm pain was 0.95 (SD = 1.40) for BS, 0.05 (SD = 0.87) for OU, and 0.95 (SD = 1.60) for SS for *Study 1*. For *Study 2* it was 1.25 (SD = 1.55) for FN, 0.65 (SD = 0.79) for DB, and 1.15 (SD = 1.35) for SPT.

It is important to note, however, that participants' experiences with fatigue and pain were quite varied and inconsistent. A significant percentage of participants did not report any changes in the intensity of symptoms, and a minority even experienced improvements — for Study 1, the number of participants (out of 20) who indicated a decrease in symptom intensity ranged from 0 to 6, depending on the symptom/game. For Study 2, improvements were noted in up to 3 participants (out of 20) per symptom/game. However, when observing only those participants who did experience an increase in symptoms (more detail on the frequency of occurrence of particular symptoms in Subsection 5.4.5), the reported intensity of symptoms increased by 2 or more points on the Borg scale (on average) for all games except for DB and SPT. While the majority of participants only experienced increases of 1 to 2 points, for each game except for OU, a small number of individuals reported worrying levels of musculoskeletal discomfort and fatigue. For example, following the FN gaming session six different participants (30%) reported increases of 4-5 points on the Borg scale for at least one of the symptoms in this symptom group. Likewise, five participants (25%) experienced increases of at least 4 and up to 7 points following the SPT scenario.

Discussion

Arm fatigue and pain were the only significantly increased symptoms for SS and DB. The only symptom that significantly increased for OU was neck fatigue, likely because participants had to turn their head more considering the 360-degree horizontal play angle. Although both swordplay-based games are very similar in terms of mechanics, which caused significant increases in arm pain and fatigue for both, BS resulted in significantly increased upper back pain, while FN caused significantly increased lower back fatigue. A possible explanation for this result lies in differences in target behavior and positioning. In FN, the player is expected to hit multiple targets in a single swing of the blade, with target objects spawning sporadically along the circular arc on the floor before the player. Expelled target objects shoot up vertically and fall under the effect of gravity, each of them moving along its individual path which means that different objects often happen to be placed at different heights at the same time. The comparatively wider horizontal play angle and diagonal slashes necessary to hit multiple targets at once may require rotations that start from the waist and therefore affect the muscles of the lower back, as opposed to BS, in which targets move in a more predictable manner and are more constrained in terms of horizontal and vertical positioning. With BS, however, the rhythmic nature of the game calls for a higher frequency of upper body movement, possibly affecting upper back muscle pain. The game that seemed to trigger the most diverse range of ergonomic symptoms was SPT, which produced statistically significant increases in arm fatigue, arm pain, upper back fatigue, lower back fatigue, and lower back pain. To avoid enemy attacks, users had to frequently bend over or crouch down, stressing their lower back. Continuously keeping their arms raised at an awkward angle whilst shooting at enemies positioned above eye level is the likely cause behind increased upper back fatigue.

The minority of participants who reported some level of slight musculoskeletal discomfort even prior to VR use, but later noted a decrease in the intensity of pain and muscle fatigue, may have experienced analgesic effects that were shown to occur with VR use [204]. This effect may have been compounded by the fact that the games we used (especially swordplay-based games and SPT) can be considered active VR games due to their physical demands, which likely enhances any analgesic effects of VR alone [205]. Nevertheless, some participants still experienced significant increases in pain and fatigue in different muscle groups. As studies have shown, light exertion in VR games is preferable to passive gameplay [23] and there are many benefits to VR design that promotes physical activity. However, even though significant worsening of musculoskeletal discomfort was reported by only several participants per each game, it is important to keep in mind the small scale of our study. If these results are in any way representative of a wider range of existing or potential consumers — if only 20 minutes of gameplay (played at a non-challenging level) can therefore produce musculoskeletal pain/fatigue increases of 4-5 points on the Borg CR-10 scale (e.g., causing discomfort to increase from non-existent to somewhat strong or strong) in 30% of young, healthy FN players, or increases of 4 up to 7 points in 25% SPT players (e.g., causing discomfort to increase from non-existent to somewhat strong or very strong) — then **this is not only a health-and-safety issue that needs to be communicated to potential consumers, but also a serious problem for the VR gaming industry**, implicating potential losses in terms of both player-base and revenue. Even increases as low as 1-2 points, especially when talking about pain rather than fatigue, indicate the **need for further improvements in ergonomic design of VR games and reconsideration of existing comfort-rating systems**, as further discussed in [206].

5.4.3 Device-related discomfort

Findings

Figure 5.4 presents the mean Borg CR-10 scores for discomfort caused by weight of the HMD, its temperature, fit (i.e., HMD feeling too tight or too loose) and display quality, as well as annoyance with the HMD cable, considering we were using a tethered headset.

Participants struggled with adjustable straps of the HMD, with HMD tightness appearing to be a more bothersome issue compared to HMD looseness. However the scores for fit-related issues remained below the *moderate* category for both studies and there were no significant differences in fit-related scores between the games.

For both studies, the shooter game scenario produced the highest level of weight-related discomfort, although Friedman test only showed significant differences between games for *Study 2* ($\chi^2(2) = 6.82$, p = 0.03), as post hoc analysis with Wilcoxon signed-rank tests and Bonferroni correction showed a statistically significant difference between FN and SPT (Z = -2.69, p = 0.007).

A significant difference between games ($\chi^2(2) = 10.05$, p = 0.006), more specifically between SS and BS (Z = -2.75, p = 0.006), was found for HMD temperature in *Study 1.* For the same study, Friedman test also showed a statistically significant difference in discomfort attributed to display quality ($\chi^2(2) = 8.60$, p = 0.01), as post hoc analysis confirmed a significant difference between OU and BS (Z = -2.43, p = 0.02), as well as SS and BS (Z = -2.96, p =

0.003).

Annoyance with the HMD cable was the only device-related VRISE that was shown to significantly differ between games for both studies. For *Study 1*, the Friedman test ($\chi^2(2) = 21.03$, $p < 0.001$) followed by post hoc analysis showed significant differences between OU and BS (Z = -3.13, p = 0.002), as well as SS and OU (Z = -3.56, $p < 0.001$). For *Study 2*, the Friedman test ($\chi^2(2) = 10.44$, $p = 0.005$) followed by post hoc analysis confirmed statistically significant differences between DB and FN (Z = -2.48, p = 0.01), as well as between SPT and FN (Z = -3.14, p = 0.002).

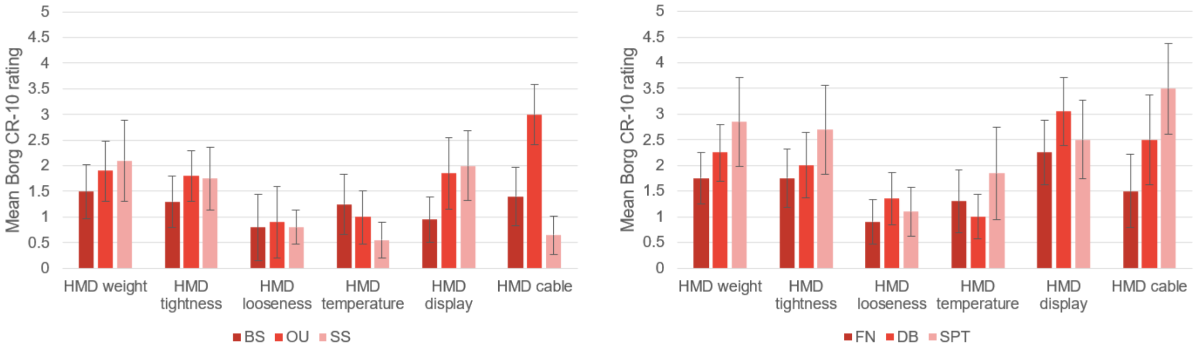


Figure 5.4: Mean Borg CR-10 ratings (95% CI) for discomfort and annoyance related to different device factors (*Study 1* – left, *Study 2* – right; taken from [30])

Discussion

Even though identical hardware was used for the entirety of both studies, the intensity of device-related symptoms varied between games, with differences in experiencing particular symptoms even reaching statistical significance. This is to be expected as the overall experience of device-related VRISE occurs as a result of combining specific hardware features with specific motor requirements of the particular application.

Annoyance with the cable was the only device-related VRISE that significantly differed between games in both studies. Playing a game with a 360-degree play angle (as is the case with OU) may result in players getting tangled in the long cable — an effect that does not occur with purely front-facing games. Similarly, the footwork-intense secondary mechanics of avoiding incoming projectiles in SPT occasionally resulted in users tripping and stepping on the cable. The safety implications of these findings are relevant for game developers, hardware manufacturers and user experience researchers alike, however, the rapidly growing popularity of portable HMDs confirms that cable-related hazards can be expected to become less of an issue for future generations of VR devices.

While this was only statistically significant for *Study 2*, HMD weight was most bothersome in the shooter gaming scenarios of both games. This is likely due to target positioning and behaviour — targets in shooter games are more dynamic, characterized by appearing and disappearing, moving in unpredictable ways, and attacking the player. As a result of their dynamics, the player is constantly scanning the environment, which often includes not only ocular movements, but head/neck movements as well. Furthermore, in both SS and SPT — but especially SPT — targets are placed above the player's eye-level, requiring the player to adjust the pitch orientation of their head. Thankfully, this did not seem to produce significant increases in reported neck pain, but it may provide the most fitting explanation for the increase in perception of HMD weight and related discomfort.

Generally speaking, although many physical features of VR HMDs can be assessed in a fully objective manner (e.g., weighing the HMD, noting its resolution, measuring the range of adjustable parts), these results highlight the importance of including subjective metrics and different scenarios of usage when attempting to evaluate the quality of specific devices. Likewise, these results show that **it makes sense to include device-related metrics of discomfort in studies evaluating different tasks or applications, even if they are using the same hardware setup**. Analyzing VRISE that occur using different combinations of HMDs and software, especially active games, is likely to be of benefit to hardware manufacturers, software developers, and — above all — to users. Such approaches may serve as a potential step toward preventing injuries resulting from the unbalanced combination of heavy devices and in-game requirements for fast-paced movements, which is considered to be the likely cause behind the fractured vertebra in the case described by [190].

5.4.4 Cybersickness

Findings

To gain a better understanding of the impact of tested VR games on the intensity of cybersickness, VR-induced differences in individual symptoms were analyzed before moving onto further calculations. Mean differences between post-VR and pre-VR ratings for individual SSQ symptoms, as well as Wilcoxon signed-rank test results comparing post-VR and pre-VR ratings, are presented in Table 5.5. Overall, games in *Study 2* produced significantly larger increases in reported symptoms compared to games in *Study 1*. While statistically significant increases in *Study 1* were found only for 2 to 3 symptoms per game, in *Study 2* they were found for 7 (both FN and DB) and 10 (SPT) out of 16 symptoms in the SSQ.

None of the symptoms experienced a statistically significant post-VR increase for every single tested game. Statistically significant increases, however, were found following the majority of tested games (i.e, 4 out of 6 games) for these particular symptoms: general discomfort, eye strain, difficulty focusing, and blurred vision. Sweating significantly increased following 3 games, while statistically significant increases were found for 2 out of 6 games for fatigue,

Table 5.5: Mean post-pre differences (PPD) in symptoms (presented alongside accompanying standard deviations) and Wilcoxon signed-Rank test (WSRT) Z-scores comparing post-VR and pre-VR symptom scores (* $p < 0.05$, ** $p < 0.005$, *** $p < 0.001$); (taken from [30])

				Study 1									Study 2								
	Symptom of cybersickness (as measured by the SSQ)	Included in:		BS			OU			SS			FN			DB			SPT		
				PPD		WSRT	PPD		WSRT	PPD		WSRT	PPD		WSRT	PPD		WSRT	PPD		WSRT
		CSQ	VRSQ	M	SD	Z	M	SD	Z	M	SD	Z	M	SD	Z	M	SD	Z	M	SD	Z
1	General discomfort		X	0.30	0.64	-1.90	0.15	0.73	-0.90	0.45	0.67	**-2.50***	0.60	0.66	**-3.00****	0.70	1.00	**-2.57***	0.55	0.50	**-3.32*****
2	Fatigue		X	0.25	1.09	-0.91	0.10	0.77	-0.58	0.30	0.84	-1.50	0.50	0.74	**-2.50***	0.40	0.92	-1.81	0.55	0.97	-2.23*
3	Headache	X	X	0.10	0.54	-0.82	0.00	0.55	0.00	0.05	0.38	-0.58	0.15	0.36	-1.73	0.30	0.56	**-2.12***	0.40	0.73	**-2.07***
4	Eye strain	X	X	0.40	0.86	-1.90	0.65	0.79	**-2.97****	0.45	0.67	**-2.50***	0.40	0.58	**-2.53***	0.30	0.90	-1.39	0.65	0.85	**-2.81****
5	Difficulty focusing	X	X	0.45	0.74	**-2.31***	0.30	0.64	-1.90	0.20	0.75	-1.15	0.50	0.67	**-2.64***	0.40	0.66	**-2.31***	0.55	0.97	**-2.18***
6	Increased salivation			0.10	0.30	-1.41	-0.05	0.22	-1.00	-0.05	0.22	-1.00	0.05	0.22	-1.00	0.10	0.30	-1.41	0.20	0.51	-1.63
7	Sweating			1.05	0.74	**-3.52*****	0.05	0.50	-0.45	0.25	0.70	-1.51	0.60	0.66	**-2.97****	0.25	0.54	-1.89	0.65	1.01	**-2.41***
8	Nausea	X		0.10	0.44	-1.00	-0.10	0.54	-0.82	0.05	0.22	-1.00	0.15	0.36	-1.73	0.15	0.48	-1.34	0.05	0.22	-1.00
9	Difficulty concentrating			0.05	0.86	-0.31	0.25	0.83	-1.31	0.05	0.80	-0.26	0.05	0.59	-0.38	0.10	0.62	-0.71	0.45	0.80	**-2.18***
10	"Fullness of the head"	X	X	0.20	0.51	-1.63	0.25	0.89	-1.19	0.20	0.60	-1.41	0.30	0.71	-1.73	0.70	0.95	**-2.66***	0.55	0.74	**-2.60***
11	Blurred vision	X	X	0.30	0.46	**-2.45***	0.15	0.48	-1.34	0.15	0.73	-0.90	0.30	0.46	**-2.45***	0.40	0.80	**-2.11***	0.55	0.80	**-2.64***
12	Dizzy (eyes open)	X		0.15	0.36	-1.73	0.15	0.36	-1.73	0.10	0.30	-1.41	0.10	0.54	-0.82	0.20	0.40	**-2.00***	0.40	0.80	**-2.06***
13	Dizzy (eyes closed)	X	X	0.05	0.22	-1.00	0.20	0.40	**-2.00***	0.00	0.32	0.00	0.05	0.74	-0.28	0.40	0.66	**-2.27***	0.15	0.48	-1.34
14	Vertigo	X	X	0.00	0.00	0.00	-0.10	0.44	-1.00	0.00	0.00	0.00	0.20	0.40	**-2.00***	0.20	0.51	-1.63	0.10	0.30	-1.41
15	Stomach awareness			0.00	0.55	0.00	-0.05	0.50	-0.45	-0.05	0.38	-0.58	0.15	0.36	-1.73	0.10	0.30	-1.41	0.10	0.30	-1.41
16	Burping			0.05	0.22	-1.00	0.00	0.00	0.00	0.10	0.30	-1.41	-0.10	0.44	-1.00	0.00	0.00	0.00	0.00	0.00	0.00

Table 5.6: Mean post-pre differences (PPD) in symptom group scores (presented alongside accompanying standard deviations) and Wilcoxon signed-Rank test (WSRT) Z-scores comparing post-VR and pre-VR symptom group scores (* $p < 0.05$, ** $p < 0.005$, *** $p < 0.001$); (taken from [30])

		Study 1									Study 2								
Questionnaire -Dimension	Symptoms included in the calculation	BS			OU			SS			FN			DB			SPT		
		PPD		WSRT	PPD		WSRT	PPD		WSRT	PPD		WSRT	PPD		WSRT	PPD		WSRT
		M	SD	Z	M	SD	Z	M	SD	Z	M	SD	Z	M	SD	Z	M	SD	Z
SSQ-N	1,6,7,8,9,15,16	15.74	20.07	**-3.19*****	2.39	17.03	-0.90	7.63	9.82	**-2.84****	14.31	18.47	**-2.89****	13.36	17.95	**-2.78***	19.08	18.84	**-3.27****
SSQ-O	1,2,3,4,5,11	14.02	27.25	**-1.97***	12.13	25.07	**-2.15***	12.51	19.07	**-2.45***	18.95	18.80	**-3.22****	19.71	21.09	**-3.36*****	28.05	27.13	**-3.56*****
SSQ-D	5,8,9,10,11,12,13,14	17.40	22.82	**-2.84****	11.83	32.73	**-2.14***	9.74	25.33	-1.65	22.27	35.32	**-2.50***	34.10	40.22	**-3.50*****	32.71	37.16	**-3.44*****
SSQ-T	all symptoms	17.77	24.86	**-2.77***	10.10	24.67	**-2.57***	11.78	17.99	**-2.47***	20.94	23.37	**-3.13****	24.12	25.16	**-3.71*****	30.11	26.56	**-3.74*****
CSQ-D	3,8,12,13,14	0.32	0.74	-1.71	0.19	1.06	-1.26	0.16	0.56	-1.02	0.45	1.60	-1.12	0.96	1.41	**-2.88****	0.76	1.17	**-2.55***
CSQ-DF	4,5,10,11	0.99	1.36	**-2.62***	0.85	1.38	**-2.33***	0.67	1.44	**-2.07***	1.00	1.12	**-3.18****	1.14	1.30	**-3.18****	1.47	1.38	**-3.43*****
VRSQ-O	1,2,4,5	11.67	19.79	**-2.23***	10.00	17.60	**-2.45***	11.67	15.00	**-2.67***	16.67	14.19	**-3.67*****	15.00	17.00	**-2.80***	19.17	20.43	**-3.16****
VRSQ-D	3,10,11,13,14	4.33	7.39	**-2.43***	3.33	13.90	1.84	2.67	8.27	-0.98	6.67	10.33	**-2.34***	13.33	16.06	**-3.40*****	11.67	12.63	**-2.97****
VRSQ-T	1,2,3,4,5,10,11,13,14	8.00	12.75	**-2.37***	6.67	14.70	**-2.07***	7.17	10.71	**-2.41***	11.67	11.27	**-3.40*****	14.17	13.66	**-3.55*****	15.42	13.25	**-3.82*****

headache, fullness of the head, and dizziness with eyes open as well as closed. No significant increases were found for salivation, nausea, stomach awareness, and burping.

The scores were also calculated for different symptom groups. SSQ scores were calculated as explained in [114] (note, the formula used to calculate SSQ-T includes the brackets that were missing from the original publication, as described by [207]), CSQ scores were calculated as explained in [115], and VRSQ scores were calculated based on [116]. As different dimensions of cybersickness for the three questionnaires were calculated, post-VR increases in individual symptom ratings were compounded as multiple symptoms were joined together and multiplied with weighing factors. This resulted in statistically significant increases in calculated SSQ dimensions, as well as CSQ and VRSQ dimensions, especially for games in *Study 2*. Mean differences between post-VR and pre-VR ratings for calculated symptom groups, as well as Wilcoxon signed-rank test results comparing post-VR and pre-VR scores, are presented in Table 5.6. It is worth noting that both total cybersickness scores (SSQ-T and VRSQ-T) showed statistically significant increases following VR gameplay for all tested games across both studies.

While there were significant differences comparing pre-VR and post-VR scores of each symptom, when comparing the post-pre differences in cybersickness between games, Friedman tests followed by Wilcoxon signed-rank tests with Bonferroni correction failed to identify significant differences for either study, except for sweating in Study 1 ($\chi^2(2) = 20.31$, $p < 0.001$), as BS differed significantly from both OU ($Z = -3.70$, $p < 0.001$) and SS ($Z = -2.89$, $p = 0.004$). In line with this, significant differences between games were not found for either calculated group of symptoms (calculated for SSQ, VRSQ, and CSQ).

Discussion

More significant increases of reported symptoms for all games in *Study 2* compared to *Study 1* are likely attributable (at least in part) to different participant populations. Although similar in terms of gender and age distribution, according to self-reported data, participants in *Study 2* were reported greater susceptibility to motion sickness and cybersickness compared to participants in *Study 1*. Moreover, participants in *Study 1* were more experienced gamers, and even slightly more experienced with the use of VR, which could have led to cybersickness adaptation.

In addition to general discomfort — a rather vague symptom — symptoms that statistically increased for the highest number of games were predominantly those that are related to eyes and vision (eye strain, difficulty focusing, blurred vision). Consequently, all six of the tested games produced statistically significant increases in all individual questionnaire dimensions that used at least three of these four symptoms in their calculation — SSQ-O, VRSQ-O, CSQ-DF. Two of these symptoms were also present in the calculation for SSQ-D, which yielded statistically significant increases for 5 out of 6 games. Notable increases in intensity following VR gameplay were found for the majority of remaining symptoms included in the calculation of the SSQ-D score, although they only reached statistical significance for either one or two out of the six games, depending on the symptom.

An equally high number of statistically significant increases (i.e., 5 out of 6 games) was found for the SSQ-N dimension. Based on this score, one would expect that participants in our studies struggled with gastrointestinal symptoms — nausea, burping, salivation, and stomach awareness. However, it was quite the contrary, as these symptoms happened to be the least prominent of all symptoms measured by the SSQ. Instead, increases in SSQ-N scores can be attributed to significant increases in general discomfort and sweating, with both symptoms being somewhat open to interpretation.

Considering that the original questionnaire was designed to be used with simulators and not active games, the inclusion of sweating in the SSQ presumably pertains primarily to diaphoresis, i.e., increased sweating which is not caused by physical activity or environmental factors (temperature, humidity). In case of an otherwise stationary VR game that has vection-producing locomotion mechanics and is being played in comfortable environmental conditions, a notable increase in sweating could indeed be classified as diaphoresis, and as such it would indicate the onset of cybersickness. However, with physically active standing/room-scale games determining whether this symptom comes as a result of exertion or sickness — or some combination of the two — is not as straightforward. Moreover, sweating may also occur as a normal physiological response to increased HMD temperature during a prolonged period of usage.

Results presented in this chapter showed that the post-VR increase in sweating did not correlate significantly with total cybersickness (calculated according to the VRSQ-T formula because of its independence from the influence of sweating) for either study, but positive correlation was found between sweating and PHYS (SIM-TLX) for both *Study 1* ($r_s = 0.25$, $p = 0.046$) and *Study 2* ($r_s = 0.46$, $p = 0.004$). A significant positive correlation between sweating and discomfort related to HMD temperature was found only for *Study 2* ($r_s = 0.43$, $p < 0.001$). With this in mind, it can be assumed that, in this case, the post-VR SSQ-N score was likely a result of increased physical activity or headset temperature rather than cybersickness.

Both CSQ and VRSQ omitted the ambiguous sweating item present in the SSQ, which makes both of them more resistant to the confounding influence of physical exertion that occurs during active VR gaming, and thus possibly more suitable for use in studies such as ours. However, as explained by [23], sweating has a notable negative impact on user experience with active VR games, and as such we still feel that including it as an item in questionnaires evaluating VRISE could provide valuable information. A potential solution to the ambiguity of this symptom is to separate it into different symptoms, similarly to the distinction made in the Motion Sickness Assessment Questionnaire (MSAQ) [208], which separates the feelings of being hot and sweaty from cold sweat/clamminess. Similarly, **finding alternative ways to better define other ambiguous symptoms such as fatigue or general discomfort — which**

could also be influenced by physical exertion alone — may aid researchers with identifying the etiology of certain symptoms.

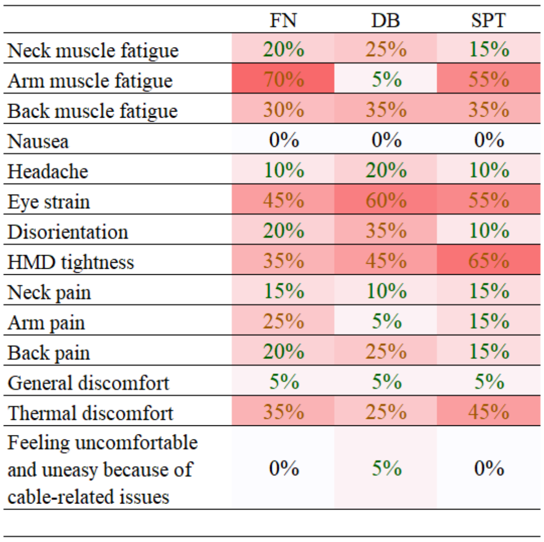
As previously mentioned, sweating is one of the SSQ symptoms that were excluded from both CSQ and VRSQ. Both questionnaires also exclude increased salivation, stomach awareness, and burping — symptoms that did not increase significantly for any of the tested games — as well as difficulty concentrating, which was only significant for SPT. Based on the presented results, these items did not contribute to valuable findings regarding VRISE during active VR games with no in-game locomotion, so their inclusion may not be necessary, although their importance would likely increase in case of more vection-provoking games. Although there are some differences between the two SSQ-variations, the majority of symptoms that were most influenced by VR gaming were included in both versions.

5.4.5 Overall prevalence and ranking of reported VRISE

Findings

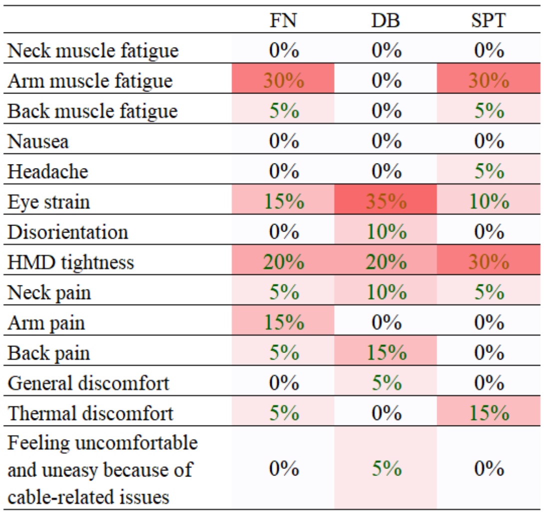
Although the occurrence of different VRISE was evaluated through individual specialized questionnaires, participants were also provided with an opportunity to evaluate their experience as a whole. Therefore, for *Study 2*, participants were asked to report all VRISE experienced during gameplay, as well as the single most bothersome VRISE for the particular session. While a predefined list of VRISE was provided for participants to choose from, they were also encouraged to add their own answers in case they experienced other VRISE. Only one participant decided to add a non-predefined option (*"feeling uncomfortable and uneasy because of cable-related issues"*), only in case of DB. The overall prevalence of VRISE (reported as the overall percentage of participants) for each session is depicted on the heat map in Figure 5.5. Nausea was the only symptom that was not reported by any participant for any game. Broadly speaking, the most prevalent VRISE were muscle fatigue in the arms and back, eye strain, thermal discomfort, and HMD tightness, however, in certain cases the prevalence of individual symptoms was highly dependant on the game (e.g., arm muscle fatigue ranging from 5% for DB to 70% for FN).

Figure 5.6 presents percentages of participants who reported each symptom/effect as the most bothersome of all experienced VRISE. Except for eye strain, symptoms commonly attributed to cybersickness were generally not considered to be the most bothersome for any of the three games tested in this study. Instead, the largest number of participants (6) reported arm muscle fatigue as the most bothersome symptom for FN, with the majority of other participants choosing either HMD tightness, eye strain or arm pain. Similar results were obtained for SPT, although a larger percentage of participants reported thermal discomfort, while none opted for arm pain. The results for DB were more distinguished, with the majority of participants choosing eye strain, followed by HMD tightness and back pain.



	FN	DB	SPT
Neck muscle fatigue	20%	25%	15%
Arm muscle fatigue	70%	5%	55%
Back muscle fatigue	30%	35%	35%
Nausea	0%	0%	0%
Headache	10%	20%	10%
Eye strain	45%	60%	55%
Disorientation	20%	35%	10%
HMD tightness	35%	45%	65%
Neck pain	15%	10%	15%
Arm pain	25%	5%	15%
Back pain	20%	25%	15%
General discomfort	5%	5%	5%
Thermal discomfort	35%	25%	45%
Feeling uncomfortable and uneasy because of cable-related issues	0%	5%	0%

Figure 5.5: Heat map depicting the overall prevalence of VRISE during gameplay (taken from [30]) — darker color indicates higher percentage



	FN	DB	SPT
Neck muscle fatigue	0%	0%	0%
Arm muscle fatigue	30%	0%	30%
Back muscle fatigue	5%	0%	5%
Nausea	0%	0%	0%
Headache	0%	0%	5%
Eye strain	15%	35%	10%
Disorientation	0%	10%	0%
HMD tightness	20%	20%	30%
Neck pain	5%	10%	5%
Arm pain	15%	0%	0%
Back pain	5%	15%	0%
General discomfort	0%	5%	0%
Thermal discomfort	5%	0%	15%
Feeling uncomfortable and uneasy because of cable-related issues	0%	5%	0%

Figure 5.6: Heat map depicting the percentage of participants who reported a particular symptom/effect as the most bothersome of all VRISE (taken from [30]) — darker color indicates higher percentage

Furthermore, one of the research goals in *Study 2* was to examine whether VRISE influenced the participants' intention to continue playing. An overview of participants' willingness to terminate the gaming session or continue playing is presented in Figure 5.7. Participants who reported they would not be willing to continue playing after the 20 minute session, as well as those who stated they wished that the session was terminated sooner, were asked to specify the reasons behind their preference (i.e., reasons for game termination). The collected qualitative data was analyzed and each answer was coded as belonging to one or more of the three broad categories: *VRISE*, *playability issues* (PI) and *technical issues* (TI).

Among the five answers collected after the FN session, one included only PI, two included a combination of VRISE and PI, and two included only VRISE as a reason for game termination. Thus, the mention of VRISE was present in four out of five answers. For SPT, the mention of

VRISE was present in all four obtained answers, two of which also included PI. The results differed for DB. Five out of 10 answers only mentioned PI, a single answer only mentioned TI, and two answers mentioned only VRISE. The remaining two answers involved a combination of TI and VRISE. Thus, the mention of VRISE was present in four out of ten answers. Combining all 19 answers across the three games, we found 18 distinct mentions of different VRISE as reasons for game termination. We categorized these mentions as follows (the number in parentheses represents the the number of mentions): muscle fatigue/muscle pain (5), overall fatigue (4), headache (3), eye strain (3), general discomfort (2), and disorientation (1).

Discussion

When observing the overall prevalence of VRISE, it can be concluded that VR gaming triggered a diverse set of uncomfortable symptoms, although many were experienced only by a minority of participants. Certain symptoms had a similar overall prevalence regardless of the game in question (e.g., back muscle fatigue, neck pain, general discomfort), while others varied significantly between games depending on their mechanics — e.g., arm muscle fatigue was reported by a single player in DB, as opposed to 14 out of 20 players in FN.

Focusing on the most bothersome symptom, there was less diversity in the results, as several symptoms — such as neck muscle fatigue, nausea, headache, and general discomfort — were either highlighted by a single participant or none at all. However, we did not find a specific symptom that was overwhelmingly chosen as the most bothersome. Instead, the results varied between games and individual participants, showcasing the diversity of the VR experience. It is still worth mentioning that HMD tightness, arm muscle fatigue, and eye strain stood out as more frequently reported compared to other symptoms.

Similarly to [25], these results showed that symptoms of cybersickness (aside from eye strain, which is one of the symptoms included in the SSQ) did not stand out as either the most frequent or the most bothersome VRISE for the tested games. Considering that cybersickness is often connected to vection, it is not unusual for it to be less pronounced in games with no in-game locomotion, such as those used in our study. However, knowing that the SSQ is often used as a primary measure of VRISE [25] — even when evaluating games without in-game locomotion (e.g., [196, 209]) — it is important to highlight that **other VRISE (which are generally less researched) may pose a larger threat to the comfort and safety of VR gamers compared to cybersickness, and should thus be included in studies dealing with VRISE, especially when examining user experience with physically active games.**

Reported reasons for game termination included both VRISE and playability issues for all three games, while technical issues were only reported for DB. **According to the presented results, VRISE were the leading (i.e., most frequently reported) cause behind the participants' desire to terminate the gaming experience for the two more physically demanding**

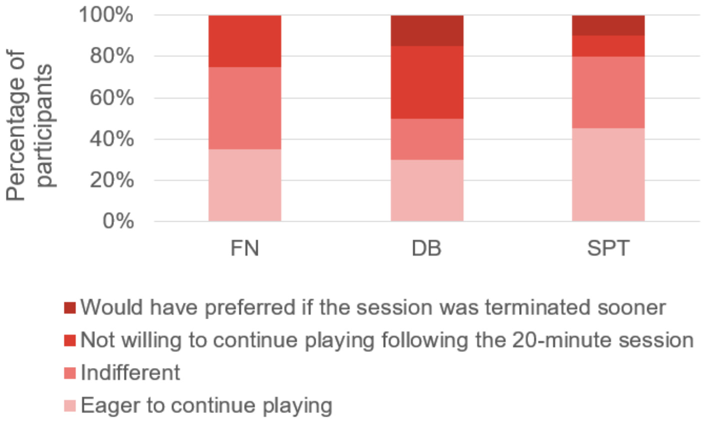


Figure 5.7: Willingness to terminate game/continue playing following each 20-minute gaming session (taken from [30])

games (FN and SPT). Based on their frequency of occurrence in the pooled results (i.e., combined reports from all three games), musculoskeletal symptoms (muscle pain and fatige) and general fatigue were the main culprits. The intensity, as well as the prevalence, of aforementioned VRISE is expected to increase as play time surpasses 20 minutes (i.e., gaming session duration across both studies), but even at such a short interval, these findings support the need for further advancements. However, a significant majority of participants in FN and SPT did not feel the need to terminate the game following the 20 minute gaming session, which is encouraging.

5.4.6 Reaction time

Findings

For *Study 1*, all games except for SS resulted in a longer post-VR SRT, with an average post-pre difference of 14.95ms (SD = 27.44ms) and 22.53 (SD = 25.67ms) for BS and OU, respectively, while the mean difference for SS was negligible (M = 1.15ms, SD = 20.16ms). However, the only game that resulted in a statistically significant increase in SRT was OU (Z = -3.17, p = 0.002). According to the Friedman test, no statistically significant difference was found in post-pre differences for the three games ($\chi^2(2) = 3.90$, p = 0.14).

Similarly, for *Study 2*, two of the games resulted in a notable mean increase in post-pre differences — DB (M = 29.98ms, SD = 28.87ms) and SPT (M = 18.68ms, SD = 35.85ms) — with a negligible mean difference for FN (M = 1.65ms, SD = 27.81ms). Again, only the pick-and-place game (DB) produced a statistically significant increase in SRT (Z = -3.51, $p < 0.001$). Friedman test ($\chi^2(2) = 6.7$, p = 0.04) followed by post hoc analysis with Wilcoxon signed-rank tests and Bonferroni correction showed a statistically significant difference between DB and FN (Z = -2.98, p = 0.003).

Table 5.7: Spearman correlation coefficients (r) between post-pre difference in SRT and different dimensions of workload for each game (taken from [30])

	Study 1			**Study 2**		
r_s	**BS**	**OU**	**SS**	**FN**	**DB**	**SPT**
MENT	0.35	-0.04	-0.13	-0.31	**-0.67***	-0.47
PHYS	0.42	0.13	-0.20	-0.02	-0.26	-0.37
TEMP	**0.49***	0.10	0.03	0.01	-0.34	-0.33
FRUS	**0.47***	0.21	0.06	-0.18	-0.19	**-0.58***
CMPX	**0.48***	-0.06	**-0.47***	**-0.69***	-0.44	-0.40
STRS	**0.60****	0.03	0.00	-0.39	-0.38	-0.37
DIST	0.12	-0.35	-0.16	-0.05	-0.17	-0.40
PERC	0.04	-0.39	-0.13	0.21	0.07	-0.36
CONT	0.17	**0.50***	0.20	-0.35	0.02	-0.26
SUM	**0.57****	0.12	-0.15	-0.34	-0.43	-0.55

* $p < 0.05$, ** $p < 0.005$

It is important to note that changes in reaction speed vary across players, and an individual player's cognitive performance may also vary greatly after exposure to different games/mechanics. Overall, the majority of participants experienced an increase in RT after VR. This effect was more pronounced for pick-and-place games, as 80% of users experienced varying increases in RT after OU, and 90% after DB. For some individuals, VR exposure resulted in a RT increase of up to 93ms for DB and 112ms for OU and SPT. However, a significant number of participants showed faster responses after certain game scenarios. Post-VR RT decreases were noted in 45% of participants following SS and 50% of participants following FN, and certain individuals experienced post-VR RT improvements of up to 40-50ms following SS, FN, and SPT.

Table 5.7 presents Spearman correlation coefficients between post-pre differences in SRT and workload and its individual dimensions. Several statistically significant moderate to strong correlations were found across different games and different dimensions of workload. However, BS was the only game for which we found statistically significant correlations between post-pre SRT difference and multiple dimensions of workload, as well as the total workload score. After calculating Spearman correlation coefficients between post-pre differences in SRT and post-pre differences in overall cybersickness and its subscales (calculated for SSQ, VRSQ, and CSQ), the only statistically significant correlation ($r_s = 0.58$, $p = 0.007$) was the one between post-pre SRT difference and post-pre SSQ-N difference for the OU scenario.

Discussion

The overall magnitude and statistical significance of RT changes varied based on genre/mechanics. [193] listed adaptation to latency introduced by the system as one of the potential reasons for VR-induced impairment of reaction speed, which may explain why playing pick-and-place games resulted in slower reaction times. [177] have shown that throwing an object in the real world is superior to throwing it in virtual reality in terms of both precision and accuracy. Although the motion of the ejected projectile was found to be physically plausible, participants in their user study reported problems with the timing of releasing the projectile from the virtual hand, which likely affected their throwing performance. This slight delay can be attributed to the action of releasing the grip of the controller trigger, as it takes some time for this action to be physically performed by the user and subsequently registered by the system. Even though participants in both studies did not perform throw mechanics, both pick-and-place games were similar in terms of controller mappings for grabbing and releasing the target (i.e., pulling and releasing the controller trigger, respectively). As opposed to pick-and-place games, the use of trigger press and release in shooters is more in line with the corresponding real-life action of shooting, while playing swordplay-based games did not involve the use of discrete controls (i.e., triggers and buttons) at all. Moreover, both shooters and swordplay-based games used mediated interaction mechanics, with the controller serving as a physical substitute for a tool/weapon. With pick-and-place mechanics as realized in our setup, the controller was used for tracking the virtual hand which is supposed to represent the player's own hand, but their alignment is far from perfect. In their attempt to control the movement of their virtual hand despite this visuomotor disturbance, participants needed to alter their reaching movements to compensate for the misalignment which could have resulted in prolonged RT [193, 194].

While these theories provide possible explanations for the increase in RT following pick-and-place simulation games, highlighting the genre's effect on cognitive performance compared to other genres, there is no way to provide a cohesive explanation for cognitive performance changes (or lack thereof) following exposure to other tested mechanics. The comparatively unimpaired post-SS reaction time can potentially be explained by previous work showing that playing shooter games during training sessions causes a significant reduction in reaction time (RT) compared to playing a control game [210], although any effects noted in this thesis were measured immediately following short-term exposure, as opposed to a long multi-episodic training period.

Another aspect that may have influenced our results is the temporary acceleration of reaction speed which was shown to occur shortly after a period of exercise or active VR gaming [196, 211], potentially explaining why BS and FN caused smaller impairments compared to OU and DB. Both theories, however, fail to explain why partaking in SPT, a shooter as well as a physically demanding game, produced a fairly large mean increase in RT. Nevertheless, these results indicate VR-induced changes in reaction speed can not be explained only by cybersickness/visual motor cues, but are rather diverse in magnitude and dependent on factors such as interaction mechanics, control modality, and workload imposed by the particular VR application. Moreover, **the implications that pick-and-place tasks produce the most significant changes in RT can be explored in future research, and potentially be used as a**

benchmark for the naturalness and the overall quality of a particular implementation of pick-and-place mechanics. It would also be interesting to see whether using hand tracking or a controller that facilitates more natural grasp/release actions (e.g., Valve Index) would lead to improvements in RT for games with pick-and-place mechanics.

Although changes in reaction time following virtual reality exposure are commonly attributed to cybersickness [191, 192], this explanation may be better suited for VR applications that present the users with very obvious motion clues and less control over their locomotion. As chosen games were based on frequent manual interaction but without in-game locomotion, these results are more in line with sources that highlight the influence of other factors [193, 196]. **According to these results, workload measurements obtained using the SIM-TLX questionnaire reveal more significant relationships with changes in SRT in comparison to different SSQ-based measures of cybersickness.**

5.5 Key takeaways

Over the years since its conception, manufacturers of VR technology have been facing challenges pertaining to its impact on user comfort. One of the biggest challenges, which is still ongoing, is in preventing the state known as cybersickness, which presents itself with symptoms such as disorientation, nausea, and oculomotor difficulties. A large body of work pertaining to this issue has highlighted its potential causes and possible solutions, which were later incorporated into guidelines and best practices for VR developers of today to abide by. However, increased efforts into cybersickness and its causing factors came at the expense of research into other symptoms and effects that may occur with VR use, such as musculoskeletal discomfort, issues with headset design, and impact on cognitive performance, which are especially relevant for VR scenarios characterized by significant physical activity.

Fortunately, the need for a more holistic approach to evaluating and eventually preventing and/or minimizing negative effects of VR use, which has long remained fairly unrecognized by the research community, is finally starting to gain traction, as witnessed by a rapidly increasing number of recently published related works that highlight the importance of VRISE other than cybersickness. This is not to suggest that the large body of research exploring cybersickness has been unnecessary, or its focus misplaced. After all, the issue of cybersickness likely has a smaller impact on VR users of today precisely because of decades-long research efforts toward understanding its causes and minimizing its symptoms, both in terms of intensity and frequency of occurrence.

While numerous articles in the field have focused on individual symptoms of VR use, the aim of this chapter was to paint a more complete picture of the user experience by extending the focus to explore a number of different VRISE at once, from workload and musculoskeletal issues, to device-related discomfort, cybersickness, and changes in reaction time. When collecting and analyzing data, the choice was made to incorporate state-of-the-art VR-specific measures (SIM-TLX, CSQ, VRSQ) as opposed to the more frequently used questionnaires (NASA-TLX, SSQ). Focusing on the gaming use-case and recognizing its diversity with regard to navigation techniques, visual effects and game mechanics, this methodology was used to evaluate a total of six games (three per study) that fit a pre-defined set of standards (i.e., standing or Room-scale games with controller based isomorphic controls, belonging to popular genres with commonly used mechanics).

After conducting two studies, with a total of 40 young, healthy participants, it was found that 20 minutes of VR gameplay produced statistically significant increases in a variety of different VRISE. In addition to producing a *Moderate* to *Somewhat high* workload, VR gaming contributed toward increased muscle fatigue and/or pain in the arms and back. Different mobility requirements imposed by different games led to varying intensities of device-related discomfort even though the same HMD was used during the entirety of both studies. While notable post-VR increases in symptoms of cybersickness were found, they were mostly constrained to those outside of the gastrointestinal category, with most significant increases found for oculomotor symptoms. This is in line with the selection of symptoms included in the more recent SSQ-based questionnaires — CSQ and VRSQ. Combined with the issue of ambiguous effects (e.g., sweating) that can be misinterpreted as cybersickness despite being caused by other factors, these findings question the suitability of the original SSQ for the evaluation of active VR gaming scenarios.

Overall, exposure to VR was shown to increase reaction time by a small margin, an effect that — while still unclear — has already been discussed in other sources. Presented results showed that this increase reached statistical significance only in case of pick-and-place games, which calls for further research into this type of mechanics. Furthermore, although several other works in the field attribute changes in reaction time to cybersickness, these results suggest a possible link with workload instead. With regard to the prevalence of different VRISE, it is worth noting that different individuals experience the same game differently, and different games may produce very different effects on the same person. Still, the diversity of these results further confirms the necessity of tracking multiple VRISE in addition to cybersickness.

Results presented herein were collected during short VR sessions, using a small sample of fairly homogenous and inexperienced individuals, and a limited set of games. Extending beyond those individual limitations in future work would likely lead to more relevant conclusions. However, findings obtained in such a controlled setting enabled the identification of the possible connections between specific game elements and VR-induced discomfort to be explored in future work. Further dissecting each of these elements (e.g., play angle, target placement) and evaluating their parameters in future studies may provide insights that could help developers with designing experiences that strike a balance between fun and challenging while remaining mindful of the comfort and safety of their users.

5.6 Modeling the impact of game genre/mechanics on workload and VRISE

Chapter 3 proposes three low-level models of VR gaming QoE. The first model, labeled as VR_QOE_LLM_1, suggests that certain game characteristics (best described as the combination of genre and game mechanics) may influence various dimensions of workload, as well as different types of VRISE. To inspect whether the results of Study 1 and/or Study 2 support the hypothesized relationships presented in the model, we considered the outputs of statistical analyses presented in the Results section of this chapter.

During the process of analyzing collected data, two distinct approaches to statistical analysis were taken (the results of both are included in the thesis). In the first approach, the focus was on the effect of each individual game on measured features. Where possible, measures taken after playing a particular game (post-VR measures) were compared to baseline measures taken before (pre-VR measures), and statistically significant changes in QoE feature scores were investigated using the Wilcoxon signed-rank test. This approach was useful in determining the overall ability of individual games — and VR gaming overall — to induce substantial changes in observed features, namely those pertaining to VRISE.

The second approach focused on looking for significant differences *between games* using the Friedman test and the following post hoc analysis. This approach was straightforward for measures that were taken only once, following VR use (all dimensions of workload, HMD-related discomfort). For features that were measured both before and after VR use (reaction time, cybersickness symptoms, pain and muscle fatigue), the pre-VR result was subtracted from the post-VR result to find the post-pre difference. The Friedman test was then used to compare these post-pre differences between games.

Because the proposed low-level model of gaming QoE focuses on the impact of game (genre/mechanics) choice, the second approach was utilized for the identification of significant relationships. The outcomes are presented in Table 5.8. Identified significant relationships between QoE IFs and QoE features are indicated by the X symbol. Figure 5.8 presents the adjusted model of VR gaming QoE (VR_QOE_LLM_1SR), based on the initial proposed model (VR_QOE_LLM_1), but condensed to include only significant relationships.

Table 5.8: The impact of game (genre/mechanics) choice on QoE features pertaining to QoE constituents workload and VRISE (based on the proposed VR_QOE_LLM_1 model)

		Tested set of games (Study ID)	
QoE constituent	**QoE feature**	**BS, OU, SS (Study 1)**	**FN, DB, SPT (Study 2)**
Workload	Mental demand		
	Physical demand	X	X
	Temporal demand	X	X
	Frustration	X	
	Task complexity		X
	Situational stress		X
	Distractions	X	
	Perceptual load		
	Task control difficulty	X	X
VRISE	Reaction time		X
	Cybersickness	X (sweating)	
	Pain and muscle fatigue		
	Device-related discomfort	X (HMD temperature, display quality, HMD cable)	X (HMD weight, HMD cable)

X denotes QoE features significantly affected by game (game genre/mechanics) choice

5.7 Chapter summary

Focusing on active VR gaming, this chapter presents the results of two user studies with a total of 40 participants. With state-of-the-art VR-specific measures (SIM-TLX, CSQ, VRSQ) incorporated into the methodology, the focus was on assessing different dimensions of workload, musculoskeletal discomfort, device-related discomfort, cybersickness, and changes in reaction time. Using a set of six different active VR games (three per study), the aim was to quantify and compare the prevalence and intensity of VR-induced symptoms across different genres and game mechanics. Varying between individuals, as well as games, the diverse symptoms reported in both studies highlight the importance of including the measures of VR-induced effects other than cybersickness into VR gaming user studies, while questioning the suitability of the SSQ —— arguably the most prevalent measure of VR discomfort in the field —— for use with active VR gaming scenarios. While the results of the two studies provide relevant insights and serve as further input for the validation and development of the proposed low-level model of VR gaming QoE, they also call for a more detailed dissection and evaluation of popular VR interaction mechanics, which is provided in the following chapter.

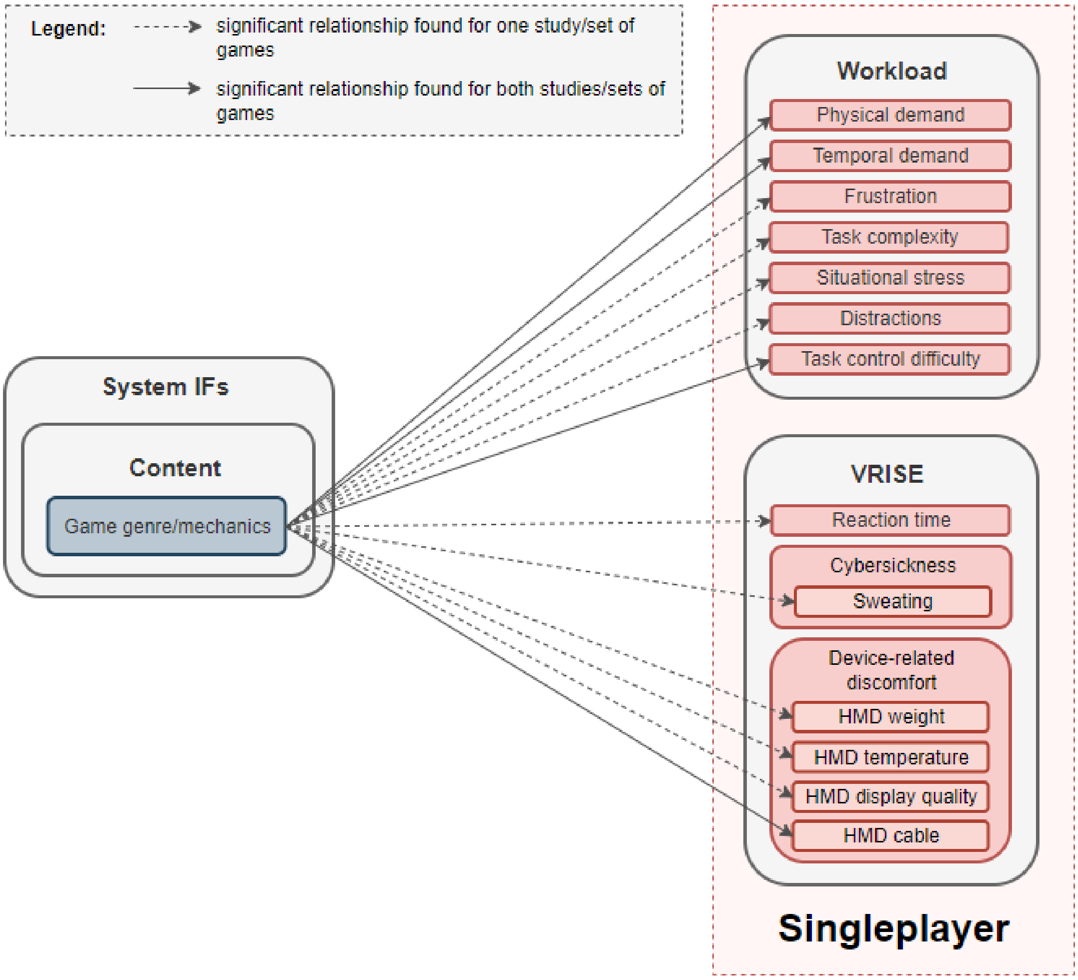


Figure 5.8: Partial low-level model (VR_QOE_LLM_1SR) of QoE for VR gaming depicting the significant relationships between game genres/mechanics and chosen QoE features for a singleplayer scenario (based on the proposed VR_QOE_LLM_1 model)

Chapter 6

Evaluating game mechanics for VR games

6.1 Introduction

The results of Study 1 and Study 2 demonstrated that the prevalence and intensity of features pertaining to workload and VRISE may vary significantly between games. Interestingly, even games that were very similar with regard to genre and game mechanics (e.g., SS and SPT) were shown to elicit different effects regarding these features. The need to take a closer look — i.e., to further dissect and analyze multiple elements of the game, such as play angle and target placement — was presented as one of the key takeaways of the previous chapter. Addressing this issue and focusing on the implementation of IMs, this chapter lays out multiple influence factors and potential configurable parameters, describes the methodology for evaluating IM quality using custom test applications, and presents the findings of three user studies focused on exploring the impact of different IM configurations on selected features of workload, VRISE, and player experience. With Study 3 focused on the slash IM, Study 4 focused on the pick-and-place IM, and Study 5 focused on the shoot IM, the outcomes of presented research are utilized for the formulation of useful guidelines for the implementation of VR IMs. Sections 6.2, 6.3, and partially Subsection 6.4.1 (along with the accompanying figures) have previously been published as part of a journal article [10], alongside the contents of Chapter 4.

6.2 Considerations regarding the implementation of IMs

The process of performing core interaction mechanics (including the preparatory mechanics) consists of multiple stages (Fig. 6.1), even in case of simple mechanics that do not require excessive strategizing or complex movements. After first perceiving the target (often preceded by active searching), the player chooses their subsequent action. In most cases, this includes multiple cognitive and motor operations necessary to aim, track, and interact with the target in some way, e.g., by touching it, picking it up, shooting at it, etc. Following the preparatory stages such as perceiving the target, as well as aiming and tracking, the player instigates the process of interacting with the target. The resulting contact with the target can be categorized as either instant (e.g., touching the target with a sword) or delayed (e.g., shooting an arrow and waiting for it to reach its destination), as well as discrete (e.g., bullet hits and immediately destroys the target object) or continuous (e.g., slicing through a large object).

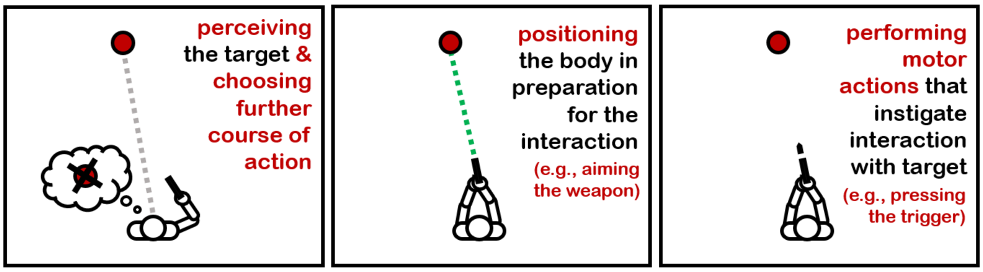


Figure 6.1: The multi-stage process of performing core interaction mechanics (taken from [10])

During this process, the player adapts their behaviour depending on various conditions provided by the game in relation to themselves and the means afforded to their in-game character. The player's action is a combined reaction to the state of the environment, target characteristics and positioning with respect to the player in the particular moment in time, and features of currently available tools. This reaction also depends on the fixed (e.g., size, fitness) and current (e.g., fatigue, mood) characteristics of the player combined with their personal tastes and preferences. A concise overview of various factors to be considered for the implementation and evaluation of interaction mechanics is presented below.

6.2.1 Perceiving a target

The way in which users allocate visual attention across the virtual environment can be described by the SEEV (salience effort expectancy value) model [212, 213], which will be explained in the remaining part of this paragraph. The *area of interest* (AOI) refers to the "*physical location*" in which the user can find specific information related to a certain task. Each environment is characterized by *salience*, which pertains to the physical characteristics of the AOI (e.g., color, size) that make it stand out from its environment. As the user's attention shifts between different AOIs, the process of scanning their surroundings requires a certain level of *effort*. The SEEV model also includes *expectancy* as one its factors, as the user is more likely to scan locations in which they expect certain changes to occur, and they base this expectation on the frequency of changes that occurred so far. The last factor of the SEEV model refers to the *value* of a certain AOI in the context of the overall task. The SEEV model, which is concerned with scanning the visual environment, serves as a basis for the NSEEV model [214], which describes the characteristics of a discrete event that have an impact on whether it is actually going to be noticed, such as the events' *eccentricity* (i.e., the offset between the fixation location and the event), its expectancy, as well as its salience.

The effort required to successfully shift between AOIs will differ based on their offset, as explained in [215]. If both areas fall within 20 degrees of visual angle (i.e., within the so-called *eye-field*, the necessary effort will be considerably lower compared to a situation in which head rotation is required (i.e., within the *head-field*, up to 90 degrees of visual angle). Likewise, an even greater angle will yield more effort, necessitating full-body rotation. A wider angle requires a lot more energy, especially with regular switching between the AOIs — thus, people are more likely to experience fatigue and discomfort, or even resist putting in the necessary effort.

As players scan the virtual environment, they are essentially performing visual search, using their eyes to scan the environment with the goal of finding a particular target, or targets. There are multiple factors affecting this process, such as the number of targets, distinguishing features of the target, the presence of non-targets, along with their overall number and their heterogeneity with respect to each other, and the dispersion of elements across the environment, which defines the necessary visual scanning distance [215].

6.2.2 Interacting with a target

The term *tracking task* can be used to describe a task that requires manually steering a controller through the virtual environment with the goal of reaching a target with an adequate level of precision [215]. Fitts' Law [216] describes the relationship between movement time, distance traveled, and target size for stationary targets. Initially referring to one-dimensional tasks, the original Fitts' Law has been extended to three dimensions in works such as [217], and, more recently, [218], highlighting the impact of direction/angle of target placement with respect to the user. The difficulty of tracking a target is further increased in case of a moving target, especially in case of fast, unpredictable movement across multiple axes [215].

Different target locations were shown to influence player comfort in different ways. Depending on target location, the task of following and manually interacting with targets requires different levels of muscle activity, and may even lead to neck and shoulder discomfort, with targets placed at eye height, or up to 15 degrees below eye height, appearing to be the most comfortable [187].

In their testbed for studying object manipulation methods in virtual reality, Poupyrev et al. [219] focused on pick and place tasks involving primarily stationary objects. The authors listed several parameters that are of interest for this type of task. Parameters that were considered to be relevant for the process of target selection included the number of objects, their size and their distance from the user, direction to the target with respect to the user, and occlusion of the target object, but the authors also mention target dynamics, density of objects surrounding the target, and target object bounding volume. In addition to aforementioned parameters pertaining to target selection, Poupyrev et al. [219] listed parameters related to the task of positioning objects, such as the initial object distance and direction with respect to the user, as well as the distance and direction to its terminal position, required precision, visibility, dynamics and occlusion of the terminal position. Furthermore, the authors list parameters related to the task of orienting objects such as initial and final orientation of the object, its distance and direction with respect to the user, as well as the required precision of orientation.

6.2.3 Tools and tool usage

Physical properties of handheld tools — such as their dimensions — may prove to be a highly relevant factor, especially when it comes to interaction mechanics that utilize tools as extensions of peripersonal space. Aside from the fact that an increase in tool length enables interaction with targets further outside of the player's reach, an increase in tangential velocity achieved for distal parts of longer tools may be relevant in the context of certain game mechanics, affecting the perceived utility and ease of use of the tool, and potentially impacting the challenge, as well as the outcome of the game. Furthermore, impaired accessibility of distant targets that comes as a result of using shorter tools may impose greater mobility/physical activity requirements on the player, which may also affect their overall experience of the game. With projectile-based interaction methods, an increase in projectile size facilitates hitting the target. Furthermore, a higher initial velocity of a projectile will result in a longer point-blank range, meaning that the player will have to elevate their weapon to compensate for the parabolic trajectory of the projectile only in cases of long distance targets. In the physical world, one may observe differences in trajectory and the eventual damage inflicted on the target in case of fired projectiles of different shapes and weights, an effect that may be simulated in a game scenario.

The use of aiming assistance may significantly improve the precision of the player, potentially contributing to the increase of their perceived level of competence. The task of selecting a remote target object in a VR game may be performed through linear of parabolic pointing, with research indicating that the use of a linear ray-cast pointer may lead to a better performance [220]. Furthermore, although the inherent precision of VR controllers does not necessitate their inclusion, game developers dealing with projectile-based interaction mechanics may still choose to incorporate some of the aiming assistance mechanisms, such as those that have already been evaluated in 3D games research [221]. For example, *bullet magnetism*, a technique that subtly adjusts the trajectory of a projectile toward the target object, has already been considered as a method of improving player performance of projectile-based VR interaction mechanics (both *throw* and *shoot*; [222]). While the aforementioned methods rely on the modification of the interaction in a way that can only be achieved within an artificial virtual environment (hypernaturalism), projectile-based interactions may also benefit from aiming assistance methods that are commonly used with existing weapons in the physical world, e.g., sighting devices such as telescopic or laser sights (literalism).

In addition to aiming aids, there are additional visual traits of weapon design that may influence user experience and performance. For example, a visual gunfire effect (a so-called muzzle flash) may appear as a reaction to a trigger press, along with a sound effect, which signals to the player that a bullet has been fired. However, because a bullet travels at a very high speed, and is thus difficult to follow, developers may choose to omit its in-game visual representation, only signaling its subsequent effect on the in-game entity that has been shot (e.g., an enemy crumbles to the floor, a wooden crate smashes to pieces). In this thesis such projectiles are described as having an invisible ballistic trajectory. In other cases, the player is able to keep an eye on the projectile from the moment it exits the barrel to the moment it reaches its destination, which allows them to form the impression of its velocity and its susceptibility to in-game physical forces, and grow more aware of their own shooting precision. Note, the trajectory of a traveling projectile can be made more obvious by incorporating an accompanying projectile trail effect (e.g., a vapor trail) that serves as its visual extension.

6.2.4 Player-related factors

It is important to note that preferences regarding interaction mechanics may vary between subjects as people differ in terms of dimensions (e.g., height, weight, arm length), visual acuity and color perception, disability, fitness and energy level, strength, motor skills, reaction speed, level of experience with virtual reality and familiarity with different activities that may affect their performance for a particular game (e.g., the skills of a real-life tennis player are likely to be useful in a VR sports game). Moreover, as hand-held controllers only track the movement of terminal segments of the kinetic chain, other segments are free to move in whichever way the player decides, which means that a vast array of movements and body positions will be afforded by the game as long as the movement of the terminal segment is successful at performing the task at hand. In other words, a game does not care if the player chooses to perform an underhand or an overhand throw, or whether they choose to keep their arms outstretched or bend their elbows as they shoot at the enemy. Likewise, the player may choose to kneel, crouch or bend at the waist in order to reach a low target or avoid an incoming projectile. These alternative approaches to performing a task may lead to different levels of discomfort or success.

6.3 Designing a platform for the evaluation of interaction mechanics quality

The taxonomy of interaction mechanics presented in Chapter 4, combined with the existing knowledge of mechanics-related factors that may influence user experience, can serve as the basis for further development of a methodology for the evaluation of interaction mechanics quality. This section focuses on presenting a framework for developing platforms to be used as tools for user research, providing a way for UX/QoE researchers to easily modify different aspects of interaction mechanics, as well as capture and export multiple performance metrics. While researchers in the field often implement their own platforms for user research purposes, there is a deficit of standardized VR test tasks and VR applications [77] as tools to be used by different research groups in order to enable direct comparisons between study results. Thus, providing a formalization of requirements for such applications could be a step in this direction. Further focusing on the gaming use-case, a summary of the most important parameters and measures that could be included in these types of applications is presented below, with a more concrete implementation of our concept described in Section 6.4.1.

6.3.1 Design principles and user requirements — the INTERACT framework

Poupyrev et al. [219] explain that a platform for the evaluation of VR manipulation techniques should include the implementation of test tasks and the necessary visual stimuli. The given platform should be configurable — i.e., it should allow the experimenter(s) to configure relevant task settings and subsequently automatically adjust the properties of the virtual environment according to the given configuration. Furthermore, the platform is supposed to aid in identifying and mitigating the effects of nuisance and confounding factors that may interfere with the results of the experiment.

Following and expanding upon the principles described by the authors, this section presents a novel framework for the implementation of a platform for the evaluation of VR interaction mechanics, presented under the acronym INTERACT (Fig. 6.2). A more detailed description regarding each guideline in the INTERACT framework is presented below.

- **provide independent interfaces:** the configuration interface should be kept separate from the VR platform used in the study, e.g., the interface could be displayed on a desktop monitor, with a mouse and a keyboard as input devices for the study administrator, while the participant wearing a head-mounted display would only be allowed to observe and interact with the configured virtual environment;
- **neutralize nuisance/confounding variables:** in addition to the customizability of con-

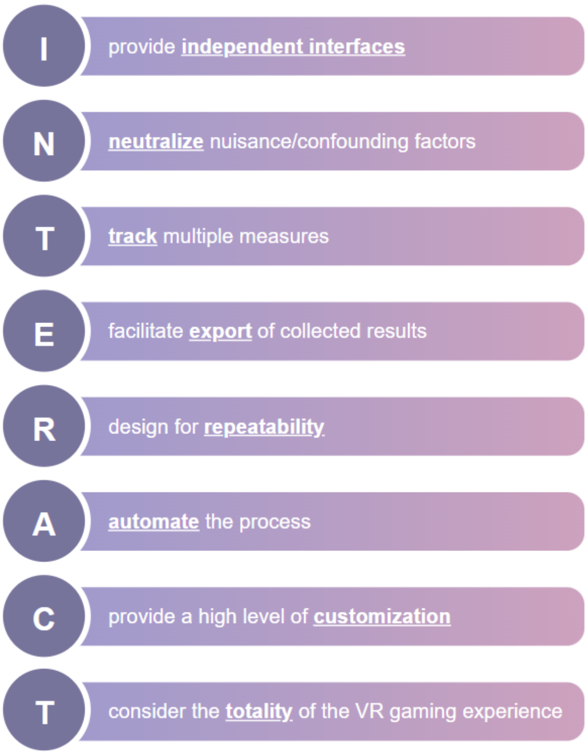


Figure 6.2: The INTERACT framework: designing platforms for the evaluation of the quality of VR interaction mechanics (taken from [10])

figuration settings, which inherently allows the experimenter to block particular nuisance variables by choosing to keep certain parameters constant across multiple scenarios, it is advisable to also include a certain degree of randomization in the implementation of the test scenario as a way of preserving internal validity, as long as it respects the constraints of predefined configuration settings;

- **track multiple measures:** the platform should be able to track and record multiple different measures for each of the individual sub-tasks in a single test scenario to provide a context for further analysis (e.g., by solely obtaining the overall accuracy score of a study participant, a researcher may overlook the implications of speed-accuracy tradeoff);
- **facilitate export of collected results:** collected metrics should be exported to a database or saved in a compact text file format (e.g., CSV, JSON, XML) and stored in a way that allows multiple records to be grouped based on the configuration/scenario, as well as based on the participant; to facilitate further analysis of the overall performance for a study scenario, the platform should provide the experimenter with a report that summarizes the results obtained for each of the sub-tasks;
- **design for repeatability:** the platform should provide the option of saving the chosen settings in a configuration file, so that they can be easily accessed (loaded) for the repeated evaluation of different participants;
- **automate the process:** any relevant spatial, temporal, or other interaction-related aspects

of the test virtual environment should automatically adjust based on the parsed configuration file;

- **provide a high level of customization:** the virtual environment should be implemented with flexibility in mind, supporting the entire spectrum of possible choices and combinations provided by the configuration interface;
- **consider the totality of the VR gaming experience:** designing a configurable virtual environment and choosing the right configuration of parameters for the evaluation of VR gaming experience requires a thorough consideration of various aspects and how they relate to each other; the goal is to select a reasonable number of scenarios, each of which presents a task that is within the limits of what is feasible and comfortable for the intended study population; efforts invested in platform development should thus be targeted toward building a virtual environment that provides a fully functional and well-balanced integration of implemented elements, such as various sensory stimuli, player controls, and the design of the task itself, while taking into account state-of-the-art industry standards, as well as perceptual, cognitive, and motor abilities of its intended users.

6.3.2 Examples of configurable parameters

Based on different interaction-related factors that may influence the gaming experience, a number of parameters that may need further examination is propose, distinguishing between target-related, task-related, and tool-related parameters.

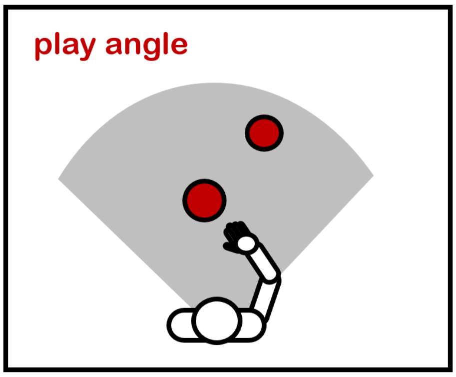


Figure 6.3: The play angle (taken from [10])

Firstly, it is important to highlight the importance of the play angle, which defines the area within which the player is supposed to interact with the targets (Fig. 6.3). Further, it is important to note the number of available targets at any given time, their sizes, as well as their distance and/or angle with respect to the player. It is worth noting that the angle in this context refers to both the altitude/elevation angle and the azimuth.

Targets are hardly ever the only elements in the environment. Their salience with respect to their surroundings is what makes them easier to spot, but what makes this more difficult is the presence of non-targets, or distractors. It is important to note the number of distractors, whether they differ from each other, and how much they differ from the target. Having a mix of targets and non-targets in the same environment may lead to visual clutter, but their number and their dispersion affects how difficult it is to find and access the target, especially if it is occluded by other objects. Fig. 6.4 summarizes the aforementioned target-related parameters.

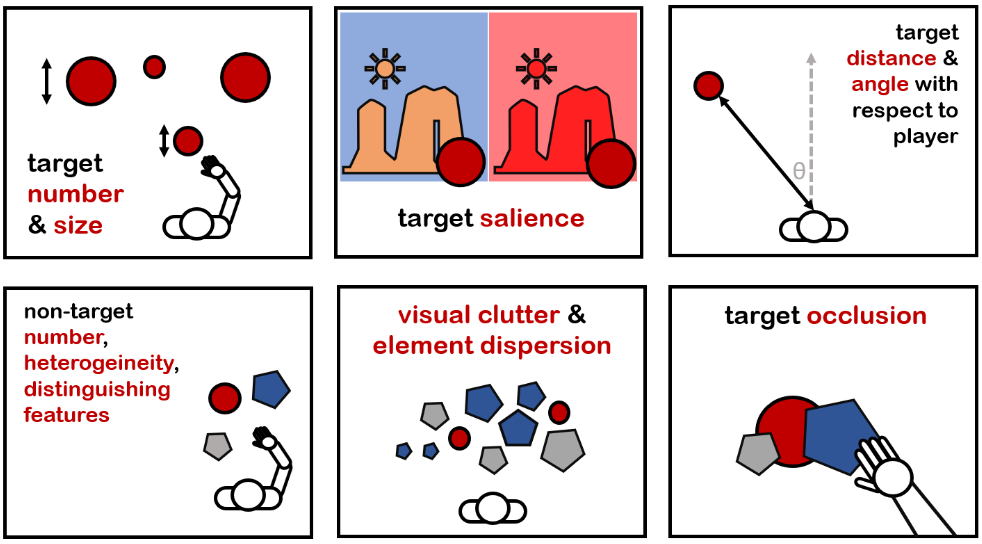


Figure 6.4: Examples of target-related parameters (taken from [10])

Depending on the game, a target may be completely passive and stationary (i.e., a stationary object), or it may be programmed with a specific behavior (e.g., enemies in a shooter game). While it is not feasible to attempt to list all possible properties of more complex behaviors in the scope of this paper, we focus on the most generic temporally variable features of target behavior. A target may move which may impact user experience to varying degrees, based on the speed and direction of movement, as well as whether it occurs along a single axis (e.g., a target only moves left to right) or multiple axes (e.g., a target may fly up and down, shift left to right, run toward or away from the player). Furthermore, the game may include target-related events — a target may spawn, or disappear with time, it may change the trajectory of its movement, etc. The frequency with which this happens may play a considerable role in player performance. These temporally variable target-related features are depicted in Fig. 6.5.

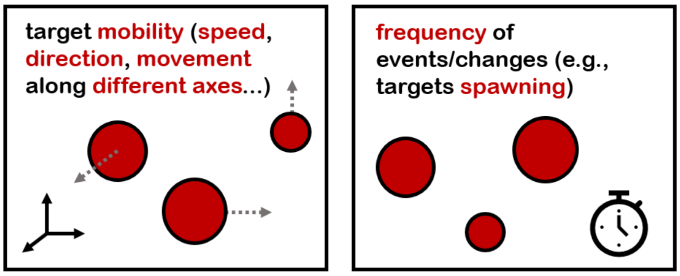


Figure 6.5: Temporally variable target-related parameters (taken from [10])

Because of the seemingly endless number of possible tasks across various games, this thesis does not aim to propose very specific task-related parameters. However, with *pick-and-place* as a basic interaction mechanic, the following parameters (Fig. 6.6) can be highlighted: position and orientation of initial object placement, the goal position and orientation, and required precision of grabbing and positioning (i.e., the accuracy with which an object has to be picked up or placed in order for it to be accepted by the game as a successful attempt). To prevent the user from strenuous movement, many games include the *"remote grab"* option, providing the player with the opportunity to access distant objects, which extends the interaction space from peripersonal to extrapersonal, decreases the challenge of obtaining distant objects, while potentially increasing player comfort.

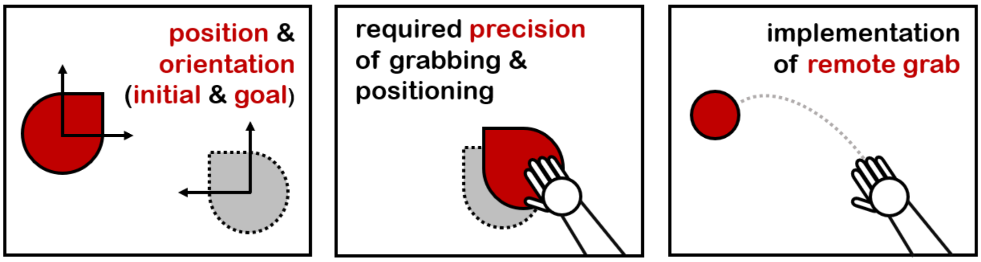


Figure 6.6: Examples of task-related parameters relevant for the implementation of *pick-and-place* mechanics (taken from [10])

For tool-mediated mechanics there are multiple tool-related parameters to consider — physical properties of handheld tools are more detrimental for mechanics that utilize tools extensions of peripersonal space, while projectile-based interaction mechanics are more affected by physical properties of projectiles. Another factor that is likely to affect the experience of using a handheld tool as an extension of peripersonal space is the use of a flexible tool, or added joints, as they alter the kinetic chain of the player's body. With ranged tools and weapons, the interaction with a target is going to be affected by the physics of projectile propulsion, as well as any aiming aids or visual effects that impact the overall precision. Tool-related parameters are listed in Fig. 6.7.

6.3.3 Examples of subjective and objective measures

Playing a VR game encompasses a number of elements that contribute toward the overall user experience, from subjective constructs such as fun and immersion, to objective aspects such as in-game performance. Furthermore, the intrusiveness and physicality of the VR platform requires researchers and developers to consider the issues of workload and physical discomfort. Because of the multidimensionality of the VR gaming experience, in this subsection, the aim is not to cover all possible measures that may be incorporated into a VR interaction mechanics evaluation platform, but rather provide only examples.

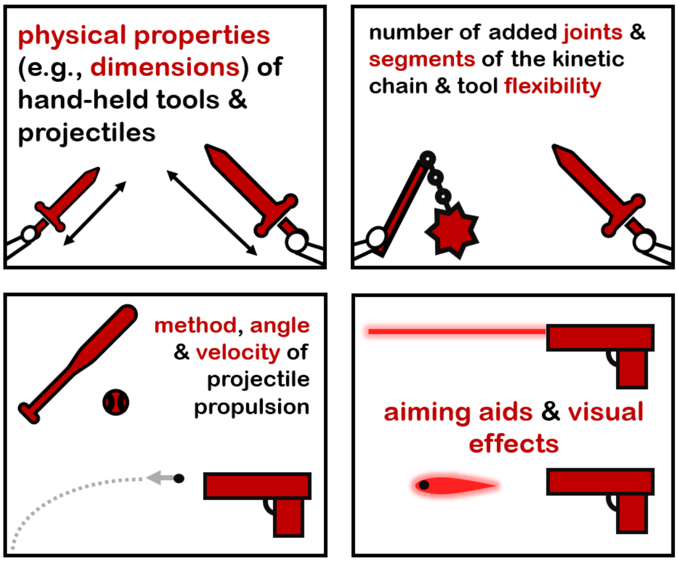


Figure 6.7: Examples of tool-related parameters (taken from [10])

In terms of subjective measures, the method of collecting data that is arguably most suited for use with evaluation platforms is the use of single- or multiple-item questionnaires. The VR Neuroscience Questionnaire (VRNQ [106]) includes items related to the overall user experience, game mechanics, in-game assistance and VR induced symptoms and effects, but would have to be adapted to suit game mechanics other than *pick-and-place*. Even though they may not be VR-specific, questionnaires developed specifically for the evaluation of gaming experience (e.g., the Game Experience Questionnaire – GEQ [103]) may also be used as they often include questions related to dimensions such as competence and challenge, both of which are likely to be affected by changes in the implementation of interaction mechanics, while some may even include specific control-related questions (e.g., the Player Experience Inventory – PXI [105]). Depending on the mechanics, the user may be required to undertake different levels of workload, which could be assessed using the Simulation Task Load Index, abbreviated as SIM-TLX [160], which was designed for use with VR.

Previously discussed cybersickness questionnaires, such as the SSQ [114], are often used in VR user studies as a way to gauge the level of discomfort in users, but they may be more suitable for use with platforms that include in-game navigation. However, even in an otherwise static virtual environment, particular implementations of game mechanics could potentially trigger certain symptoms which are often related to cybersickness, such as eye strain, headache, and disorientation. Because of the manual interaction mechanics' reliance on repetitive gross motor movements, it is advisable to include questions related to muscle pain, strain, or exertion, which may be assessed using Borg rating scales [199]. Questionnaires like those listed above could be used independent of the evaluation platform (e.g., in paper form), but could also be incorporated into the platform itself. Examples of assets that enable such integration of questionnaires into VR applications are VRate [118] and VRQuestionnaireToolkit [223].

In terms of objective measures, as listed in existing sources focused on the evaluation of VR interaction quality, such as [42, 219, 224], typical examples include the duration needed to accomplish a task, the overall accuracy, and error rate. However, depending on the explored mechanics, these metrics may be collected in different ways. For example, the overall error rate may be calculated based on a number of missed targets in a shooting scenario, or a number of objects that were left intact because the player's swing of an edged weapon was too weak to destroy them. In a *pick-and-place*-based puzzle game, what we consider an error may involve an object placed in the wrong spot, or placed too far from the center of its intended terminal position. Likewise, duration needed to accomplish a task and the overall accuracy can be calculated in various ways, depending on the particular mechanics and the context of the task. Performance measures may be supplemented by physiological measures such as tracking eye movement or heart rate, as well as subjective measures of discomfort, perceived workload, perceived challenge, etc.

6.4 Methodology

Even though the scope of possible configurations of VR IMs is too broad to be covered in this thesis, this section presents the methodology of three studies investigating the impact of chosen parameter values on different QoE features of the singleplayer experience through the use of a custom test application.

6.4.1 Materials

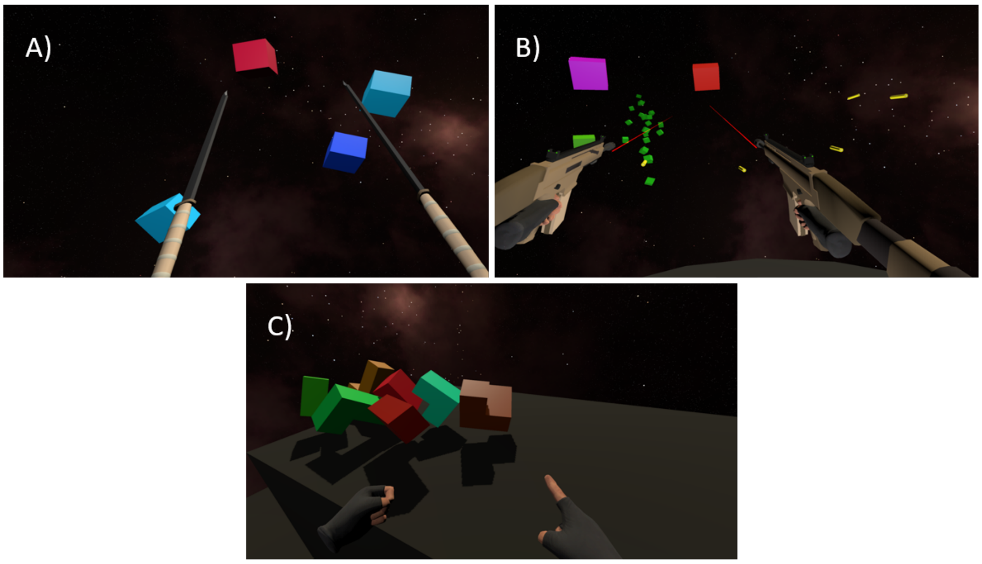


Figure 6.8: Screenshots from the test application depicting different VR interaction mechanics: A) slash; B) shoot; C) pick and place (taken from [10])

As the main research objective of Studies 3-5 was to explore the effects of specific IM implementation configurations, using commercially available games as test material would not provide the required level of customization, nor would it enable the collection of detailed task performance metrics. Therefore, a special platform was designed to serve as a tool for the exploration of user experience with three different types of bimanual VR interaction mechanics: *pick and place* as a non-mediated type of mechanics, *slash* as an example of mechanics involving tools as extensions of peripersonal space, and *shoot* as an example of projectile-based interaction mechanics. The application was designed specifically for the purpose of this research in line with the INTERACT framework, and implemented with the help of students Filip Nemec and Monika Matokanović [225] using the Unity game engine with the Virtual Reality Interaction Framework (VRTK) package. The application provides a desktop user interface allowing the administrator to define the parameter space for multiple mechanics-specific parameters, which could then be saved for future use. These parameters are used to generate the virtual environment which is experienced through a head-mounted display, with its dynamic behavior randomized within administrator-defined boundaries. Performance measures collected during gameplay are stored in a local folder and organized based on identification provided by the administrator. Avoiding all superfluous detail that could possibly divert the users' focus away from the implemented interaction mechanics, the application's environmental design is constrained to a limited set of low-poly objects set against a celestial nocturnal backdrop. Due to its minimalist aesthetic, all three customizable scenes (one for each implemented IM type) provide a similar visual impression.

The customizable scene for the evaluation of slash mechanics, inspired by Fruit Ninja VR, places the user on a platform surrounded by a designated number of cannons positioned either radially all around the user or along a smaller circular arc, as determined by the chosen configuration. Depending on defined parameter values, each cannon will sporadically expel a cuboid box — either straight up into the air, or at an angle — which serves as a target object to be destroyed by the user. Object destruction is performed by slashing the object using an elongated handheld weapon. If the user fails to make contact — or fails to use the adequate force whilst making contact — the object will eventually fall down, influenced by gravity, and shatter upon impact with the ground.

The customizable scene for the evaluation of pick-and-place IMs was designed as a three-dimensional puzzle game in the style of a solid dissection puzzle known as the Soma cube. Upon entering the scene, the user is placed next to a desk with scattered puzzle pieces. Each puzzle piece is a polycube, which consists of individual cubic units joined together in a randomized spatial arrangement, reminiscent of tetrominoes in a Tetris game if they were converted into three dimensions. When properly assembled, all provided puzzle pieces form a singular large cube. Users are also provided with a solution cube, a large cube the size of a completed puzzle, divided into smaller cubic slots. Upon picking up a puzzle piece, a set of cubic slots corresponding to its shape will light up, indicating its solution space, i.e., the designated placement of that puzzle piece within the solution cube. Clearly highlighting the correct solution space of each puzzle piece ensures that the cognitive effort necessary to solve the puzzle is eliminated, so that participants can be fully focused on the straightforward task of positioning and orienting virtual objects. Users are then expected to slide the piece into place, fitting it inside the solution cube, and repeat this step with all other puzzle pieces until the finalized cube is assembled inside of the solution cube. As the user makes progress on the puzzle, the task of fitting complex polycubic shapes into place increases in complexity, as their efforts are hindered by existing puzzle pieces blocking the placement. Depending on the selected configuration, the task can be made easier by providing a small buffer space between the assembled puzzle pieces. An illustration of described puzzle components is depicted in Figure 6.9.

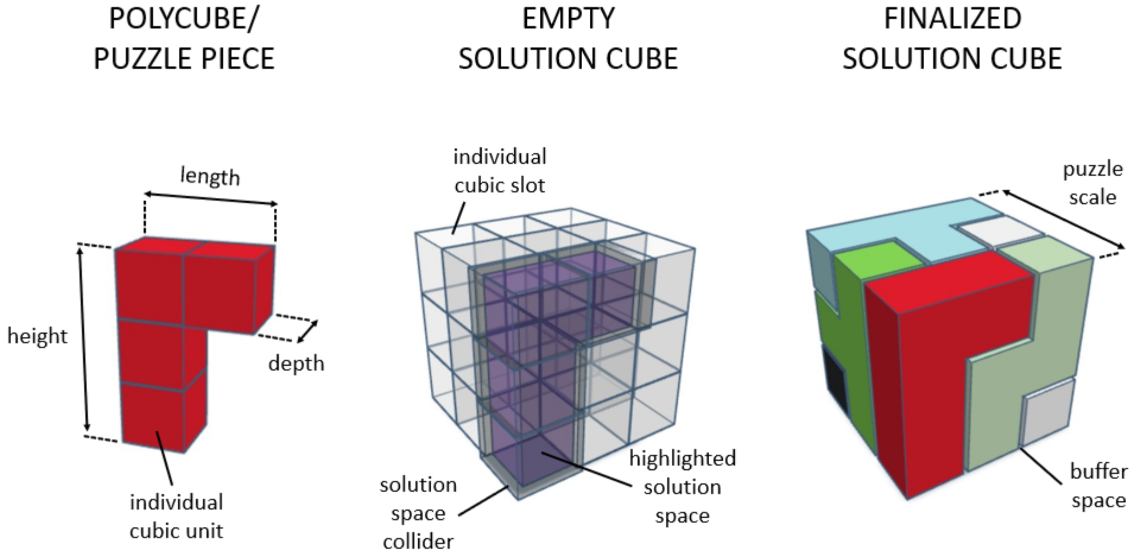


Figure 6.9: An illustration of puzzle components in the customizable scene for the evaluation of pick-and-place IMs.

Inspired by games such as Serious Sam VR: The Last Hope and Space Pirate Trainer, the customizable scene for the evaluation of shoot mechanics places the user on a platform with a view of floating (stationary) or flying (moving) targets. Visually contrasting the night sky, these targets are sporadically spawned at random positions within the boundaries of a volume defined by a predetermined spawn angle and an acceptable range of altitudes and distances with respect to the player. The targets are destroyed immediately after being hit with a projectile expelled from the user's handheld weapons. In addition to customizing target specifications and behaviour, the application provides a significant level of customization pertaining to ballistic properties of selected weapons, as well as enabling the use of optional visual aiming aids.

Examples of parameters and measures included in the prototype platform for the evaluation of interaction metrics quality are presented in Table 6.1, and screenshots of each mechanics

Table 6.1: Examples of configurable parameters and collected measures included in the custom platform for the evaluation of VR interaction mechanics quality (adapted from [10])

	slash	**pick-and-place**	**shoot**
target-related parameters	range of target object positions, target number, target size, target movement, target event frequency	range of target object positions, target number, target size	range of target object positions, target number, target size, target movement, target event frequency
task- and tool-related parameters	tool dimensions, minimum force necessary to destroy target	precision of grabbing, precision of positioning, remote grab	projectile expulsion, aiming aids, visual effects
objective measures	successful and unsuccessful attempts at slashing, average slashing force, target lifetime	initial and terminal position and orientation, number of times object was grabbed, time to correct placement, overall success rate	total shots fired, total hits and misses, target lifetime

implementation are presented in Fig. 6.8. All three studies utilized the HTC VIVE Pro Eye headset paired with Valve Index controllers.

6.4.2 Procedure

This subsection presents the combined description of methodological choices and study procedure performed for three distinct studies — referred to as Study 3, Study 4, and Study 5 throughout this thesis — as they share many similarities with respect to methodology. All studies were conducted in the same laboratory setting under the supervision of a study administrator, using the same test material, with each study focusing on one type of IM.

The study procedure for all three studies was as follows. Upon entering the laboratory premises, each participant was asked to provide informed consent for the participation in the study. Participants were warned about the possibility of VRISE and encouraged to pause or terminate the experiment if they started to feel uncomfortable. Participants filled out an online pre-study questionnaire, providing their demographic information (age, sex, dominant hand) and reporting their level of experience with VR, as well as their sentiment toward VR technology. Participants were then given instructions describing how to adjust and use VR hardware and entered VR to partake in a short tutorial session to grasp the controls and the task at hand. The tutorial scenario was set to default values of the application, some of which are presented in Table 6.2, while a more detailed overview of default values for all parameters is given in [225].

Following the tutorial session, participants were asked to complete either three or four groups of scenarios (depending on the study/IM), each consisting of two to four separate scenarios. All scenarios within a particular group were identical in terms of configuration (which was set to default values), except for a single parameter value that was manipulated for that group. The information regarding all scenarios is presented in Table 6.2. Scenario groups were experienced in the order presented in the table, while scenarios within each scenario group were experienced in a different, randomized order for each participant. Conceived as a series of short interactive tests [62], each individual scenario was set to last 90 seconds. The only exception

Table 6.2: An overview of manipulated parameters and individual scenarios tested in Study 3, Study 4, and Study 5; default values of each manipulated parameter are typed in bold letters and denoted by an asterisk in the *Scenario label* column

Study ID (**IM**)	**Order of presentation**	**Manipulated parameter(s)**	**Parameter description**	**Scenario label**	**Parameter value**
Study 3 (Slash)	1	Target spawn angle	A horizontal angle determining the boundaries of the space in which the targets will be spawned, determined in relation to participants' initial position and rotation (e.g., a 90-degree spawn angle encompasses the region spanning 45 degrees to the left and 45 degrees to the right of the player).	**90_DEG***	*Spawn angle* is set to 90 degrees.
				180_DEG	*Spawn angle* is set to 180 degrees.
				360_DEG	*Spawn angle* is set to 360 degrees.
	2	Force to destroy	The minimum force that needs to be exerted on the target object to successfully destroy it, expressed in Newtons.	0N	*Minimum force* is set to 0 newtons.
				2N*	*Minimum force* is set to 2 newtons.
				6N	*Minimum force* is set to 6 newtons.
	3	Weapon length	The length of a handheld weapon (in case of Study 2, a katana) used to slash target objects. The weapon held in the right hand and the weapon held in the left hand can be adjusted separately. For the purpose of Study 2, both weapons were set to identical dimensions.	**1U***	*Weapon length* is set to 1 Unity unit.
				0_7U	*Weapon length* is set to 0.7 Unity units.
				1_3U	*Weapon length* is set to 1.3 Unity units.
Study 4 (Pick-and-place)	1	Puzzle scale	Edge length of a finalized puzzle cube, expressed in Unity units. Individual puzzle pieces are scaled to fit within the finalized cube, depending on their number (in case of Study 2, the number of puzzle pieces to be placed was 7).	0_1U	*Puzzle scale* is set to 0.1 Unity units.
				0_4U*	*Puzzle scale* is set to 0.4 Unity units.
				0_7U	*Puzzle scale* is set to 0.7 Unity units.
	2	Collider scale	The cumulative scale of the colliders in the solution space, expressed in Unity units. Entering the solution space is a requirement for successful placement of puzzle pieces. Individual colliders within the solution space are scaled to size.	0_2U	*Collider scale* is set to 0.2 Unity units.
				0_5U*	*Collider scale* is set to 0.5 Unity units.
				1U	*Collider scale* is set to 1 Unity unit.
	3	Remote grab	While puzzle pieces can otherwise only be accessed if they are picked up with the virtual hand, enabling the remote grab option allows participants to pick up puzzle pieces from a distance using a pointer.	NO_GRAB	*Remote grab* is set to TRUE.
				REM_GRAB*	*Remote grab* is set to TRUE.
	4	Scale offset	This parameter determines the final scale of a puzzle piece when allowing for buffer space between puzzle pieces in a solution cube. The maximum value of this parameter, which is set to 1, completely removes the buffer space between puzzle pieces.	0_8S	*Scale offset* is set to 0.8.
				0_9S	*Scale offset* is set to 0.9.
				1S	*Scale offset* is set to 1.
Study 5 (Shoot)	1	Visual aiming aids	Turning on the laser equips the weapon with a laser sight. Turning on the visibility of the bullet trajectory creates a visual effect that trails behind the projectile (akin to a comet tail), increasing its visibility.	LAS_TRAJ	*Use laser* is set to TRUE; *Show bullet trajectory* is set to TRUE.
				LAS	*Use laser* is set to TRUE; *Show bullet trajectory* is set to FALSE.
				TRAJ*	*Use laser* is set to FALSE; *Show bullet trajectory* is set to TRUE.
	2	Target spawn angle	A horizontal angle determining the boundaries of the space in which the targets will be spawned, determined in relation to participants' initial position and rotation (e.g., a 90-degree spawn angle encompasses the region spanning 45 degrees to the left and 45 degrees to the right of the player).	**90_DEG***	*Spawn angle* is set to 90 degrees.
				180_DEG	*Spawn angle* is set to 180 degrees.
				360_DEG	*Spawn angle* is set to 360 degrees.
	3	Shoot force	The force exerted on the projectile, expelling it from the muzzle of the weapon. As it dictates the initial velocity of the projectile, this parameter also determines its range and the shape of its trajectory.	20N	*Shoot force* is set to 20 newtons.
				40N*	*Shoot force* is set to 40 newtons.
				80N	*Shoot force* is set to 80 newtons.

were the scenarios in Study 4 (which focused on pick-and-place mechanics), as they would be terminated earlier in case the puzzle was completed prior to the expiration of the allotted time. Participants were not able to track the time whilst being immersed in a scenario.

Following each scenario, participants were asked to complete the post-scenario questionnaire which was presented on the headset display and filled in with the use of controllers to avoid the laborious task of adjusting the VR equipment between multiple scenarios. The post-scenario questions included several items pertaining to chosen features of core QoE constituents:

- **player experience:** challenge, competence, and fun (items inspired by the GEQ core module [19, 20]);
- **workload:** physical demand, mental demand, task control difficulty (items adapted from the SIM-TLX questionnaire [160]);
- **VRISE:** pain and muscle fatigue, overall sense of physical discomfort.

Even though multiple items were adapted from existing questionnaires, to further simplify the cumbersome process of repeatedly completing the post-scenario questionnaire during the course of a study session, the same 5-point scale (ranging from *"1 - not at all"* to *"5 - very much"*) was used for collecting the evaluations of each feature's intensity. Participants were also asked to rate the overall QoE on a scale from *"1 - bad"* to *"5 - excellent"*, and to report whether they were willing to continue playing in the presented conditions (*yes/no*).

Table 6.3: Task performance measures used in Study 3, Study 4, and Study 5

Study ID (IM)	Task performance measure	Task performance measure description
Study 3 (Slash)	Average force	Average force used to slash target objects, expressed in newtons.
	Accuracy	Percentage of target objects that were successfully destroyed.
Study 4 (Pick-and-place)	Percentage of successful participants	Percentage of participants who successfully completed the puzzle prior to the expiration of allotted time.
	Average percentage of puzzle pieces placed	Percentage of puzzle pieces that were successfully placed by the expiration of allotted time.
	Average duration	Average duration calculated based on data collected from all participants.
	Average duration of successful completion	Average duration calculated based on data collected from successful participants only.
Study 5 (Shoot)	Number of shots fired	Overall number of expelled projectiles.
	Accuracy	Percentage of shots that successfully destroyed the target object.

Objective task performance measures were also collected; however, due to specific characteristics of each implemented IM, they differed between studies. Measures included in this thesis are listed and described in Table 6.3. After completing each scenario group, participants were given a post-group questionnaire, in which they were asked to pick what they considered to be the best and the worst scenario from that group.

6.4.3 Participants

As clarified in Section 1.3, the participant sample in Studies 3-5 was chosen to be fairly homogenous, comprising of young adults without mobility or other limitations that are likely to impact their experience, with a fairly balanced sex distribution. All studies used the same sample size of 30 participants, with the following demographic characteristics:

- **Study 3:** 13 female and 17 male participants, aged 18-33 (M = 23.43, SD = 3.31);
- **Study 4:** 16 female and 14 male participants, aged 19-31 (M = 23.07, SD = 2.60);
- **Study 5:** 13 female and 17 male participants, aged 18-34 (M = 22.13, SD = 3.21).

Due to limited access to regular VR users, participants were generally inexperienced with VR, as seen in Table 6.4. Despite their lack of experience, participants in all three studies expressed a generally positive attitude toward VR technology, with this item receiving the average score of 4.20 (SD = 0.66) from participants in Study 3, 4.10 (SD = 0.66) from participants in Study 4, and 4.27 (SD = 0.74) from participants in Study 5.

Table 6.4: Reported frequency of VR usage for participants in Studies 3-5, expressed as the number of participants picking each option

	Study 3	**Study 4**	**Study 5**
I have never used VR technology prior to this study.	8	6	9
I have only ever tried using VR technology on a few occasions, about 1 to 3 times.	17	16	16
I occasionally use VR technology, but no more than once a month.	2	7	3
I use VR technology at least monthly.	3	1	2

6.5 Slash interaction mechanics

This section summarizes the results of slash IM tests conducted in Study 3, with a more detailed presentation of descriptive statistics and performed statistical tests available in the Appendix. Participants' preferences for the best and worst scenario in each scenario group are listed in Figure 6.10.

With regard to target spawn angle, participants clearly preferred the 90_DEG scenario, which also scored the highest in terms of QoE (M = 4.4, SD = 0.62) compared to 180_DEG (M = 3.97, SD = 0.96) and 360_DEG (M = 3.27. SD = 1.26). On average, the highest scoring scenario was also found to be the least challenging, easiest to control, and least mentally and physically demanding. Furthermore, it was found to be more fun and made the participants feel more competent compared to the other two scenarios. While neither scenario seemed to cause significant VRISE, 90_DEG still scored the lowest compared to the other two. On the opposite end of the scale, the 360_DEG scenario consistently scored the worst for all tested items. While 97% of participants were willing to continue the 90_DEG scenario, and 90% of

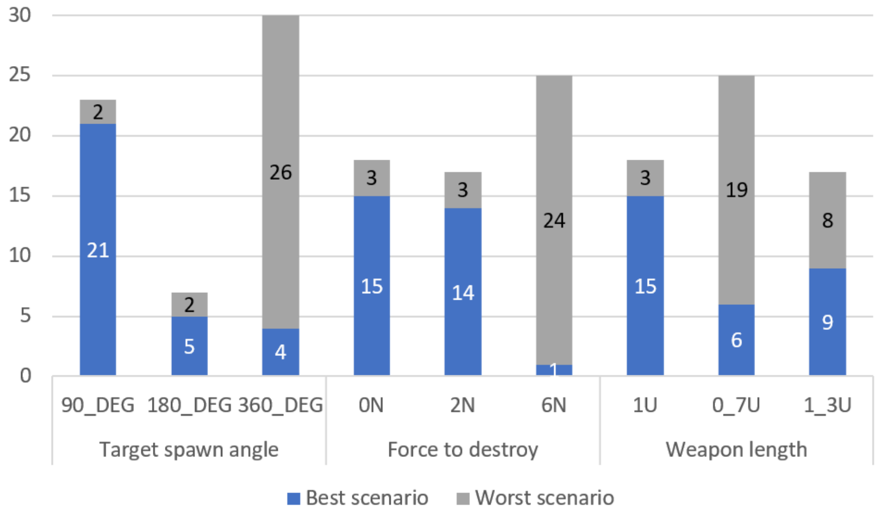


Figure 6.10: Participant preferences for each slash IM scenario group

participants were willing to continue the 180_DEG scenario, this percentage dropped to only 47% for 360_DEG. Accuracy also suffered with an increase in target spawn angle, dropping from the average of 68.50% and 56.42% for 90_DEG and 180_DEG respectively, to 33.05% for 360_DEG. Differences in accuracy between all three scenarios were found to be statistically significant.

In terms of force required to destroy a target object, 0N and 2N scenarios received relatively similar scores, both in terms of QoE — with an average score of 4.2 (SD = 0.85) for 0N and 4.3 (SD = 0.75) for 2N — and participant preferences. Even though the 0N scenario was found less physically demanding, as well as significantly less challenging and easier to control, the 2N scenario was considered more fun, which may have influenced a higher number of participants (93%) to report their willingness to continue this scenario, as opposed to 0N (80%). For the 6N scenario, which received the lowest QoE score in this scenario group (M = 3.17, SD = 1.11), participants reported experiencing significantly less fun and feeling significantly less competent compared to the other two scenarios, with only 47% of participants willing to continue playing with this configuration. This scenario was also found to be significantly more challenging, physically demanding, and difficult to control, in addition to causing significantly more pain and muscle fatigue. Task performance also differed significantly between scenarios. While participants attempted to slash the target objects with a significantly higher average force (M = 5.05N, SD = 0.76N) compared to 0N (M = 2.91N, SD = 0.84N) and 2N (M = 3.53, SD = 0.76) scenarios, their performance suffered with, on average, only 34.29% of objects successfully destroyed, compared to 81.09% and 71.36% for 0N and 2N, respectively.

The least significant results were obtained for the weapon length scenario group. Even though a higher number of participants seemed to prefer the 1U scenario over 1_3U, both scenarios received the average QoE score of 4.23, with 0_7U trailing closely behind (M = 4.1, SD

Table 6.5: Friedman test results for all slash IM scenario groups (* p < 0.05, ** p < 0.005, *** p < 0.001)

QoE constituent	**QoE feature**	$\chi^2(2)$		
		Target spawn angle	**Force to destroy**	**Weapon length**
Player experience	Competence	**39.39*** **	**36.98*** **	**6.17* **
	Challenge	**33.49*** **	**42.14*** **	**8.27* **
	Fun	**17.20*** **	**19.43*** **	0.38
Workload	Mental demand	**20.72*** **	4.22	4.75
	Physical demand	**10.17* **	**41.95*** **	**7.14* **
	Task control difficulty	**25.89*** **	**28.42*** **	0.62
VRISE	Pain and muscle fatigue	**6.86* **	**19.02*** **	1.37
	Overall sense of physical discomfort	**10.00* **	1.63	2.92

= 0.76). Interestingly, a slightly higher percentage (93%) of participants was willing to continue the 1_3U scenario, compared to the other two (87% for both). Even though differences in scores between the scenarios were fairly small for this scenario group compared to other scenario groups, the shortest weapon length was reported as posing the highest challenge and presenting the most physically demanding task, also slightly affecting participants' perceived competence. The issue was somewhat reflected in the accuracy scores, as average accuracy of 0_7U (M = 67.63%, SD = 8.18%) was slightly lower compared to 1U (M = 71.31%, SD = 10.87%) and 1_3U (M = 70.85%, SD = 8.92%) scenarios.

While the presented results provide useful information regarding user preferences, they also confirm that manipulating chosen parameters of the slash IM implementation yields significant differences in observed QoE features belonging to core QoE constituents of the singleplayer experience. This was confirmed by the results of Friedman tests which were used to find statistically significant differences in observed QoE features for each scenario group, as presented in Table 6.5. A more detailed analysis including Wilcoxon signed-rank tests with Bonferroni correction is included in the Appendix.

6.6 Pick-and-place interaction mechanics

This section summarizes the results of pick-and-place IM tests conducted in Study 4, with a more detailed presentation of descriptive statistics and performed statistical tests available in the Appendix. Participants' preferences for the best and worst scenario in each scenario group are listed in Figure 6.11.

The first observed parameter was related to the scale of the three-dimensional puzzle. It is worth noting that puzzle pieces — when scaled to fit the solution cube of a size determined by this parameter — were comprised of individual cubes with an edge length of approximately 3cm (0_1U), 10cm (0_4U), and 20cm (0_7U). Each puzzle piece was comprised of multiple

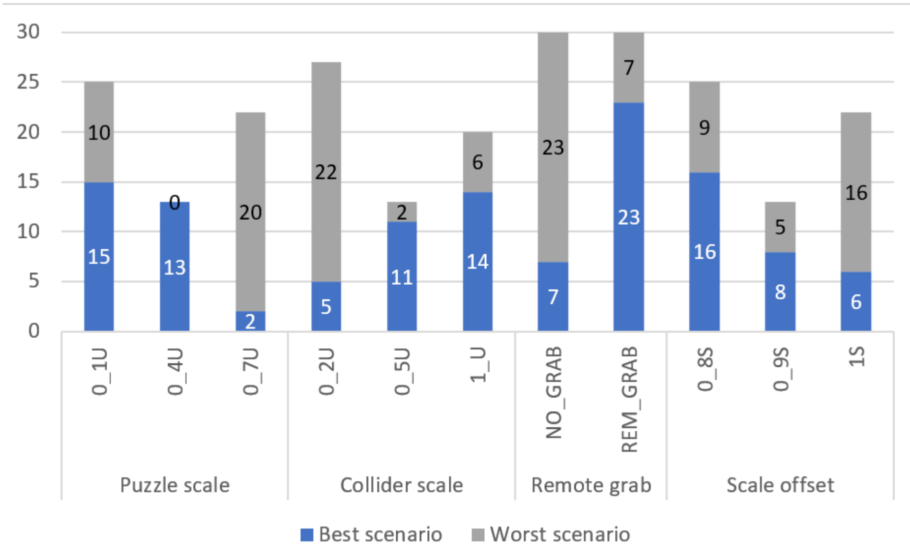


Figure 6.11: Participant preferences for each pick-and-place IM scenario group

Table 6.6: Friedman (χ^2) and Wilcoxon signed-rank test (Z) results for all pick-and-place IM scenario groups (* $p < 0.05$, ** $p < 0.005$, *** $p < 0.001$)

QoE constituent	**QoE feature**	**χ^2(2)**			**Z**
		Puzzle scale	**Collider scale**	**Scale offset**	**Remote grab**
Player experience	Competence	**13.01****	**29.85*****	4.96	-1.89
	Challenge	**10.29***	**17.34*****	**10.93****	**-3.41*****
	Fun	2.20	3.60	0.79	**-2.56***
Workload	Mental demand	1.55	**12.52****	5.64	**-2.11***
	Physical demand	4.69	4.57	4.50	0.00
	Task control difficulty	**10.83****	**31.14*****	**20.28*****	**8.07****
VRISE	Pain and muscle fatigue	2.00	2.00	4.00	0.00
	Overall sense of physical discomfort	1.00	2.00	/	0.00

individual cubes, but no more than three in each dimension (i.e., each piece was 1-3 cubes wide, 1-3 cubes high, and 1-3 cubes deep). Participants appeared to dislike the largest puzzle scale (0_7U), with 20 out of 30 participants choosing it as the worst option in the puzzle scale scenario group. Moreover, this scenario received the worst average QoE score (M = 4.00, SD = 0.83) out of the three scenarios. While slightly more participants rated the 0_1U scenario as the best scenario compared to 0_4U, it was considered the worst by a third of all participants in this study. Considering that 0_4U (M = 4.50, SD = 0.68) also received a higher QoE score compared to 0_1U (M = 4.23, SD = 0.86), the medium puzzle scenario appears to have garnered a more universally positive response. With both 0_1U and 0_7U scoring higher in terms of challenge and all measures of workload, participants felt most competent following the 0_4U scenario, while also experiencing slightly more fun. Participants' subjective experience was reflected in collected objective metrics, as the 0_4U scenario was successfully completed by 80% of participants, significantly exceeding the success rate of both 0_1U (53%) and 0_7U

(27%). However, despite fairly low success rates, the majority of participants — 87% and 93% for 0_7U and 0_1U, respectively — expressed their willingness to continue playing both of the lower rated scenarios. Only one participant was not willing to continue playing 0_4U.

On the subject of collider scale, the most conclusive results were obtained for 0_2U as the least preferred option. Chosen for the worst scenario in the group by 73% of participants, this scenario received a significantly worse mean QoE score (M = 3.60, SD = 1.28) compared to both 0_5U (M = 4.40, SD = 0.72) and 1U (M = 4.30, SD = 0.75). Due to the implications of smaller colliders on placement precision, the 0_2U scenario significantly lowered participants' perceived competence, in addition to making the task significantly more challenging, mentally demanding, and difficult to control compared to the other two. This was evidenced by objective measures, as the percentage of successful participants for this scenario (40%) was considerably smaller compared to 0_5U (87%) and 1U (93%). Moreover, the results of both temporal measures (average duration and average duration of successful completion) further indicate that participants were struggling with the placement of puzzle pieces in the smallest collider scale scenario. Remarkably, the percentage of participants willing to continue this scenario (83%) was comparable to the percentage of participants willing to continue the other two (86% for both), which is somewhat surprising considering that, in addition to its negative effects on player performance, the 0_2U scenario also received the lowest mean fun score in the group (although there were no statistically significant differences between scenarios in this case).

Even though both scenarios in the remote grab group received identical average QoE scores (4.50), a significant majority of participants preferred the REM_GRAB scenario, as seen in Figure 6.11. However, this preference was not evident from the average scores of examined QoE features, as REM_GRAB was deemed to be more challenging, mentally demanding, and more difficult to control, making participants feel less competent. Whereas all participants successfully completed the puzzle during the NO_GRAB scenario, only 70% of participants managed to do the same for the REM_GRAB scenario. Moreover, on average, completing the NO_GRAB scenario required significantly less time compared to REM_GRAB. These results may be somewhat unexpected, as the remote grab option was supposed to serve as an aid to participants. However, as evidenced by participants' comments, enabling this option made it easier to accidentally grab the wrong puzzle piece, or shift assembled pieces out of place. Still, the scenario received a significantly higher fun score compared to NO_GRAB, which is likely the main reason for it being chosen as a clear favourite as participants praised its novelty and magical qualities. Furthermore, despite the increased workload of picking up and placing the right pieces that came with the REM_GRAB scenario, multiple participants expressed their satisfaction with the ability to grab puzzle pieces from a distance, further stressing the importance of introducing accessibility improvements to VR games.

As seen in Figure 6.11, smaller scale offsets were preferred over the 1S scenario, which left no buffer space between puzzle pieces assembled in the solution cube. The preference was stronger for the 0_8S scenario, although its QoE score (M = 4.43, SD = 0.77), was only slightly higher compared to 0_9S (M = 4.33, SD = 0.71) and 1S (M = 4.27, SD = 0.87). Despite unremarkable differences between scenarios, the 1S scenario consistently stood out from the other two, as it made participants feel less competent by significantly increasing the challenge and task difficulty compared to other scenarios. This is due to significant maneuvering required to place a new puzzle piece among existing puzzle pieces, densely packed inside of the solution cube. Moreover, this scenario's scores for mental and physical demand were also slightly higher compared to the others, while its average fun score was slightly lower. The 1S scenario also had the lowest success rate (63%), compared to 0_8S (100%) and 0_9S (93%). In general, however, participants were still mostly willing to continue playing either scenario in the group, with 87% reporting their willingness to continue 1_S, and 90% for the other two scenarios.

6.7 Shoot interaction mechanics

This section summarizes the results of shoot IM tests conducted in Study 5, with a more detailed presentation of descriptive statistics and performed statistical tests available in the Appendix. Participants' preferences for the best and worst scenario in each scenario group are listed in Figure6.12.

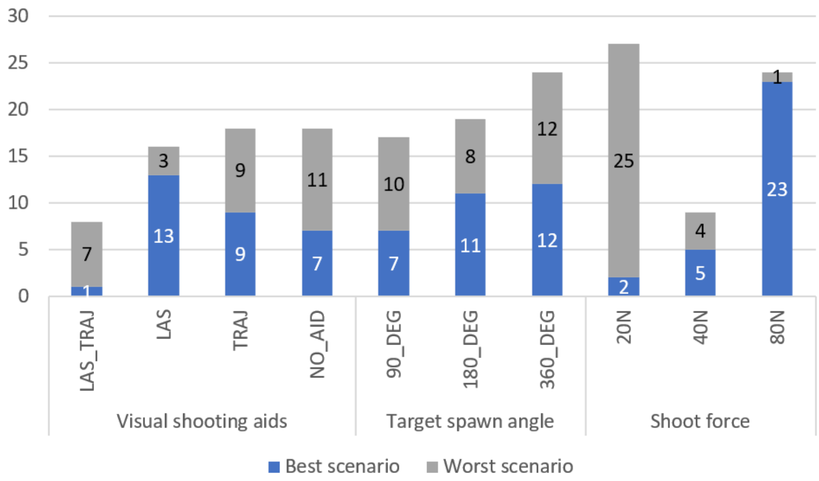


Figure 6.12: Participant preferences for each shoot IM scenario group

Exploring the use of visual shooting aids, the highest number of participants preferred the use of laser. Even though participants avoided listing LAS_TRAJ as the best scenario in the group, its QoE score (M = 4.40, SD = 0.81) was close to those of LAS (M = 4.47, SD = 0.90) and NO_AID (M = 4.43, SD = 0.77), while TRAJ received the lowest average QoE score (M = 4.17, SD = 0.91). While reported preferences and QoE ratings for this group of scenarios are mixed, individual ratings of different QoE features show significant differences between scenarios. Overall, scenarios that did not involve the use of laser (TRAJ and NO_AID) resulted in lower perceived competence, along with increases in challenge, pain and muscle fatigue, and all measured categories of workload compared to the other two scenarios. As expected, the NO_AID scenario was deemed the most demanding. Even though the LAS_TRAJ involved the use of two aiming aids in conjunction, participants found it more challenging than LAS, which only used laser. This may have been somewhat distracting to participants as using both aiming aids at once highlighted the offset between the straight-line laser beam and the slightly curved trajectory of the projectile as it succumbs to gravity. The results of subjective measures addressing competence, challenge, and workload were confirmed by task performance metrics, with the LAS scenario resulting in the accuracy of 80.21%, followed by LAS_TRAJ at 71.72%, TRAJ at 61.42% and NO_AID at 51.43%. However, even though it produced the worst performance of the scenario group, the NO_AID scenario (M = 4.43, SD = 0.86) was comparable with LAS (M = 4.40, SD = 0.97) and slightly higher from TRAJ (M = 4.33, SD = 0.88) in terms of fun, with LAS_TRAJ obtaining the lowest score in the group (M = 4.07, SD = 1.17).

As with visual shooting aids, results for target spawn angle were also mixed. Although participants were more likely to choose a wider spawn angle for their favourite scenario, the QoE rating of the 90_DEG scenario (M = 4.43, SD = 0.82) was higher than the 360_DEG scenario (M = 4.20, SD = 0.96), although both were exceeded by 180_DEG (M = 4.7, SD = 0.6). Likewise, participants were slightly more inclined to go on with playing the 90_DEG scenario, with 90% of participants stating their willingness to continue compared to 87% for 180_DEG and 77% for 360_DEG. This is likely due to the increased challenge, mental and physical workload, and task control difficulty of wider target spawn angles. Furthermore, increases in target spawn angle also increased the overall sense of physical discomfort, although there were no statistically significant differences for this feature. However, when it comes to fun, participants seemed to enjoy the 180_DEG scenario (M = 4.53, SD = 0.73) more than 90_DEG (M = 4.17, SD = 0.95) and 360_DEG (M = 4.23, SD = 0.90). With regard to task performance metrics, the mean accuracy of expelled shots/projectiles for all three scenarios was relatively close, ranging from 60.80% for 360_DEG to 65.48%. It is necessary to point out, though, that the number of shots fired decreased with wider spawn angles, dropping from 174.17 for 90_DEG down to 103.93 for 360_DEG, which means that, overall, significantly more targets were successfully hit during the 90_DEG scenario.

As opposed to mixed results for visual shooting aids and target spawn angle, results for the examination of shoot force were very conclusive, with a large majority of participants preferring the 80N scenario. This scenario received the highest QoE score (M = 4.47, SD = 0.68), with 40N receiving the average score of 4.23 (SD = 0.77), and 20N receiving the average score of 3.63 (SD = 1.00). Furthermore, 97% of participants were willing to continue playing the 80N scenario, compared to 80% for 40N and only 50% for 20N. A higher shoot force also helped participants experience more fun and perceive themselves as more competent, while the lowest shoot force significantly increased the overall challenge of the scenario, as well as its mental demand and task control difficulty. With regard to VRISE, there were no significant differences between scenarios. Significant differences were found for task performance, as 80N resulted in the highest accuracy (64.53%) with 192.2 shots fired on average, although this performance was closely followed by the 40N scenario with the average of 186.27 shots fired and the mean accuracy of 62.73%. The increased workload of the 20N impaired the participants' performance, with the average of 171.97 shots fired at the average accuracy of 51.92%.

As with the other two studies, the results of this study also confirmed the influence of manipulating the configuration of the IM implementation on the observed QoE features. The results of Friedman tests for this scenario group are presented in Table 6.7, while a detailed presentation of the results of Wilcoxon signed-rank tests with Bonferroni correction is included in the Appendix.

Table 6.7: Friedman test results for all shoot IM scenario groups (* $p < 0.05$, ** $p < 0.005$, *** $p < 0.001$)

QoE constituent	**QoE feature**	χ^2**(3)**	χ^2**(2)**	
		Visual shooting aids	**Target spawn angle**	**Shoot force**
Player experience	Competence	**30.07*** **	0.64	**27.59*** **
	Challenge	**34.15*** **	**27.83*** **	**22.71*** **
	Fun	3.18	**6.74* **	**16.00*** **
Workload	Mental demand	**21.42*** **	**11.41** **	**26.79*** **
	Physical demand	**17.84*** **	**6.40* **	**6.00* **
	Task control difficulty	**27.97*** **	**12.25** **	**24.11*** **
VRISE	Pain and muscle fatigue	**19.68*** **	1.00	1.56
	Overall sense of physical discomfort	1.20	3.89	0.11

6.8 Discussion

6.8.1 General remarks

Analyzing all three IMs in parallel, it is evident there is a lot of obvious overlap between measures of player experience and different measures of workload, which is to be expected. General response to the challenge item had a tendency to increase along with the increases in some or all measures (e.g., certain tasks were considered mentally, but not physically demanding and vice versa) of workload, while competence — for the most part — increased in the opposite direction. Aforementioned subjective measures were generally in line with the objective measures of task performance. Another measured aspect of the player experience — fun — varied between scenarios, but was less consistent with the other measures, which will be further discussed later in this section.

With regard to VRISE, even though significant differences between scenarios were noted, this effect was not observed for every type of mechanics. Furthermore, measured levels of experienced VRISE were generally quite modest. This is likely due to particular study design choices, as well as specific characteristics of the tested platform. With regard to study design, the duration of each scenario was short and immediately followed by a short break. Moreover, the intensity of the material was varied, with constant switching between configurations that were less demanding, and those that were more demanding, meaning that even the most intense parts of the VR experience were very temporally limited. Additionally, because the application used to test these mechanics was not an actual game, it lacked the actual pacing and dynamics of an actual VR game, and omitted physically demanding secondary game mechanics such as physically interacting with another in-game entity, avoiding enemy attacks, or moving with respect to a specified rhythm. This likely reduced its physical demands, mitigating the risk of extensive VRISE.

Overall, it can be observed that there is no one-size-fits all approach to implementing each of the tested mechanics. However, in a significant number of cases, participants were inclined toward less demanding scenarios. Nonetheless, we did note some parameters for which participants preferred scenarios that were higher in terms of workload. In those cases, participants' opinions were swayed by the level of fun experienced during the more challenging scenarios. Therefore it can be noted that, in the context of our studies, tested parameters may be roughly classified into three groups based on the observed relationship between workload (or challenge) and fun with respect to opinions and preferences held by the majority of enrolled participants. The first group (*entertaining workload*) consists of parameters for which added workload was deemed entertaining, and thus, the most demanding scenario was preferred. The second group (*ideal-point workload*) consists of parameters for which some added challenge was deemed entertaining, but extensive workload started to detract from the experience. In such ideal-point cases, the option with the moderate workload was generally preferred. The third group (*draining workload*) consists of parameters for which added workload was deemed draining (i.e., increased workload leads to a less fun experience), and participants favoured the least challenging option. Based on the trends observed in our results, parameters may be roughly categorized as presented in Table 6.8.

6.8.2 Guidelines for the implementation of VR IMs

Approaching each individual parameter, our results indicate certain findings regarding user preferences that could be utilized in the formulation of guidelines for the implementation of IMs.

Table 6.8: Classification of interaction mechanics parameters according to the relationship between workload and fun with respect to participant opinions and preferences

Category	**Interaction mechanics**		
	Slash	**Pick-and-place**	**Shoot**
Entertaining workload	/	remote grab	/
Ideal-point workload	force to destroy	/	target spawn angle
Draining workload	target spawn angle, weapon length	puzzle scale, collider scale, scale offset	visual aiming aids, shoot force

First and foremost, it is evident that there is no one universal solution equally embraced by all participants, which motivates the primary guideline (labelled as G.G1) that can be generalized to all mechanics and parameters.

G.G1 Given the diverse range of user needs, preferences, abilities, and situational variables of each gaming session, it is advisable to provide several options to each implemented interaction mechanic, going deeper and beyond the usual difficulty levels commonly encountered in digital gaming. For example, accessibility of the application could be increased by introducing a properties menu that allows users to adjust specific values of relevant parameters such as those pertaining to the play angle, target positioning and scale, realistic and hyperrealistic aiming/grabbing aids, etc.

Following the specification of this general-purpose guideline, Table 6.9 presents a set of IM-specific guidelines related to the implementation of slash (G.SLX), pick-and-place (G.PPX), and shoot (G.SHX) mechanics. These guidelines are based on the presented findings of our three studies and may not reflect the opinions of a broader audience of consumers, which calls for further research.

6.9 Key takeaways

This chapter further dissects the design factors pertaining to interaction mechanics described in Chapter 4. Aiming to supplement theoretical findings with a more practical exploration of described concepts, this chapter introduced the INTERACT framework — a proposed set of design principles/guidelines for the implementation of applications to be used as tools for the perceptual assessment of VR IMs. After highlighting relevant implementation parameters and listing potential measures that may be of interest to fellow researchers, the chapter presents the results of three user studies involving a total of 90 participants. Considering that the research objective of investigating VR IMs requires a more in-depth approach, instead of using commercial games as test material, the choice was made to utilize a custom application specifically designed for this purpose. Inspired by the games used in Studies 1 and 2, and developed in line

Table 6.9: Guidelines for the implementation of VR interaction mechanics

IM	Considered IM parameter	Label	Guideline
Slash	Target spawn angle	G.SL1	Horizontal target spawn angle should fit within the FoV of the player. 90 degree spawn angles are preferred, while spawn angles beyond 180 degrees should be avoided.
	Force to destroy	G.SL2	Players prefer slash implementations that do not require excessive physical effort. It is advisable to implement delicate targets that can be destroyed with minimal force. However, the preference pertains primarily to targets that are destroyed with a light slashing force, as opposed to targets that are destroyed immediately upon contact with the weapon regardless of slashing movement.
Pick-and-place	Puzzle scale	G.PP1	Objects sized between 10 and 30 centimeters in each dimension (height, length, and depth) tend to be well received by players in a task where placement precision is of high priority. When choosing the sizing of objects to be handled in a similar task, significantly larger objects should be avoided, while objects smaller than the listed dimensions may be well tolerated.
	Collider scale	G.PP2	When setting a task of fitting a puzzle piece into its designated position, it is advisable to allow for a solution space collider that is slightly larger than the solution space itself. This lowers the precision requirements of the puzzle placement task, and thus reduces the workload necessary for puzzle pieces to be accepted into the solution space.
	Remote grab	G.PP3	Introducing hyperrealistic or fully magical elements, such as the remote grab option, should be considered as a way to improve the accessibility of the game that involves handling objects. Furthermore, such elements may improve the entertainment factor of the application.
	Scale offset	G.PP4	In a task that requires objects to be picked up, positioned, and rearranged, it is advisable to leave a buffer space between objects. Designated positions where objects need to be placed need to be spaced-out as well. This minimizes the need for extensive maneuvering on the player's part, while mitigating the risk of objects being accidentally pushed out of place by another object.
Shoot	Visual aiming aids	G.SH1	The use of visual aids (laser sights and/or projectiles with a visible trajectory) improves shooting performance. Implementing a laser sight in combination with projectiles without a visible trajectory is generally favored over weapons with no visual aid capabilities, as well as over weapons that expel projectiles with a visible trajectory (with or without the use of laser sights).
	Target spawn angle	G.SH2	Even though it may not be the easiest in terms of challenge, workload, and task performance, the 180 degree horizontal target spawn angle is preferred over significantly smaller (i.e., 90 degrees), as well as significantly larger (i.e., 360 degrees) angles.
	Shoot force	G.SH3	Players strongly prefer long range weapons with projectiles that travel along a seemingly straight line, as opposed to an noticeable parabolic trajectory. Implementations with a modest shoot force, i.e., those that expel projectiles with a low initial velocity, should thus be avoided as they impair player experience in addition to negatively affecting shooting performance.

with the INTERACT framework, the application provides the means to evaluate different configurations of slash (Study 3), pick-and-place (Study 4), and shoot (Study 5) mechanics, while also recording detailed task performance measures.

Subjective measures pertaining to player experience, workload, and VRISE were supplemented with IM-specific objective measures ranging from accuracy and force exerted by players to duration needed for completing the task. After analyzing collected results, we found that acceptance of particular parameter values varied between participants, indicating that an ideal solution to IM implementation likely does not exist. However, we identified certain trends regarding participant preferences, which served as a foundation for the proposed set of practical guidelines for the implementation of VR IMs. Moreover, the results of our studies show that changing selected parameters of IM implementation tends to produce significant differences in subjective measures of observed QoE features. This finding supports the conceptual low-level model of VR QoE presented in Chapter 3, in addition to further motivating subsequent research on this topic.

The application used as test material provides the means for future experiments by offering a broad range of additional parameters and measures that have not yet been explored in the scope of this thesis or otherwise. Moreover, the implementation of current IMs can be further refined and the application can be extended with novel IMs, parameters, and objective measures. Likewise, there is room for further research utilizing additional subjective measures.

6.10 Modeling the impact of IMs on player experience, workload, and VRISE

In addition to being utilized for the formulation of guidelines for the implementation of VR IMs, the findings presented in this chapter can also be used to support the proposed low-level model of VR gaming QoE (VR_QOE_LLM_2). The initial proposed model did not provide specific detail regarding the exact parameters that may exhibit a relevant influence on measured QoE features pertaining to player experience, workload, and VRISE, but the parameters listed in the theoretical parts of this chapter can be considered as a starting point. In this instance, however, the focus is on the parameters that were tested in the scope of Study 3, Study 4, and Study 5.

Statistical tests used in the analysis of obtained results identified statistically significant differences between QoE feature scores given for scenarios with different IM parameter values. IM parameters that were shown to exhibit a significant impact on the measured features are listed in Table 6.10. These significant relationships are further indicated in Figure 6.13, which presents the adjusted version (VR_QOE_LLM_2SR) of the proposed low-level model of VR QoE (VR_QOE_LLM_2).

Table 6.10: IM parameters exhibiting a significant impact on measured QoE features pertaining to QoE constituents (based on the proposed VR_QOE_LLM_2 model) player experience, workload, and VRISE

QoE constituent	QoE feature	Interaction mechanics		
		Slash	Pick-and-place	Shoot
Player experience	Competence	target spawn angle, force to destroy, weapon length	puzzle scale, collider scale	visual shooting aids, shoot force
	Challenge	target spawn angle, force to destroy, weapon length	puzzle scale, collider scale, scale offset, remote grab	visual shooting aids, target spawn angle, shoot force
	Fun	target spawn angle, force to destroy	remote grab	target spawn angle, shoot force
Workload	Mental demand	target spawn angle	collider scale, remote grab	visual shooting aids, target spawn angle, shoot force
	Physical demand	target spawn angle, force to destroy, weapon length	/	visual shooting aids, target spawn angle, shoot force
	Task control difficulty	target spawn angle, force to destroy	puzzle scale, collider scale, scale offset, remote grab	visual shooting aids, target spawn angle, shoot force
VRISE	Pain and muscle fatigue	target spawn angle, force to destroy	/	visual shooting aids
	Overall sense of physical discomfort	target spawn angle	/	/

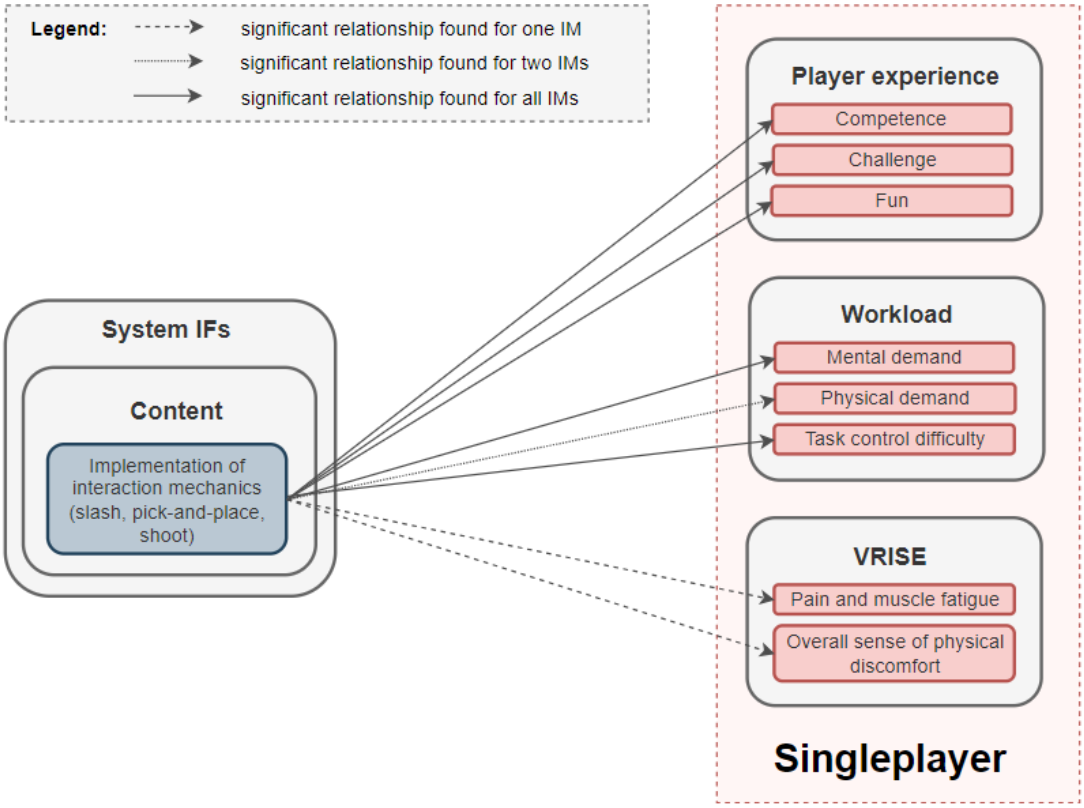


Figure 6.13: Partial low-level model (VR_QOE_LLM_2SR) of QoE for VR gaming depicting the significant relationships between the implementation of interaction mechanics and chosen QoE features for a singleplayer scenario (based on the proposed VR_QOE_LLM_2 model)

6.11 Chapter summary

This chapter presents different influence factors that may influence user experience with VR IMs. In addition to introducing the INTERACT framework — a set of guidelines created to serve as a conceptual foundation for creating applications to be used as tools for user research — the chapter also lists key IM parameters and measures that may be considered for future research in the field. Putting the described concepts into practice, the results of three different studies are presented and utilized for the formulation of QoE-driven guidelines for the implementation of VR IMs, while also serving as input for the proposed conceptual model of VR gaming QoE.

Chapter 7

Exploring the multiplayer VR gaming experience

7.1 Introduction

In previous chapters, the impact of different factors on various VR gaming QoE features was investigated with experimental setups that were organized to involve only one participant at a time. However, switching to a networked multiplayer mode brings a range of new challenges that extend beyond those of a singleplayer experience. Based on the proposed sub-model of VR gaming QoE (as described in Chapter 3), this chapter focuses on the impact of network and social factors on QoE features pertaining to perceived quality of networking, perceived social interaction and interplayer involvement experience. The chapter begins with a brief overview of two preliminary studies, referred to as Study 6 and Study 7. The findings of these studies provided motivation, as well as useful input for the more comprehensive Study 8, which is the main focus of this chapter. A more comprehensive analysis of Study 6 and Study 7 was published in two conference papers presented in the scope of the author's doctoral research [45, 226].

Analyzing the impact of different system (network type, levels of latency), player (level of expertise), and context (relationship between co-players) IFs, the remaining part of the chapter (which reports on Study 8) highlights the importance of investigating multiplayer VR experiences through divergent exploratory studies with the goal of identifying challenges to be further addressed in more focused studies. Furthermore, it discusses the implications of such factors on the design of user studies and the process of participant recruitment. The results of Study 8 were first published in a conference paper [227]. Extended from [227], this chapter (except for Section 7.2) contains the material from a recently accepted journal article [12].

7.2 Motivation and preliminary studies

This section presents a concise outline of the objectives, methodology, results, and findings of Studies 6 and 7. It is important to note that Study 6 preceded Study 7, and the described methodological differences between studies are the results of conscious decisions made as a way of addressing identified potential issues in Study 6.

The objective of both studies was to determine the impact of network factors — namely network latency — on QoE during VR gaming. Given that fast-paced action games and games with a first-person perspective (such as typical FPS games) are considered to be among the most latency sensitive games for non-immersive platforms [228], a hypothesis was made that the immersivity and embodied multimodal interactivity of the VR platform might render VR FPS games even less robust to changes in network performance due to players expecting more realistic —i.e., immediate — game responses to their actions. However, after conducting Study 6, obtained results indicated that levels of latency as high as 300 ms were imperceptible to most participants. Attempting to find explanations for these unexpected findings, several aspects were identified as potential confounding factors: level of difficulty, weapon choice, and social context. Thus, the methodology design of Study 6 was reworked into the new methodology for Study 7, updated with changes to identified factors, along with technical upgrades (i.e., a different latency simulator), a wider range of tested network latency values, and additional objective and subjective measures. A summary of differences between the two studies is given in Table 7.1.

Both studies were conducted in a laboratory environment, using the same VR game — **Serious Sam VR: The Last Hope**. Considering the same game was used as an example of a VR FPS in Study 1, its general overview was already provided in Chapter 5. However, as opposed to Study 1, in which a single-player mode was used, Studies 6 and 7 utilized its two-player cooperative mode, in which both users were positioned side by side, joining forces as they fight off waves of enemies. While the same in-game map/arena was used in all studies, Studies 6 and 7 differed based on the level of difficulty. In Study 6, participants were swarmed with numerous enemies, resulting in a very chaotic experience. Participants were forced to shoot in all directions with no time to focus and aim their weapons, nor to wait and see whether their shooting attempts were successful. Constantly turning around toward incoming enemies, participants were noticeably distracted from their shooting performance, which likely concealed the detrimental effects of added network latency. This issue was compounded by the free choice of weapons provided to each player. As explained in Chapters 4 and 6, there are multiple parameters that may impact the user experience with each IM. Carefully analyzing the visual effects of network degradation for the shoot IM, it was found that latency may be concealed through the use of weapons with projectiles that are difficult to observe as they traverse through

Table 7.1: Comparison of Study 6 and Study 7, adapted from [45] (note: RTT refers to the round-trip time, not one-way delay)

	Study 6	**Study 7**
No. of participants	24	33
Age distribution	14 - 38 (average: 26.54)	15 - 51 (average: 25.63)
Gender distribution	10 females, 14 males	12 females, 21 males
Experience with VR	beginners: 8; intermediate: 8; advanced: 8	beginners: 17; intermediate: 14; advanced: 2
Experience with FPS games	N/A	beginners: 6; intermediate: 19; advanced: 8
Level of difficulty	6	1
Co-player activity	*active co-player*	*passive co-player*
Weapon choice	undefined	weapons displaying visible projectile trajectory
VR system used by participants	Oculus Rift (examined participant), HTC Vive (*active co-player*)	Oculus Rift (examined participant), HTC Vive (*passive co-player*)
Latency simulator	Clumsy	Net.Shark
Latency scenarios (RTT)	150 ms, 200 ms, 250 ms, 300 ms	50 ms, 100 ms, 150 ms, 200 ms, 250 ms, 300 ms
Testing procedure duration	approx. 40 minutes per pair of participants	approx. 35 minutes per participant
Observed subjective measures	overall QoE, willingness to continue playing in given conditions	overall QoE, willingness to continue playing in given conditions, perceived weapon precision
Observed objective measures	survival/*death*	survival/*death*, co-player survival/*death*

the air, which could have further contributed to the inconclusiveness of Study 6 results. Thus, for Study 7, participant choices were constrained to weapons with a visible projectile trajectory, as opposed to weapons that only displayed a muzzle flash effect (Figure 7.1).

Participants in Study 6 were paired together during gameplay, but only one of them (the one being delayed based on the used setup) was being tested, the other taking on the role of an *active co-player*. Participant roles were switched after completing all latency scenarios. Considering that participants were allowed to communicate during the session, the presence of another player served as further distraction to some participants, especially in cases where co-players were previously familiar with each other, a factor that was not controlled during the

Figure 7.1: Different types of weapon triggering different visual effects (taken from [45]): a) weapon displaying individual projectiles with a visible trajectory, b) weapon displaying only a muzzle flash effect

participant recruitment process. Moreover, as the overall in-game success was dependant on simultaneous efforts of both players, the existence of an *active co-player* hindered the analysis of task performance measures. Individual contributions of each player were difficult to identify, especially as more experienced FPS gamers naturally assumed a dominant role, aiding and defending their less experienced co-players. To eliminate the confounding impact of social context, in Study 7 the game was still being played in co-operative mode on two HMDs, but the tested player was joined by a static entity in a virtual world, as there was no-one behind the other headset. In addition to preventing the participant from taking part in social interaction (also mitigating the potential impact of prior relationships, co-player gender and skill level), this change simplified the task performance analysis as the overall in-game success was only dependent on one person. Furthermore, this choice further distinguished the impact of network latency on the delayed player's success as well as on their ability to protect their co-player from the impact of social factors.

After introducing significant changes to the methodology, the results of Study 7 were considerably more conclusive. As opposed to participants' relative indifference toward high levels of latency in Study 6, the experiences of participants in Study 7 were significantly deteriorated for scenarios exceeding 100 ms of latency added to the RTT, which corresponded with the decline in perceived weapon precision. While latency thresholds tend to vary between games as well as genres, these results indicated that VR games may be comparable to non-VR games in terms of sensitivity to network latency. However, while subjective measures indicated a latency sensitivity threshold of 100 ms, notable differences in objective measures occurred between 50 ms and 100 ms scenarios. This was especially evident for *passive co-player deaths*.

It is important to note, however, that Serious Sam VR: The Last Hope, even though it classifies as a co-operative game, does not provide a very high level of interplayer interactivity (note, interplayer interactivity in this context refers to the interaction between player roles in the context of the game; it is not to be confused with interplayer involvement, although it may provide context for the interpretation of interplayer involvement experience measures, as interplayer involvement experience may be shaped by prominent patterns of interplayer interaction as core gameplay elements, which will be discussed later on). While individual contributions of each player contribute to the overall success of the team, and players are sharing the same virtual space and fighting common enemies, this game — by design — is not allowing the players to interact directly or through shared objects. Moreover, by playing from a fixed position, side by side, players are not even required to look at each other, with the hectic gameplay further discouraging them from doing so. **Thus, it should not be presumed that competitive VR games or highly interactive co-operative VR games would necessarily exhibit the same latency threshold.**

Even though the specific setup for Study 7 fails to replicate a realistic gaming scenario due to not having an *active co-player* in a multiplayer game, its findings highlight the likely impact of contextual factors on the results of a multiplayer gaming user study. Although individual contributions of the potential confounding factors identified in this set of preliminary studies can not be quantified, they highlight a larger issue. Their main takeaway is as follows: **results obtained during a multiplayer gaming study, even when its objectives and measures are clearly focused on a specific factor (in this case - network latency impact), need to be observed and discussed with respect to its broader context involving social interaction, interplayer involvement experience, and specifics of the game itself, such as interaction mechanics and interplayer interactivity.**

The goal of delving further into the impact of network factors on user experience, while also addressing surrounding influence factors, was the driving force behind subsequent Study 8, which was significantly more divergent in focus and measures. In this study, participants were recruited in pairs so that they were either completely unfamiliar with their co-player, or bonded in a close relationship. In addition to the impact of network quality on user experience, measures pertaining to QoE constituents social interaction and interplayer involvement experience were employed. Understanding the importance of genre, interaction mechanics and interplayer interactivity in the experience of networked multiplayer gaming, Study 8 involved the use of two different VR games: **Blaston**, an FPS dueling game with similar mechanics to Serious Sam VR: The Last Hope, but with a competitive gameplay and opponents facing each other, as well as **Eleven Table Tennis**, a sports game with a high level of interplayer interactivity realized through alternate control of a mutual object (i.e., the ball).

7.3 Related work

7.3.1 Multiplayer VR gaming as a networked activity

With respect to previous research addressing networked games [228, 229, 230, 231, 232] most VR games can be characterized as having an *avatar interaction model with a first person perspective* — thus exhibiting specific traits that are known to be especially sensitive to network degradations such as latency. Coupled with the increased expectations of interaction realism and fluidity that are implied by natural interaction mechanics present in a number of VR games, it can be theorized that locally-rendered VR games would likely have a latency threshold that is either comparable to non-immersive games of similar characteristics, or possibly even lower. However, at the time of this writing, there is limited work investigating the extent to which network latency affects VR games. While studies addressing latency in VR exist [233, 234, 235, 236, 237, 238], they are often focused around investigating motion-to-photon (MTP) latency, which is more of an issue for VR cloud-gaming, as opposed to locally-rendered games.

Further focusing on locally-rendered multiplayer VR games, the impact of different levels of latency (30 ms, 100 ms, and 500 ms) on gaming QoE and related dimensions was measured by Kojic et al. [239]. According to their results, added latency had an effect on measures of QoE, presence, and flow during a simulated competitive rowing exergame; however, because a rowing game is not characterized by high interplayer interactivity, nor does it require precision, results were not as significant as they could have been if a different game was used.

Interfering with presentation constistency, physics consistency, and interaction consistency [240], the exact manifestation of network latency varies between games, but as with non-immersive games, VR games may benefit from the use of lag compensation techniques — software techniques that are used on the client, server, or both, with the goal of decreasing the impact of network latency [241]. As noted by the authors, there is a need for further research on the impact of lag compensation techniques on a broader spectrum of gaming platforms, including VR. However, considering the increased tracking capabilities of VR systems, with multiple body parts being tracked in six degrees of freedom and moving non-deterministically, mitigating the effects of latency in VR games is not an easy task.

7.3.2 Multiplayer VR gaming as an interactive interplayer competitive activity

As addressed by Kojic et al. [239], the influence of latency is likely to be more extensive in case of games that require a significant level of interaction between users. However, when choosing to measure participants' experience with a task that is highly interactive, there are multiple additional aspects that may influence QoE and related dimensions. If the methodology involves task performance metrics, it may be useful to consider the posibility that participating in a task that is competitive may improve task performance, but this effect occurs only for participants with more competitive personalities, as was the case with a sample of older adults who were trying out a VR exergame in a study conducted by Anderson-Hanley et al. [242].

Another aspect to consider in a competitive scenario is the way in which the success of each participant may reflect on their affective state, motivation, and engagement, possibly impacting their overall QoE. For example, participants in a study by Ventura et al. [243] were tasked with hitting targets in a VR game, playing alone or against a virtual opponent, which was programmed to either win, lose, or tie with the player. Results showed that player engagement was higher in case they lost to the virtual opponent compared to their engagement when playing alone, an effect that did not occur for cases when the virtual opponent lost, or the game was tied. In the context of performing QoE studies, similar effects could occur with participants playing against other humans, thus highlighting the possible benefits of appropriate skill-based matchmaking of study participants.

An obvious method of pairing participants based on this criteria is to rely on their previous level of experience. However, determining what kind of experience to look for is not straightforward, especially considering that the spectrum of multiplayer VR games is incredibly diverse, while individuals with extensive or highly specific VR gaming experience are still fairly hard to find. However, the relevance of considering prior expertise is highlighted in ITU-Recommendations pertaining to similar services. ITU-T Recommendation G.1032 [244] (gaming-specific) lists general gaming experience, as well as gaming experience with a particular game or genre as influencing factors impacting the overall gaming QoE. According to ITU-T Recommendation G.1035 [156] (VR-specific), expectations and expertise are listed as QoE influencing factors for VR services. The document highlights users' previous experiences with VR technology as possible contributors to the overall QoE, while also noting that real-life experiences and interactions may influence QoE during experiences and interactions set in the virtual world. This influence of real-life experience and perceptual, cognitive and motor skills on performance in VR was demonstrated by novice, academy, and professional soccer players who were subjected to VR soccer in [245]. The transference of skills likely goes both ways, as training with a VR table tennis game was shown to improve performance in real-life table tennis [246].

Focused on the broader context of VR use, the latter aforementioned ITU-T recommendation (G.1035) did not address previous experience with gaming (using either VR or non-immersive platforms) as a potential influencing factor for VR services. However, it is worth mentioning that experience with non-immersive gaming can influence user performance (and thus likely also influence QoE) in VR, even in case of seemingly uncomparable applications. For example, playing a first-person shooter desktop game for five weeks was shown to improve performance in a VR surgical simulator [247].

7.3.3 Multiplayer VR gaming as a social activity

The presence of other people in the context of a user study may influence different aspects of the user experience. When matchmaking participants for participation in user studies, a social context factor that may be considered is prior relationship between participants. An example of a study where participants were paired together based on this criterion, forming pairs of friends and pairs of strangers, is provided by Rivu et al. [248].

As seen in previous studies, the experience of gaming can be influenced by the level of familiarity between players participating in a multiplayer game. For example, as seen in the work of Mandryk et al., playing with friends (compared to strangers) may improve the levels of excitement and fun. On the contrary, playing with strangers as opposed to friends may increase toxicity, while decreasing relatedness [249].

Further focusing on the VR platform, a longitudinal study presented by Moustafa and Steed [141] was performed on different groups of participants defined by their previous relationships (strangers, friends, siblings, romantic couple, parent and child). Results indicate that group dynamics in a VR environment are consistent with the group dynamics from the real-world. An older study which used a CAVE environment [250] showed that, contrary to initial expectations, groups of strangers and friends were similar in terms of collaboration, behavior and enjoyment in the virtual environment.

Social context in VR was also addressed by Kojic et al. [251], who explored the effect of using a virtual environment (VE) on perceived flow and presence, along with investigating the impact of the ability to communicate on user experience, showing that the sense of social presence was increased when participants were able to communicate with their opponent and concluding that communication could improve QoE in VR exergames.

Lastly, Liszio et al. [252] explored methods of improving the sociability of VR games through the integration of social entities. The authors conducted a study in which they explored the impact of a virtual agent and a co-player on the VR player experience and its dimensions (enjoyment, social presence, perceived loneliness, immersion). Their findings indicate that including a social component may reduce loneliness and thus improve the overall experience in VR games.

7.4 Methodology

7.4.1 Materials

Two VR games were used in this study: Eleven Table Tennis (ETT)* and Blaston†, both chosen for their simple, easy-to-grasp mechanics which were appropriate for a single-session study, as well as for their fast pace, making them sensitive to explored variations in network performance. Both games are designed for two players, have a first-person perspective, and both can be classified as competitive and highly interactive in terms of interplayer interactivity. In terms of in-game locomotion, both games support the Room-scale mode (and are thus free of major cybersickness triggers [253]), as players are only expected to move within a small in-game area with respect to particular boundaries imposed by their environment (i.e., stay on top of the platforms in the middle of a dueling arena in Blaston, move around their half of the tennis table in ETT). The games differ in terms of genre, as Blaston is a dueling first person shooter game, and ETT is a sports game that simulates a real table tennis match. While both games feature projectile-based interaction mechanics [10], it is important to note significant differences in in-game physics with regard to projectile motion. The ball in ETT behaves in a completely realistic

---

[*] https://elevenvr.com/en/

[†] https://www.resolutiongames.com/blaston

Figure 7.2: Screenshots from the chosen games used in Study 8: Blaston (left), ETT (right)

manner, and the game utilizes haptic feedback capabilities of Meta Quest handheld controllers to further enhance the realism of the moments in which the ball makes contact with the racket. On the contrary, bullets and missiles expelled from Blaston's weapons are significantly slowed down compared to the motion of real-life bullets, which was a deliberate choice made by the game's developers, making it possible for players to physically dodge incoming attacks. During the game, players also switch between different weapons with projectiles of different calibers and velocities. Once the player runs out of ammunition, they are expected to discard their handheld weapon and quickly grab another from several types of weapons floating in the air around them.

Moreover, as evident from Figure 7.2, the games also differ in terms of visual design. In the scenes used during the experiment, ETT showcases a more realistic, muted indoor environment with seemingly diurnal lighting, while Blaston's design features a more stylized, darker aesthetics contrasted by bright neon details and visual effects. In terms of user representation in this study, during Blaston participants were represented by a stylized half-body avatar with visible hands, while ETT only showed the virtual representation of the user's HMD and a tennis rackets in place of the participant's controller. In terms of audio design, ETT features realistic sounds of the ball colliding with the table and the players' rackets, while Blaston features background music and dynamic shooting sound effects. Both games feature the option of using voice chat, which was used in the study so that participants partaking in a study session were able to verbally communicate with each other. Both games were played on identical Oculus/Meta Quest HMDs with the accompanying controllers.

7.4.2 Procedure

The study took place in two adjoining laboratories (Lab A and Lab B), separated by a soundproof wall. Each laboratory was kept quiet from disturbing noises and at a comfortable temperature. The space was cleared to allow for increased movement expected during active VR gameplay, and to ensure the safety of participants whose vision was occluded by the VR head-mounted displays (HMDs).

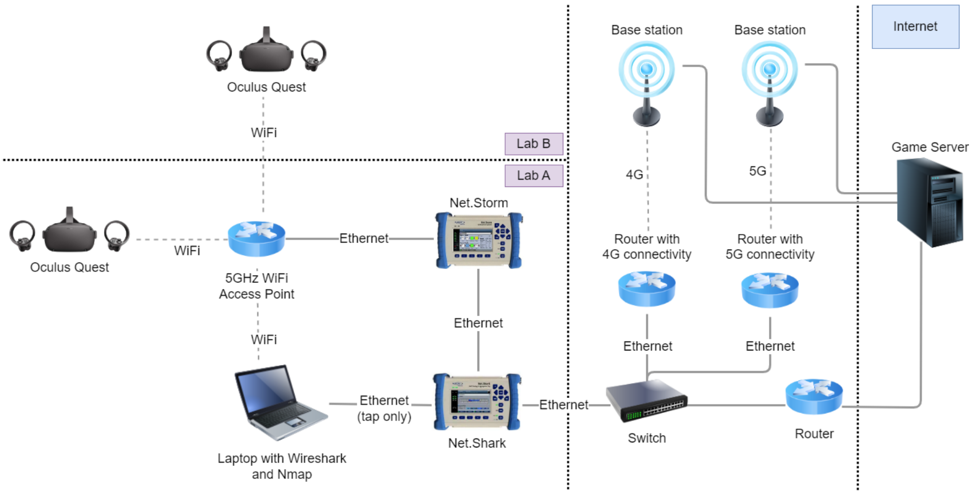


Figure 7.3: Laboratory setup: two participants, one located in each Lab (Figure taken from the original paper [227]).

As depicted in Figure 7.3, the used laboratory prototype included:

- two identical Meta Quest HMDs with accompanying handheld controllers;
- two routers, each connected to an access network (4G, 5G);
- a WiFi access point (positioned at the same distance from both Lab A and Lab B);
- an Albedo Net.Storm [‡] device used to generate network impairments;
- an Albedo Net.Shark [§] device used for network tapping;
- a laptop for measuring Round Trip Time (RTT) and storing captured network traffic.

The laptop used to conduct the RTT measurements and both VR headsets were connected to the same WiFi access point on the 5GHz band. The WiFi network was connected to the Internet via 4G and 5G, as well as via an Ethernet fiber optic line. 4G and 5G connections to the Internet were realized through a dedicated base station located on the Faculty of Electrical Engineering and Computing premises in cooperation with Croatian Telecom. This laboratory configuration was chosen with the goal of simplifying the test setup, facilitating the switching between 4G, 5G, and Ethernet test scenarios, and adding support for delay emulation and network traffic tapping, as opposed to directly connecting the VR headsets to 4G and 5G base stations, or alternatively using a mobile hotspot. Between testing scenarios, the Internet connection was manually switched by the study administrator.

Albedo Net.Storm and Net.Shark devices, which were used for inserting added delay and network tapping, respectively, were set up as intermediary nodes between Internet access and the WiFi access point. Wireshark was used to analyze network traffic data captured during gameplay. Wireshark was also used to find out the IP addresses belonging to game servers. Even though ICMP traffic to game servers was blocked, testing the reachability of acquired IP addresses was conducted by periodically sending a TCP ping using the Nmap tool.

[‡] https://www.albedotelecom.com/pages/emulation/src/netstorm.php

[§] https://www.albedotelecom.com/pages/fieldtools/src/netshark.php

Table 7.2: Test scenarios used in the QoE study (Table taken from the original paper [227])

Pair	Game	Network scenario A	Network scenario B
1	ETT	4G	5G
2	ETT	Ethernet	5G
3	ETT	Ethernet+100ms	Ethernet+200ms
4	Blaston	4G	5G
5	Blaston	Ethernet	5G
6	Blaston	Ethernet+100ms	Ethernet+200ms

During each session, network factors were manipulated for each game, switching between different access networks (4G, 5G, and Ethernet) and different levels of added network latency (Ethernet+100 ms, Ethernet+200 ms), depending on the scenario. It should be noted that when talking about the added 100 ms and 200 ms of latency we refer to latency added along the whole network path between the players. For the first scenario, 50 ms of latency was added from each client toward the server. As the traffic between clients flowed through the server, it resulted in a total latency of 100 ms. For the second scenario, the same method was applied, but the amount of latency was doubled, resulting in a total of 200 ms of added latency. Latency was introduced in this way as the state for both games is held on the clients, and not on the server which only serves as a relay for the exchanged messages. While the extent of added latency (i.e., the additional 100 ms and 200 ms of RTT between the players) was significant, it should be noted that similar levels of delay may occur in mobile networks due to congestion or reduced channel capacity.

The duration of each study session was between 70 to 90 minutes. All questionnaires were given as online forms displayed on a desktop computer. Upon arrival, participants signed informed consent sheets, filled the *pre-study questionnaire* (to be discussed in the following subsection), and signed consent forms. After receiving information on the use of VR technology, they were taken to separate rooms (Lab A and Lab B), as each member of a pair was assisted by a separate administrator. After completing a short tutorial session explaining the rules and mechanics of the first game, participants went on to complete all scenarios for that game, each lasting 2-4 minutes (each test scenario involved participants playing one set in ETT up to 11 scored points, or three game rounds in Blaston).

In addition to asking participants to provide an absolute rating for each scenario, comparison ratings between chosen scenario pairs were also collected, as defined in Table 7.2. It is important to note that the order of pairs was randomized, as well as the order of scenarios within each pair of scenarios, and network was the only aspect that was changed. Participants played both games in succession, i.e., after they were finished with all scenarios for the initial game, they moved onto the other. As with scenarios and scenario pairs, the order of games was randomized with the goal of avoiding the influence of ordering effects on subjective scores.

Between scenarios, participants' were given the *post-scenario questionnaire*, asking them to rate their overal QoE (on a 5-pt Absolute Category (ACR) scale) and the interaction quality using a modified version of the Gaming Input Quality Scale (GIPS) [254]. Following every second scenario (i.e., after each pair of scenarios was completed), participants were asked to compare that scenario to the previous one using a 7-cat. Comparison Category Rating (CCR) scale.

After all scenarios for a particular game were finished, participants were asked to fill the *post-game questionnaire*, which contained a modified version of the Core Module of the Game Experience Questionnaire (GEQ) [20]. This questionnaire was used for measuring *flow*, *competence*, *positive affect*, *negative affect*, *tension*, and *challenge*. Moreover, the *post-game questionnaire* also included a subset of questions from the GEQ Social Presence Module. Included questions were those pertaining to *negative feelings*, as we considered these questions to be of high relevance to the aims of the study pertaining to exploring competitive feelings between players. The second part of the *post-game questionnaire* included a compilation of questions and statements taken from [113, 255, 256, 257], and focused on participants' social interaction and bonding, competitive feelings, and experience of suspense during gameplay. Finally, participants were also asked to report their overall QoE and rate their own skill with respect to the skill of their opponent for that particular game, as well as to select statements that applied to them from a set of statements pertaining to their preferences regarding the choice of opponent, single or multiplayer gaming, as well as the influence of social relationships and balance of skill on their experience. Note, specific items explored in the *post-game questionnaire* will be listed and discussed in more detail along with their respective results.

7.4.3 Participants

Considering that one of the research goals of this study involved exploring social interaction during VR gameplay, a quota sampling approach was used to recruit participants based on their prior relationship. Therefore, the 32 enrolled participants were joined into 16 pairs, 8 of which were pairs of close *friends*, and 8 of which were pairs of individuals that had never met prior to participating in the study, which we refer to as *strangers*. The sample was also balanced in terms of sex, with 16 female and 16 male participants.

Information about the age of participants was reported in the pre-study questionnaire given to participants before entering VR. Participants were aged between 21 and 42, with the average age of 25 (median age 24) years. The pre-study questionnaire also included questions investigating participants' competitiveness (Revised Competitiveness Index [258]) and the Big Five personality traits (the Mini-IPIP scales [259]). However, these measures are considered out of scope for this thesis and are thus not reported.

A common practice in gaming user studies is to report participants' level of expertise with gaming (see [244, 260, 261]), while reporting previous experience with VR technology is recommended for VR studies [156] . In the context of multiplayer VR, having knowledge of this user-related factor may aid in grouping participants according to their skill level. However, knowing that certain games emulate very specific and occasionally physically demanding real-life experiences (e.g., VR sports games resembling real-life physical activities), while others also strongly resemble specific existing non-immersive game genres (e.g., VR shooter games resembling first-person shooter games on other platforms), there is a possibility that asking such general questions does not provide enough information to gauge each participants' skill with the exact type of gameplay required by a specific game. Thus, in addition to being asked to report their level of expertise with VR technology and gaming in the pre-study questionnaire, participants were also asked to report on their more game-specific skills: expertise with real-life table tennis and general expertise with first-person shooter games. All categories of expertise were rated on a scale of 1 (Beginner) to 5 (Expert). The distribution of participants according to these categories of expertise will be discussed in detail in Section 7.5.2. It should be noted that this sample of participants was generally experienced in gaming (i.e., 56.2% of participants rated themselves as either 4 or 5 on the gaming expertise scale), while only a single participant was a complete novice. The majority were PC gamers (65.5%), while the remaining participants mostly played on consoles. With regard to previous experience with VR technology, the sample was fairly inexperienced, as 46.9% of participants were complete novices (1 on the scale), while the remaining participants reported varying levels of expertise.

7.5 Results

7.5.1 Multiplayer VR gaming as a networked activity

Mean RTT values measured periodically every second for the duration of each test scenario are presented in Figure 7.4. It should be noted that the presented RTT values are between the player and the game server, so presented values should in general be doubled when considering player to player latency. Only player to server latency could be measured using the ping command and estimate player to player latency through doubling those values as both players were in the same LAN and pinging them directly would report local delays. We checked whether games sent any traffic directly between clients without using the server as a relay, but no direct traffic was detected. On average, the 5G network produced slightly lower latencies compared to 4G. For the scenario pair comparing Ethernet with 5G, 5G had higher levels of latency compared to Ethernet. It should be noted that the addition of 200 ms between the players incurred high variability in measured latency with around 30 ms between first and third quantile.

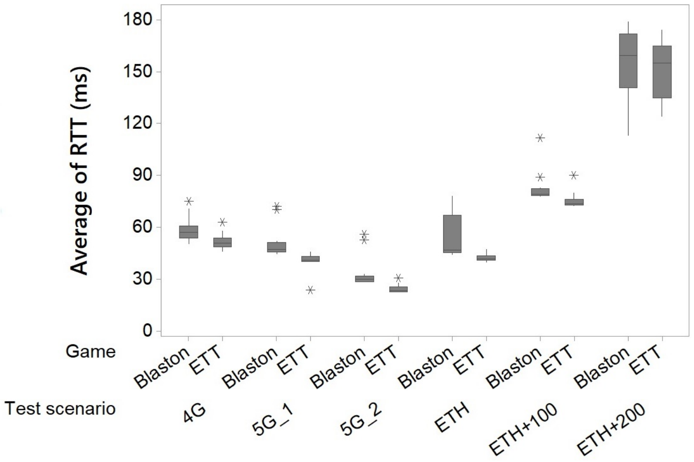


Figure 7.4: RTT measurements for tested network types, taken periodically every sec. during the test scenarios (Figure taken from the original paper [227])

The impact of access network on VR gaming QoE

Overall QoE scores given to different scenarios are presented in Figure 7.5. Comparing the 5G scenario to 4G, it can be seen that — on average — both games received higher QoE scores when they were played over 5G compared to 4G, but (according to the Wilcoxon signed-rank test) this effect was not statistically significant for either ETT ($Z = -0.91$, $p = 0.37$) or Blaston ($Z = -0.05$, $p = 0.95$), which is unsurprising considering that differences in measured latency between the two access networks were very small. Comparing 5G to Ethernet, we found that Ethernet scored higher on average, but no statistically significant differences were found for either of the tested games ($Z = -0.62$, $p = 0.54$ for ETT, $Z = -0.55$, $p = 0.58$ for Blaston), leading us to the conclusion that the performance of 5G was comparable to Ethernet according to participants' subjective ratings.

In addition to QoE scores, we also report on results obtained through CCR comparing each pair of scenarios, as presented in Figure 7.6. Even though the order of test scenarios was randomized during the actual testing, for the analysis we corrected the scores so that either 4G or Ethernet (depending on the scenario pair) are always first, and 5G is always second. Based on these findings it can be seen that scores for 4G and 5G are leaning toward positive values, indicating that 5G scored slightly better than 4G. However, the opposite was true for the 5G/Ethernet scenario, as 5G was rated slightly worse. However, due to very small differences in measured network latency and subjective scores, it can be concluded that **user studies exploring multiplayer VR gaming on portable headsets can be conducted using either access network without this factor significantly affecting the experience of participants.**

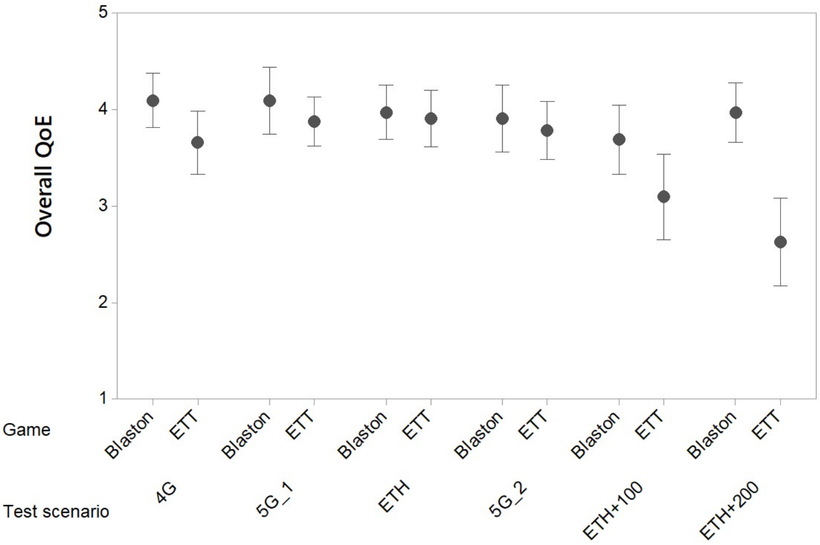


Figure 7.5: Subjective user ratings of overall QoE (95% CI) (Figure taken from the original paper [227])

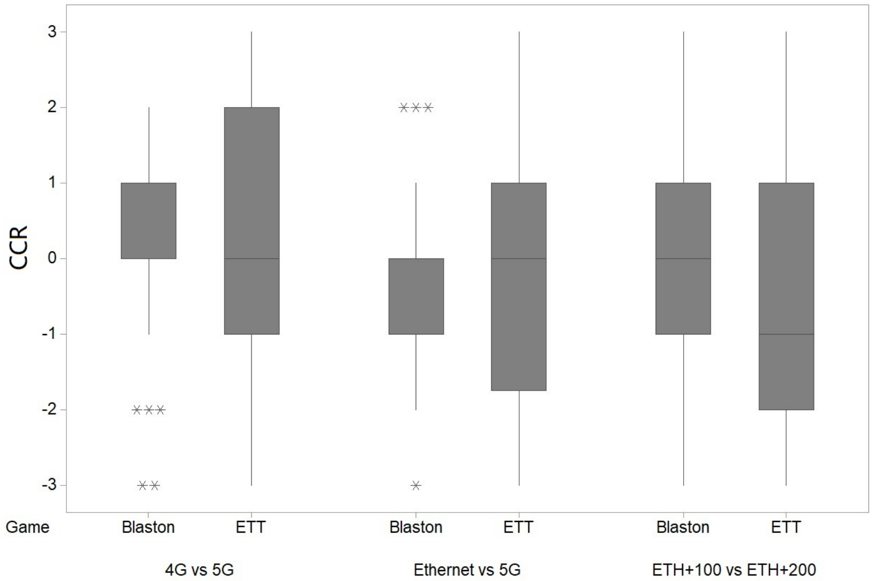


Figure 7.6: CCR for comparison of different pairs of the test scenarios (95% CI) (Figure based on the original paper [227])

The impact of network latency on gaming QoE

QoE scores given to the scenarios with added network latency are presented in Figure 7.5. For ETT, test scenarios with lower latency received better scores on average compared to test scenarios with higher latency, which is to be expected. Significant differences between scenarios with different levels of latency were confirmed by the Friedman test ($\chi^2(2) = 16.67$, $p < 0.001$) followed by a post hoc analysis with Wilcoxon signed-rank test and Bonferroni correction which showed significant differences between ETH and ETH+100 (Z = -2.79, p = 0.005) and ETH and ETH+200 (Z = -3.91, $p < 0.001$).

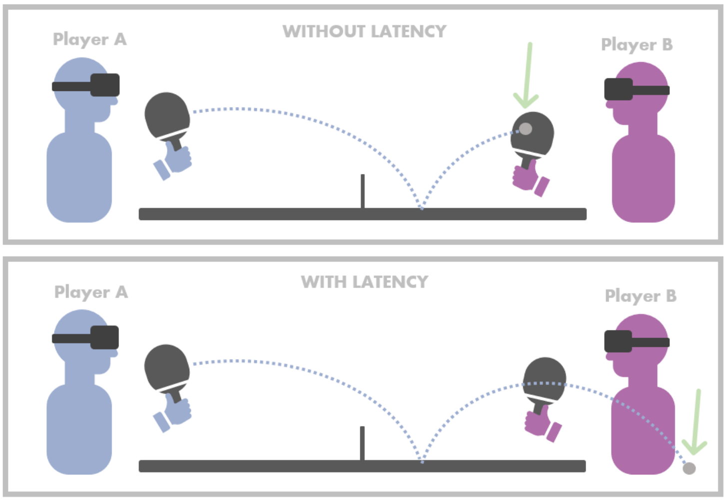


Figure 7.7: The trajectory of a table tennis ball being served by Player A and returned by Player B (as seen by Player A) — the green arrow marks the position of the ball at the exact moment in which Player B returns the ball (taken from [12])

However, Blaston yielded different results, as test scenarios with added 200 ms of latency unexpectedly received higher mean scores compared to test scenarios with added 100 ms of latency. Moreover, no statistically significant differences were found between scenarios ($\chi^2(2) = 1.83$, $p = 0.40$). In terms of CCR (Figure 7.6), as with the other scenarios, the CCR scores were corrected for the analysis so that the scenario with 100 ms latency was always first, and the scenario with 200 ms latency was always second. For Blaston, the MOS of CCR was zero, confirming that there was no discernible difference between the two delayed scenarios. For ETT, however, the MOS slightly leaned toward the negative values, indicating that the scenario with lower latency was preferred over the 200 ms latency scenario.

These results indicate that network degradations have a more noticeable effect on the table tennis game, which can be attributed to the successful use of lag compensation techniques in Blaston. Because of the specifics of the genre combined with the fact that developers did not try to make a fully realistic game, mechanics of projectile propulsion in Blaston lend themselves better to higher latencies compared to the ones in ETT. More specifically, projectile (ball) motion in ETT is alternately controlled by both players through collisions with their rackets. The ball is a single physical object, remaining on the scene as long as both players are being successful in their attempts to return it. This continuity is only being interrupted when the ball is lost, and one of the players has to serve the ball again. With physics that are otherwise fully realistic, and the ball as the sole focal object of the game, players are more likely to notice any deviations from the expected fluidity of its motion. An example of noticeable issues with the behavior of the ball when significant latency was introduced is given in Figure 7.7. Similarly to the description of the blank-period problem as described in [262], once the illustrated attacking move performed by Player A was finished, it was expected to be met with Player B's reaction,

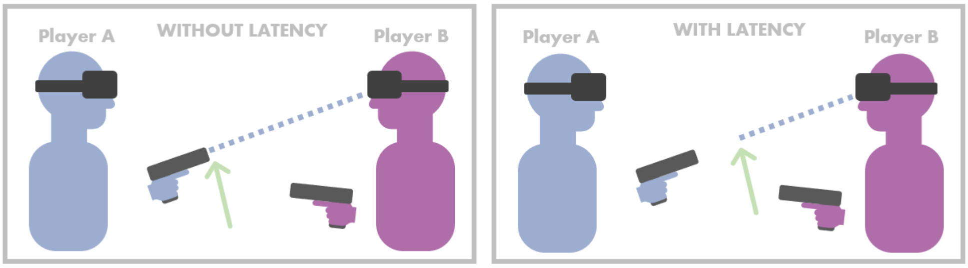


Figure 7.8: The trajectory of a bullet expelled from the weapon of Player A (as seen by player B) — green arrow marks the spawn position of the bullet (taken from [12])

but the response of Player B was delayed due to added latency. In the perspective of player A, as a way to preserve the physical consistency [240] of the virtual world, the ball was seemingly continuing along its ballistic trajectory, uninterrupted. Because of the delay, the successful reaction of Player B whose racket intercepted the ball and changed its trajectory, revealed itself to player A only after the fact, as the ball suddenly disappeared and seemingly respawned itself in a different position, traveling along the trajectory determined by the collision with Player B's racket.

With Blaston, projectiles are not objects that are continuously present on the scene. Unlike the table tennis ball that bounces back and forth propelled by both players, the motion of each projectile is generally one-directional and its velocity pre-determined by the characteristics of the weapon. Furthermore, bullets only appear once they are expelled from the weapon, and immediately disappear after reaching their target or meeting the invisible boundaries surrounding the arena. The stylized nature of the game, and the slowed-down speed in which the bullets are moving prevent players from expecting high levels of physical realism, and the simultaneous presence of multiple threatening projectiles further distracts the user from focusing on a single bullet. The combination of all mentioned factors enhances the game's method of compensating for added latency — as seen in Figure 7.8, instead of spawning from the muzzle of Player A's weapon, a delayed bullet will appear further along its trajectory. However, looking from the perspective of Player B, who is likely facing the bullet head-on, and from a certain distance, it is easy for this gap between the weapon and the newly spawned bullet to go unnoticed. Even though first-person shooters are considered a genre that is highly sensitive to latency — and thus this tolerance of high delays in Blaston could be perceived as somewhat unexpected — previous work [45, 226] involving a different VR shooter game also indicates that latencies of up to 150 ms may go unnoticed by participants. **According to obtained results, latency sensitivity (at least in terms of subjective measures) in networked VR is a more pressing issue for highly interactive games with projectile-based interaction mechanics that involve players intermittently exerting control over a particular virtual object (e.g., ball games), rather than fast-paced FPS games.**

Table 7.3: Spearman correlation coefficients between overall QoE and subjective measures pertaining to perceived delay and game responsiveness (** p < 0.001); adapted from [12]

QoE feature	**Item**	**ETT**	**Blaston**
Perceived delay	I noticed a delay between my actions and the outcomes.	**-0.61****	**-0.49****
Input responsiveness	The responsiveness of my inputs was as expected.	**0.60****	**0.61****
Input smoothness	My inputs were applied smoothly.	**0.56****	**0.64****
Perceived co-player delay	When I observed the actions of my co-player, I noticed there was a visible delay.	**-0.72****	**-0.43****
Perceived performance degradation	I feel that my performance was affected by the perceived delay.	**-0.77****	**-0.64****

The impact of network latency on the perceived quality of networking

To further confirm that discussed variations in QoE between the scenarios were likely caused by network degradations, i.e., visual manifestations of network latency as explained in the above examples, we explored the correlation between QoE scores and QoE features pertaining to participants' perception of network quality (calculated by combining the scores given to ETH, ETH+100, and ETH+200 scenarios). To measure each of the identified QoE features, a modified GIPS scale was used [254]. Using this tool, participants were asked to report their agreement with the presented statements on a scale from -3 (Strongly disagree) to +3 (Strongly agree). Measured features with the accompanying statements are presented in Table 7.3, along with Spearman correlation coefficients describing their each feature's correlation with the overall QoE score. As seen in the table, all of the features were strongly correlated with QoE, although this effect was slightly more prominent in case of ETT.

Stepping back to observe how these listed features reflect the changes in network latency, their mean values and accompanying standard deviations are listed in Table 7.4. Again, mean scores given in ETT generally correspond with the expected effects of latency. Differences in scores between the three scenarios are less significant in the case of Blaston. Although no statistically significant differences in perceived delay were identified using the Friedman test for either game, significant changes were found for all other features in case of ETT. For this game, the Friedman test identified significant differences in input responsiveness ($\chi^2(2) = 6.93$, p = 0.03), input smoothness ($\chi^2(2) = 6.53$, p = 0.04), perceived co-player delay ($\chi^2(2) = 22.16$, $p < 0.001$) and perceived performance degradation ($\chi^2(2) = 16.17$, $p < 0.001$). For Blaston, significant differences were found only for input smoothness ($\chi^2(2) = 9.03$, p = 0.01) and perceived performance degradation ($\chi^2(2) = 6.69$, p = 0.04).

7.5.2 Multiplayer VR gaming as an interpersonal competitive activity

The relationship between level of expertise and perceived competence

With the knowledge that both transferable real-life skills and gaming skills may influence task performance in VR, we chose to observe how self-reported level of expertise in different areas

Table 7.4: Mean values and corresponding standard deviations for QoE features pertaining to perceived network quality

		ETT			Blaston		
QoE feature		**ETH**	**ETH+100**	**ETH+200**	**ETH**	**ETH+100**	**ETH+200**
Perceived delay	M	-0.44	0.34	0.38	-0.44	-0.25	-0.53
	SD	1.68	2.13	2.23	1.76	1.68	2.00
Input responsiveness	M	1.41	0.56	0.34	1.59	0.88	0.88
	SD	1.46	1.98	2.22	1.29	1.83	1.70
Input smoothness	M	1.34	0.66	0.38	1.69	0.94	1.09
	SD	1.33	1.81	2.08	1.00	1.52	1.38
Perceived co-player delay	M	-0.19	1.09	1.78	-0.72	-0.22	-0.59
	SD	1.99	2.13	1.79	1.71	1.81	1.86
Perceived performance degradation	M	-0.75	0.69	1.38	-1.00	-0.19	-0.63
	SD	1.90	2.01	2.08	1.67	1.91	1.81

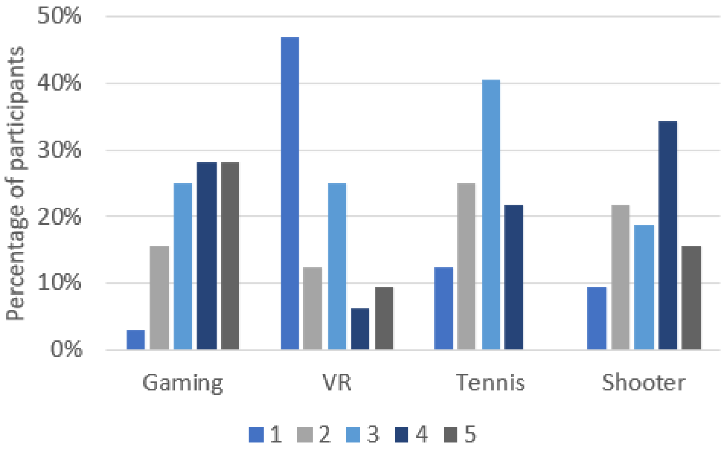


Figure 7.9: Percentage of participants who self-reported each level of expertise (1-Beginner, 5-Expert) with gaming (*Gaming*), VR technology (*VR*), table tennis (*Tennis*) and FPS games (*Shooter*), respectively (taken from [12])

(as reported at the beginning of the study, with results reported in Figure 7.9) correlated with perceived competence, as measured by the GEQ after playing each game. As seen in Table 7.5, the correlation was tested even for seemingly unrelated skills (FPS for ETT and table tennis for Blaston). For ETT, correlation between competence and previous skill with real-life table tennis was the only one that was statistically significant, although expertise with gaming, as well as specific FPS gaming skills, also showed a notable correlation. For Blaston, specific expertise with the games of the same genre displayed a statistically significant correlation, with an even higher correlation found for the more general category of level of expertise with gaming. Notably, expertise with VR showed the weakest correlation with perceived competence for both games, although it is important to note that almost 50% of participants were complete novices with VR technology, which further complicates interpretation of obtained results.

It is worth noting that significant correlation was found between certain categories of expertise, the most obvious being the correlation between general expertise in gaming and expertise in playing FPS games ($r_s = 0.73$, $p < 0.001$). Interestingly, expertise with first person shooters also correlated with expertise with table tennis ($r_s = 0.37$, $p = 0.04$), which may explain why expertise with FPS, along with general gaming expertise, showed a fairly significant correlation with perceived competence in ETT. However, the reverse was not true, as table tennis skills were not strongly correlated with perceived competence in Blaston. Considering that FPS skills were previously shown to improve performance in VR even for a completely unrelated type of application [247], these results may further indicate that skilled FPS players, even when trained on other platforms, possess a certain advantage over other demographics when it comes to adapting to a wider range of genres and applications, as well as gaming platforms. In this case, it is worth noting that ETT (like table tennis in real life) is a fast-paced game requiring quick reactions and good hand-eye coordination — skills that are well developed in experienced gamers, especially those that play FPS games [263].

Furthermore, it needs to be stressed that, because both games require a high level of interactivity between players, participants' perceived competence is likely strongly connected to the feeling of dominance over the other player. As such, perceived competence displayed statistically significant correlations with participants' ratings of their own skill with respect to their opponent — with Spearman correlation coefficients of 0.63 for ETT ($p < 0.001$) and 0.80 for Blaston ($p < 0.001$). Considering that we did not use objective task performance measures, and pairs were matched based on the criterion of familiarity rather than comparable skill level, it is impossible to know whether the results would remain similar if participants were paired differently, or if they objectively reflect participants' skill levels. Even still, these results call for further exploration of the ways in which questions regarding previous expertise should be formulated if the goal of such questions is to predict the individual's skill for a particular VR game or application used in a user study. **Including specific questions related to gaming experience or athletic skills into the initial questionnaire may provide more useful information compared to questions pertaining to previous experience with VR technology alone.** However, further efforts toward exploring this issue would require a more strategic approach, entailing a more deliberate participant sampling process, the incorporation of objective task performance metrics, as well the elimination of confounding factors, such as interplayer involvement that comes with a multiplayer setup.

Table 7.5: Spearman correlation coefficients (presented alongside corresponding p-values) representing the degree of correlation between perceived competence reported post-game and prior levels of expertise as reported in the demographics questionnaire (taken from [12])

	ETT				**Blaston**			
	Gaming	**VR**	**Tennis**	**Shooter**	**Gaming**	**VR**	**Tennis**	**Shooter**
r_s	0.33	-0.02	0.39	0.31	0.7	0.14	0.21	0.57
p	0.06	0.91	**0.03**	0.09	**<0.001**	0.46	0.25	**<0.001**

The impact of relationship with co-player on interplayer involvement experience

We analyzed whether enjoyment of competition and desire to win during VR gaming were affected by participants' prior relationship by comparing the ratings obtained from friends to those obtained from strangers. As seen in Figure 7.10, Participants who played with friends displayed a stronger desire to win compared to the other group for both games, although the effect was only statistically significant for Blaston (U = 68.00, p = 0.02). Even though friends also seemed to enjoy competition more during Blaston, both groups were very similar in terms of enjoying competition during ETT.

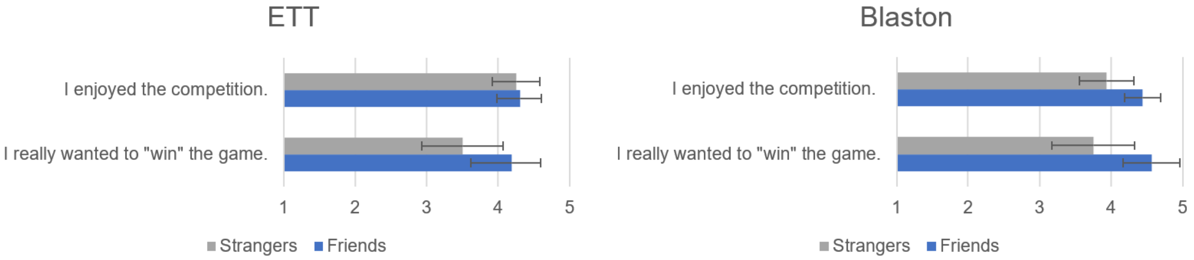


Figure 7.10: Bar charts portraying the mean score (with 95% CI) for items measuring enjoyment of competition and desire to win, as rated on a 5-point Likert scale (1-Strongly disagree, 5-Strongly agree) - Eleven Table Tennis (left) and Blaston (right); taken from [12]

Table 7.6: Mean values and corresponding standard deviations for chosen items of the Negative feelings dimension of the GEQ Social Module (adapted from [12])

		ETT				Blaston			
		Friends		Strangers		Friends		Strangers	
QoE feature	**Item**	**M**	**SD**	**M**	**SD**	**M**	**SD**	**M**	**SD**
Jealousy	I felt jealous about the other.	0.63	0.81	0.31	0.87	0.50	0.97	0.81	1.38
Revengefulness	I felt revengeful.	1.50	1.21	0.50	1.10	2.19	1.42	1.06	1.39
Malicious delight	I felt schadenfreude (malicious delight).	1.63	1.50	0.94	1.24	2.00	1.55	1.13	1.36

Negative feelings toward the opponent were also analyzed, with the focus on three specific items of the GEQ Social module, namely those measuring jealousy, revengefulness, and malicious delight. As evident from Table 7.6, there was a lot of variation in the given scores. However, there was a noticeable trend of friends exhibiting higher levels of negative feelings for the majority of examined items compared to participants playing against an unfamiliar person. These differences between scores given by friends and those given by strangers were statistically significant only for the feeling of revengefulness, with U = 62.00, p = 0.01 for ETT and U = 74.00, p = 0.04 for Blaston.

In [264], the authors report that playing with friends resulted in a stronger commitment to the in-game goals than playing with strangers in a cooperative goal structure context, but there was no difference between friends and strangers when observing competitive goals or measures of motivation in either competitive or cooperative contexts. **However, the results of Study 8**

indicate that, in a VR gaming context involving dueling games, participants may be more likely to revel in their success and feel more motivated to win or avenge their losses if they are playing with someone they have a friendly relationship with.

The impact of perceived balance of skill between co-players on VR gaming QoE

When asked to check the statements that applied to their experience, 19 out of 32 participants checked the item stating that their QoE during ETT gameplay was "dependent on the skill level of their co-player". The same was reported by 16 participants (i.e., 50% of all participants) for Blaston.

In case of ETT, 6 out of 32 participants reported (i.e., by checking the corresponding statement) they would have been more satisfied if their co-player was more skilled than they were, while 4 stated they would have been more satisfied with a less skilled co-player. In case of Blaston, 7 participants would have been more satisfied with a more-skilled co-player, while 8 would have preferred a less skilled co-player.

We further compared the overall QoE based on participants' perceived skill in comparison to their respective co-players. For ETT, very similar QoE scores were given by groups of participants who perceived themselves as less skilled (M = 4.09, SD = 0.70), equally skilled (M = 4.00, SD = 0.82), and more skilled (M = 4.14, SD = 0.67) than their co-player. The Kruskal-Wallis test found no significant differences between the three groups ($\chi^2(2) = 0.18$, p = 0.92). For Blaston, the Mann-Whitney U test (U = 86.5, p = 0.29) found no significant differences between QoE scores given by those participants who perceived themselves as less skilled (M = 4.31, SD = 0.79) and those who perceived themselves as more skilled (M = 4.64, SD = 0.50). The group of participants who perceived themselves to be equally skilled as their co-player was excluded for being too small (N = 2), but both QoE scores given by those two participants (4 and 5, respectively) did not deviate from the scores given by the other two groups. Slight differences in QoE suggest that players may be slightly more satisfied when they perceive themselves as superior to their co-player. In general, according to obtained results, **although a significant number of participants considers balance of skill to be among the factors that influence their QoE score, for the context of multiplayer user studies, it may not be necessary to pair participants based on their skill level, unless exploring related features is the primary objective of the study.**

7.6 Multiplayer VR gaming as a social activity

7.6.1 The impact of VR gaming and game genre on social interaction

Average scores for all items related to social interaction (Table 7.7) were between 3.31 and 4.31 for both games. Based on these average scores, and seeing their distribution in Figure 7.11 it can be noted that scores were generally skewed toward the games being perceived as fostering social interaction, with the highest number of given scores (for almost every item) being 4 ("Agree").

On average, ETT received higher scores compared to Blaston for all items related to social interaction except for the one regarding the feeling of doing something together, which received extremely similar scores for both games. Furthermore, the amount of participants disagreeing with the presented statements (i.e., giving out 1s and 2s) was visibly higher for Blaston. Differences between games were also statistically significant for three out of five items, as participants reported communicating with the other significantly more ($Z = -2.75$, $p = 0.006$) during ETT, and the game also received significantly higher scores for creating social bonding between participants ($Z = -2.36$, $p < 0.019$).

The shooter genre can be considered thematically more aggressive compared to table tennis, driving users away from socializing and causing them to be more performance-focused, as well as more antagonistically inclined toward their opponent. Additionally, as the more chaotic game, Blaston required users to repeatedly perform three very distinct tasks, sometimes at the same time and in no particular order — attacking the opponent, avoiding opponent's attacks and switching weapons (which also required participants to turn away from their opponent). Thus, participants could have been too distracted for conversation. Conversely, a more turn-based nature of table tennis, with each player periodically switching between proactive, passive, and reactive behaviour as the ball moves back and forth between players, likely discouraged the user from moving their attention away from the other player, thus creating playing conditions that foster contact between participants. Moreover, it is worth reiterating that Blaston is a game with very prominent sound effects and background music, while ETT is fairly quiet, which likely also influenced participants' willingness to communicate. **Therefore, it could be theorized that considering aspects such as genre, mechanics, and visual and auditory aspects of VR games may provide useful context for the interpretation of studies examining social interaction during gameplay.**

7.6.2 The impact of relationship with co-player on social interaction

Despite the expectation that participants who played with friends would give higher scores for the items related to social interaction, this was not necessarily the case (Table 7.8). In case of Blaston, friends did seem to give slightly higher scores to all social interaction items compared

Table 7.7: Mean values and corresponding standard deviations for items pertaining to social interaction and related dimensions (adapted from [12])

		ETT		**Blaston**	
QoE feature	**Item**	**M**	**SD**	**M**	**SD**
Communication	I communicated with the other participant.	4.31	0.74	3.72	1.05
Co-presence	I felt a strong sense of co-presence, as if we were in the same room.	3.88	0.91	3.78	1.16
Social bonding	The game created some sort of social bonding between me and the other player.	3.69	0.82	3.31	1.18
Togetherness	I had the feeling we were doing something together.	4.06	0.88	4.06	0.88

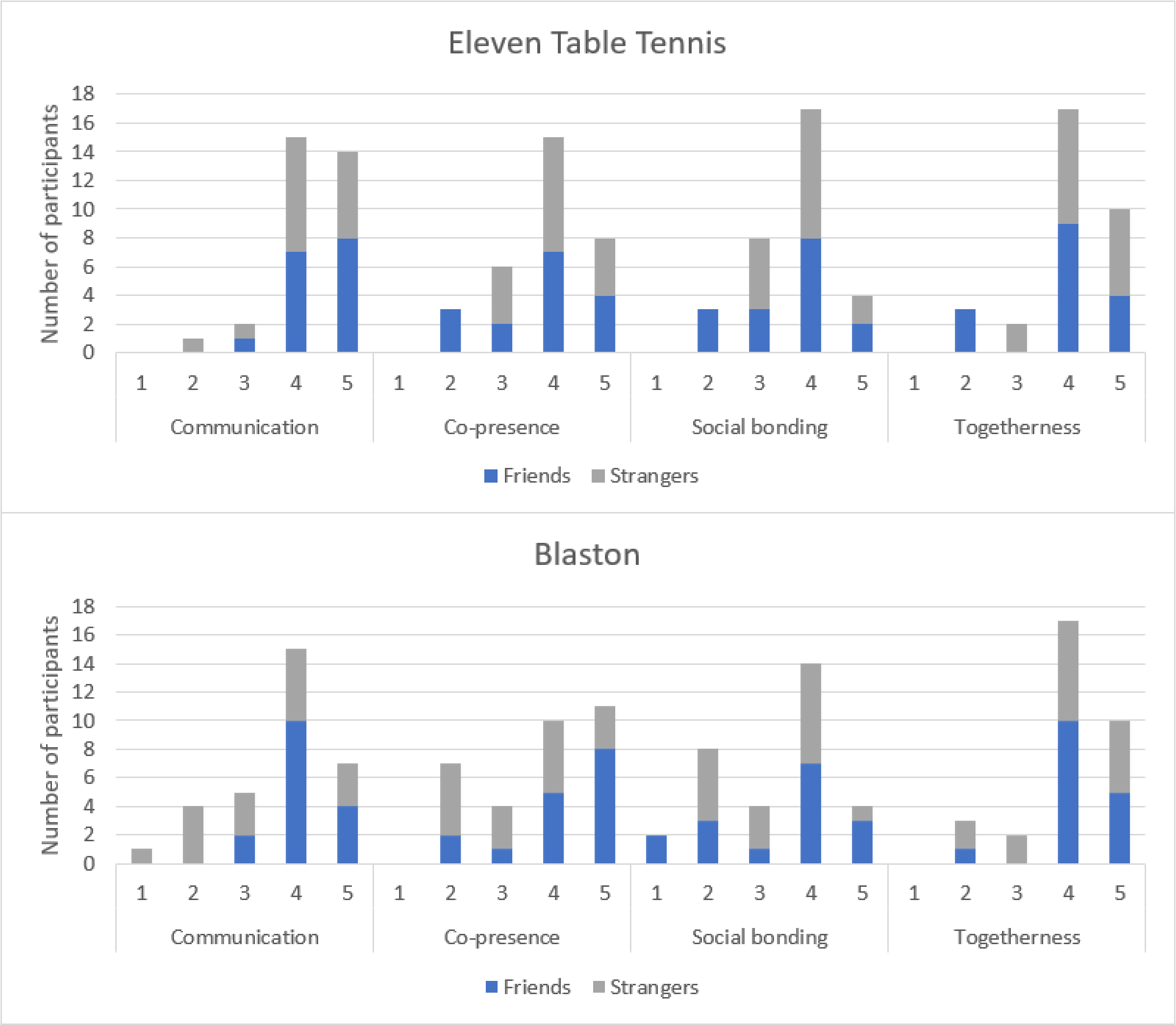


Figure 7.11: Stacked bar charts portraying the number of participants rating different aspects of VR gaming as a social experience on a 5-point Likert scale (1-Strongly disagree, 5-Strongly agree) - Eleven Table Tennis (top) and Blaston (bottom); taken from [12]

to strangers, however, the difference in co-presence scores (U = 76.00, p = 0.05) was the only one that was almost statistically significant when comparing the two groups. Surprisingly, in case of ETT, average scores for items pertaining to co-presence, social bonding, and togetherness were actually higher for the strangers group. In case of ETT, no statistically significant differences were found between groups for either item. Although these findings may seem counter-intuitive, previous research by Steed et al. [250] did find that groups of strangers and friends were surprisingly similar in terms of collaboration, behavior and enjoyment during an extended amount of time (210+ minutes) spent in a virtual environment. With that in mind, it is not completely unreasonable to assume that similar results could also be obtained in a competitive context, though shorter in duration. Furthermore, the presence of the study administrator and the formal context of a laboratory study could have discouraged friends from exhibiting the level of comfortable social behaviour they might have exhibited in a more private situation.

7.6.3 The impact of relationship with co-player on VR gaming QoE

Comparing the results from friends and strangers, it was found that, on average, overall QoE ratings as reported in the *post-game questionnaire* following Blaston were slightly higher for friends (M = 4.56, SD = 0.51) compared to strangers (M = 4.38, SD = 0.81). However, the opposite was true for ETT, as strangers (M = 4.25, SD = 0.68) rated their experience with the table tennis game higher than friends did (M = 3.94, SD = 0.68). Nevertheless, differences between friends and strangers were not statistically significant for either Blaston (U = 116.5, p = 0.67) or ETT (U = 97, p = 0.25).

While obtained results did not indicate a significant effect of prior relationship on reported QoE scores, participants were also provided with an option to confirm whether they agreed with either of the several statements pertaining to the relationship between their gaming QoE and the level of familiarity with their opponent. For both ETT and Blaston, only one participant who played with a stranger stated that they would prefer not to play that game with their opponent again. For ETT, two participants from the strangers group, as well as two participants from the friends group, stated that they would prefer to play the game in single-player mode. For

Table 7.8: Mean values and corresponding standard deviations for items pertaining to social interaction (adapted from [12])

	ETT				**Blaston**			
	Friends		**Strangers**		**Friends**		**Strangers**	
QoE feature	**M**	**SD**	**M**	**SD**	**M**	**SD**	**M**	**SD**
Communication	4.44	0.63	4.19	0.83	4.13	0.62	3.31	1.25
Co-presence	3.75	1.06	4.00	0.73	4.19	1.05	3.38	1.15
Social bonding	3.56	0.96	3.81	0.66	3.38	1.36	3.25	1.00
Togetherness	3.88	1.02	4.25	0.68	4.19	0.75	3.94	1.00

Blaston, none of the participants from the friends group would have preferred to play a single-player version of the game, but 25% of those playing with strangers would. Moreover, in case of ETT, almost two-thirds of participants from the stranger group reported that they thought their overall QoE would be significantly improved if they had a closer relationship with the other participant. The same was reported by just over a third of participants from the stranger group following Blaston.

In the context of study methodology design and participant inclusion criteria, the results indicate that considering prior relationships when choosing and matchmaking participants for QoE studies is not an absolute necessity. However, a significant percentage of participants did seem to favor a closer relationship with their opponents and were of the belief that playing against a more familiar opponent would be reflected in their QoE score, which calls for further research into this issue. It is also worth reiterating that obtained results pertaining to competitive feelings have suggested a significant impact of prior relationships, further confirming that certain study objectives may benefit from taking this factor into account.

7.7 Key takeaways

This chapter presents the results of two preliminary studies (Study 6 and Study 7) focused on the impact of network latency on multiplayer gaming QoE and a more comprehensive study (Study 8) exploring user experiences with multiplayer VR gaming from multiple angles. After reaching inconclusive results in Study 6, the initial methodology was altered, leading to more significant conclusions in Study 7. Based on this discrepency between studies, several aspects were identified as potential confounding factors: a level of challenge that was not necessarily in proportion with the skills of the participants, the implementation of interaction mechanics (i.e., the implementation of the shoot mechanics for certain weapons), and social context (i.e., the presence of another player, levels of familiarity and social interaction between co-players).

Building off of the findings explored in preliminary studies, Study 8 investigated the impact of network factors (i.e., different access networks and levels of latency) on QoE, explored whether participants engaged in social interaction and bonding during gameplay, and focused on issues such as expertise, skills, and competitive feelings in the context of VR gaming as an interplayer competitive activity. The results indicate that playing on a broadband cellular network (4G or 5G) — despite its slight latency — was not significantly different from playing on a fiber optic connection (Ethernet). Even slowing down the RTT between players by as much as 100-200 ms may be tolerated by users, however, the impact of this level of degradation is dependent on the game, especially with regard to the implementation of in-game physics and lag compensation mechanisms.

Even in the artificial context of VR gaming as a task in a laboratory study, the experience of each participant was (at least in part) still being influenced by factors such as social context. However, while participants did report communicating and bonding with their opponent and experiencing levels of co-presence and togetherness, the extent of experiencing these aspects was dependent on the game, with potential culprits such as genre, pace, objective, and sound effects left to be explored in future studies. Despite initial expectations, playing with a friend, as opposed to a stranger, did not seem to be a significant factor affecting the amount of socializing during a study session, nor did it significantly impact overall QoE. While the influence of existing relationships between players likely plays a larger role in a more realistic context, the results indicate that when recruiting participants for VR multiplayer gaming research, considering their prior relationship may not be necessary. However, when factoring in the opinions of the participants in this study, further consideration of this issue is advisable.

Moreover, as indicated by obtained results, if the goal of the study is to investigate constructs such as combativeness, revengefulness, or desire to win, pairing participants with a familiar opponent may produce more significant results compared to pairing them with a stranger. Focusing on participants' skills and skill-balancing in the context of participant recruitment, it appears that considering expertise with a specific genre or comparable real-life activities may be more useful in predicting participants' eventual competence during VR gameplay (as opposed to relying solely on expertise with VR technology). While it would be advisable to perform skill-based matchmaking of participants in future VR multiplayer studies involving highly interactive competitive games, conducted analysis did not confirm a notable influence of this factor on the overall QoE. Nevertheless, it is not possible to draw generalized conclusions regarding this issue, as participant recruitment and matchmaking criteria did not control for participants' previous experience or balance of skill, and objective measures of performance were not employed in this study.

However, while taking such a divergent research approach in this study may not provide the most definite answers regarding each individual factor and its impact on user experience, these results highlight the variety of factors that need to be considered in future studies. It is important to keep in mind that real-life experiences with VR gaming are diverse and multifactorial, which further complicates the methodology design of user studies in the field. Conducting studies that cover a broader spectrum of research questions as a way to identify potentially confounding factors may therefore be of benefit to the research community. In light of this, more divergent studies such as this one (which shed light on issues that call for further research efforts) may further inform and complement more focused studies (which explore specific issues in depth) and vice versa, with both approaches potentially yielding results that further the current understanding of user experience with such a complex and multilayered service.

Table 7.9: The impact of chosen IFs on QoE features pertaining to the following QoE constituents (based on the proposed VR_QOE_LLM_3 model): perceived quality of networking (PQN), perceived social interaction (PSI), interplayer involvement experience (IIX), and player experience (PX)

		QoE IF					
		Network latency		**Relationship with co-player**		**Level of expertise**	
	QoE feature	ETT	Blaston	ETT	Blaston	ETT	Blaston
PQN	Perceived delay						
	Perceived co-player delay	X					
	Input responsiveness	X					
	Input smoothness	X	X				
	Perceived performance degradation	X	X				
PSI	Communication						
	Co-presence				X		
	Social bonding						
	Togetherness						
IIX	Enjoyment of competition						
	Desire to win				X		
	Jealousy						
	Revengefulness			X	X		
	Malicious delight						
PX	Competence					X	X

X denotes QoE features significantly affected by chosen IFs

7.8 Modeling the impact of system, context and player IFs on chosen QoE features

In addition to providing general insights into the multiplayer experience, as well as presenting findings that may inform the methodology design of future studies, results obtained in Study 8 can be utilized as input for QoE modeling. This process builds on the last of the three proposed low-level models, the VR_QOE_LLM_3 model, which describes the hypothesized relationships between chosen system, context, and player IFs and different features pertaining to QoE constituents characterizing multiplayer gaming experiences.

Table 7.9 provides the matrix of significant findings pertaining to the proposed model and based on the results of Study 8. In cases where statistical tests (presented in the Results section of this chapter) indicated statistically significant differences in QoE feature scores with respect to the values of the observed (relationship with co-player) or manipulated (network latency) IFs, the corresponding cell of the Table is marked by an X symbol. The same symbol is used to indicate statistically significant correlations between level of expertise and competence (to maintain visual clarity, the simplified tabular preview of this relationship does not distinguish between different types of expertise). These significant relationships are presented in Figure 7.12, which illustrates the adjusted low-level model of VR gaming QoE (VR_QOE_LLM_3SR).

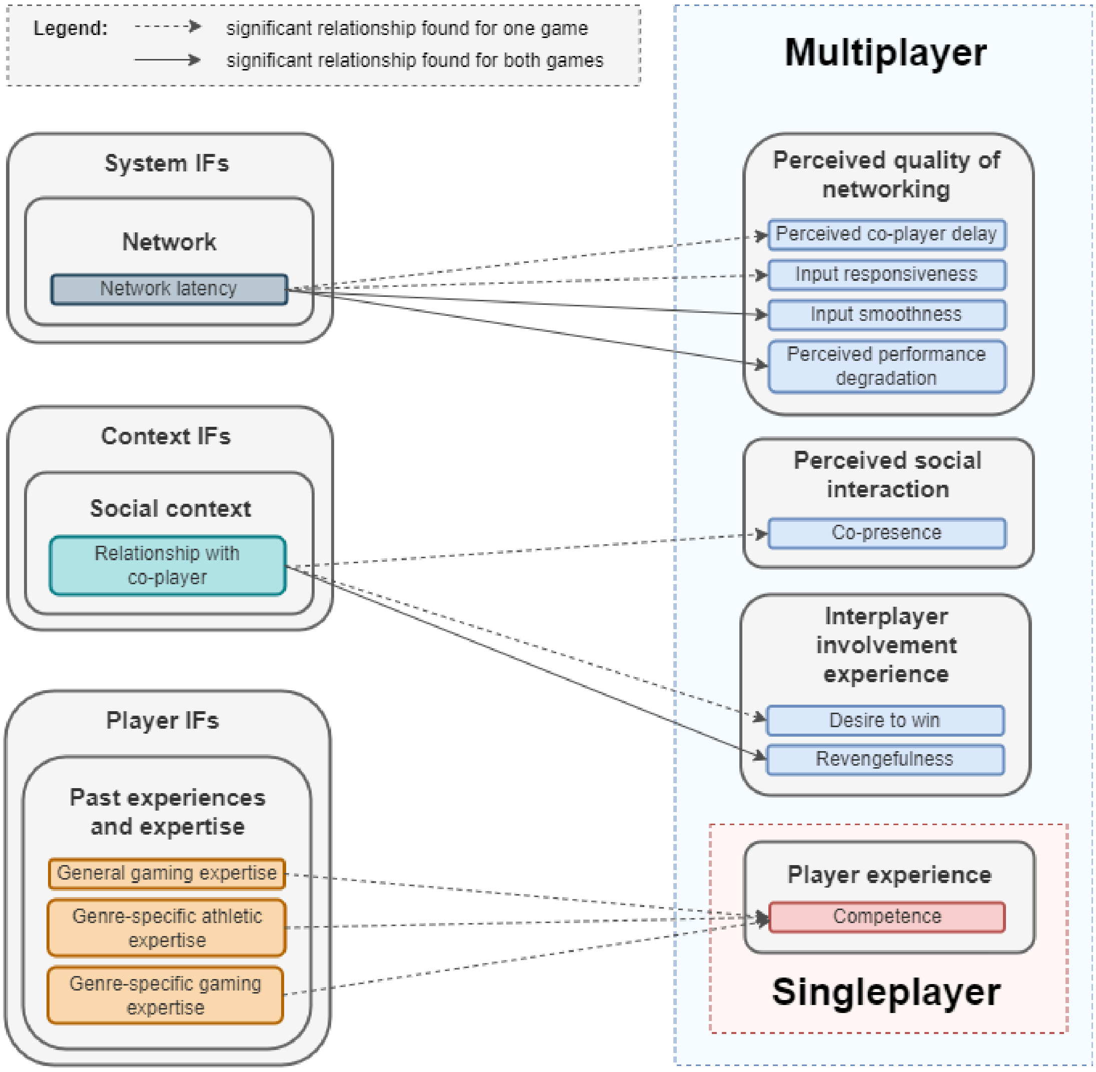


Figure 7.12: Partial low-level model (VR_QOE_LLM_3SR) of QoE for VR gaming depicting the significant relationships between network factors, social context, and past experiences/expertise, and chosen QoE features for a competitive multiplayer scenario (based on the proposed VR_QOE_LLM_3 model)

7.9 Chapter summary

Unlike previous chapters, this chapter focuses on multiplayer VR. After presenting the results of preliminary work, which highlights the complexities of conducting user studies focused on multiplayer VR gaming, the chapter goes on to discuss the results of a more comprehensive study addressing the impact of player, system, and context IFs on different QoE features, especially those pertaining to identified constituents of the proposed multiplayer sub-model of VR gaming QoE. Reported findings are presented with special consideration of the process of designing and conducting a multiplayer user study, offering useful insights pertaining to participant recruitment and methodology design, such as the choice of access network to be used for testing, and the issue of matchmaking study participants based on skill and prior relationship. Furthermore, the proposed low-level model of VR gaming QoE in multiplayer scenarios is reconsidered based on the obtained results, and a new (adjusted) model is proposed.

Chapter 8

Conclusions and future work

This chapter provides an overview of relevant findings and contributions of this thesis. The chapter is divided into two parts. The first part provides a structured summary of conclusions. The second part addresses the limitations of the thesis, while also providing suggestions for future work.

8.1 Conclusions

The summary of concluding remarks is presented in the form of answers to main research questions listed in Chapter 1.

RQ1: Which QoE influence factors should be considered as potential contributors to the overall QoE for the VR gaming use case?

The pool of potential influence factors impacting QoE for VR gaming is incredibly broad. As discussed in detail in Chapters 2 and 3, many of these factors have previously been identified in research articles and industry recommendations, and are often being grouped as either context-related, user-related (player-related), or system-related. Building on existing studies, an overview of relevant IFs was illustrated in the schematic diagram presented in Chapter 3, and further consideration of history of illness, injury, mobility limitations, and current state of the player as relevant IF categories was proposed. Chapters 5, 6, and 7 presented the results of several user studies exploring the impact of chosen IFs on the overall QoE and the selected features of the VR gaming experience. Some IFs were shown to have a definite impact on the overall QoE score, such as different implementation parameters of tested interaction mechanics and network latency. Certain factors (e.g., game genre/mechanics, social context) were identified as having a significant impact on selected QoE features, and should therefore also be considered as potential contributors to the overall QoE, either directly, or through mediating QoE features.

RQ2: Which QoE features and QoE constituents should be considered as potential contributors to the overall QoE for the VR gaming use case?

The proposed high-level model of VR gaming QoE (VR_QOE_HLM), presented in Chapter 3, is comprised of two sub-models. The singleplayer sub-model includes player experience, workload, and VRISE as highlighted constituents. The multiplayer sub-model extends the singleplayer sub-model by adding perceived quality of networking, perceived social interaction, and interplayer involvement experience as relevant constituents of the gaming experience. Each of the identified constituents can be further dissected into individual QoE features, which were proposed in the mid-level model of VR gaming QoE (VR_QOE_MLM).

RQ3: How can we classify interaction mechanics commonly found in VR games?

The proposed taxonomy of interaction mechanics for VR gaming was presented in Chapter 4. This taxonomy categorizes IMs based on interaction fidelity, hand usage, target characteristics, tool mediation, projectile use, and interaction space. It is important to note that the goal of this taxonomy was not to divide interaction mechanics into separate categories, as each IM likely belongs to several. Instead, the goal was to enhance the understanding of various aspects of each IM, and propose a systematic framework to facilitate their description.

RQ4: What is the impact of VR game mechanics on workload?

Different games, belonging to distinct genres with diverse game mechanics, were shown to vary significantly with regard to their impact on measured dimensions of workload. As presented in Chapter 5, the games used in our study all produced a similar, somewhat high, level of mental demand, and the majority of them have also produced substantial levels of physical and temporal demand. With respect to other measured dimensions of workload, their reported intensities varied between games, but were generally less pronounced. Comparing Study 1 and Study 2, games that shared primary interaction mechanics were not necessarily similar with respect to their effect on workload, which may be attributed to different participant samples, specific implementations of primary game mechanics, or the addition of secondary game mechanics. In fact, Studies 4, 5, and 6 (Chapter 6) showed that even within the constraints of the same general primary IM and in the absence of secondary game mechanics, measured workload may differ significantly based on the specific implementation details of the tested IM.

RQ5: What is the impact of VR game mechanics on subjective measures of discomfort and physiological symptoms?

Games with different game mechanics were shown to produce significantly different levels of discomfort (Chapter 5). With regard to musculoskeletal pain and fatigue, VR gaming generally had the most significant effect on the arms due to the specifics of presented primary IMs. Even though the same exact hardware was used for all games, the extent of HMD-related discomfort was shown to vary based on the game, even more than other types of workload, likely due to games differing with regard to required neck positions and movements. Participants were also bothered by perceived HMD temperature which may be affected by hardware overheating, but also participants experiencing thermal discomfort due to physical activity. While participants reported some changes in physiological symptoms that could indicate cybersickness, sweating appeared to be the most pronounced. However, it can be argued that the common approach of interpreting this symptom as a symptom of cybersickness may not be justified, especially for games that omit common cybersickness triggers, such as vection. Instead, the more likely explanation for this symptom is the physically intensive nature of VR gaming.

RQ6: What is the impact of VR game mechanics on cognitive effects?

Previous research [193] has indicated that the cognitive effects of VR use may be negligible, except for its notable negative impact on reaction time. In line with this, the simple reaction time test was chosen as the measure for evaluating the cognitive effects of VR gaming in Study 1 and Study 2 (Chapter 5). Overall, the majority of participants experienced an increase in reaction time following VR use. However, shooters and physically active VR games have actually resulted in a decreased post-VR SRT for some participants. Moreover, in both studies, only games with pick-and-place mechanics produced a statistically significant increase in SRT. While several possible theories behind SRT increases that occur after pick-and-place games were considered in Chapter 5, a definite explanation for this phenomenon was not proposed. However, it should be noted that, contrary to sources [191, 193] that identified a possible connection between cybersickness and VR-induced increases in SRT, these changes in SRT seemed to correspond more significantly with the measures of workload rather than cybersickness.

RQ7: How can we evaluate QoE for different configurations of VR interaction mechanics?

While using commercial games as test material is an established practice in gaming research and comparing user experience with games of different mechanics may provide relevant insights, commercial games are essentially black boxes from the perspective of QoE researchers.

It may therefore be difficult — or completely impossible — not only to manipulate the observed parameters of the game, but also to collect and export a comprehensive overview of detailed task performance metrics. Depending on the study objective, this may or may not be an issue. The particular objective of exploring and designing QoE-driven interaction mechanics requires a more flexible approach to manipulating, collecting and exporting data. Thus, Chapter 6 proposes a general methodology for the evaluation of IM quality, identifying potential influence factors and possible subjective and objective measures, in addition to presenting a novel set of guidelines (the INTERACT framework) for the implementation of highly customizable test platforms to be used specifically for this purpose.

RQ8: What is the impact of different configurations of interaction mechanics on task performance in VR?

The results of Study 3, Study 4, and Study 5 identified multiple IM parameters that may influence participants' task performance to a significant degree. For the slash IM, significant differences in accuracy were recorded between different target spawn angles and different amounts of force required to destroy a target object. For pick-and-place, the number of successful participants substantially differed based on the parameter values for puzzle scale, collider scale, remote grab, and scale offset. The overall duration of time necessary to successfully complete the puzzle also varied based on tested configurations. For the shoot IM substantial differences in accuracy were identified for different configurations of visual shooting aids. Differences in target spawn angles and shoot forces were also reflected in variations in shoot accuracy combined with varying numbers of expelled projectiles.

RQ9: What is the impact of different configurations of interaction mechanics on VR QoE and QoE features?

For each of the IMs (slash, pick-and-place, shoot) explored in Chapter 6, the overall QoE score was shown to fluctuate based on the tested configuration. Different parameters were identified as significant influence factors impacting QoE features pertaining to player experience, workload, and — to a lesser degree — VRISE. The overall conclusion was that there was no one-size-fits-all approach to implementing any of the test mechanics; however, in a significant number of cases, participants were inclined toward less demanding scenarios. This was not universal to all parameters, though, which underlines both the relevance and the complexity of analyzing workload as a QoE constituent. In line with this, a preliminary classification of IM parameters was proposed based on the directional relationship between perceived workload and the overall (subjective) quality. Relevant findings obtained in Study 3, Study 4, and Study

5 were summarized in a set of proposed guidelines for the implementation of VR interaction mechanics.

RQ10: What is the impact of previous experience on perceived competence during VR gaming?

Considering that previous experiences and expertise are considered to be relevant influence factors impacting user experiences with multimedia applications, inquiries regarding previous experiences with gaming in general, as well as with specific games, genres, and platforms, are often made in gaming user studies. Likewise, researchers focused on user experiences with VR often collect information regarding participants' levels of expertise with VR use.

Researchers exploring QoE of VR gaming may benefit from both of the aforementioned approaches to assessing participants' levels of relevant experience/expertise; however, there is an additional factor to consider. With the prevalence of manual, isomorphic, and controller-based interaction mechanics in VR services, designing games that are centered around emulating interactions from the "physical world" is a common practice among content developers. Thus, it is reasonable to assume that certain life skills could be transferable to the VR context, especially with regard to sports games. In Study 8 (Chapter 7), we found that previous expertise with table tennis appeared to have a significant correlation with perceived competence after playing a VR table tennis game. For the tested VR shooter game, post-VR measure of perceived competence correlated significantly with both general gaming expertise and expertise with FPS games. Previous expertise with the use of the VR platform did not have a significant effect for either game. However, it is important to reiterate that, while our findings provide the motivation for further research into this issue, they can only be considered preliminary, as relevant factors were not sufficiently controlled, and the conclusions were not grounded in objective task performance metrics.

RQ11: What is the impact of network quality on QoE and its features for multiplayer VR gaming?

Whereas differences in QoE scores for different access networks (4G, 5G, Ethernet) were not found statistically significant, the impact of added network latency on QoE was more evident (Chapter 7). However, differences between scenarios with different levels of latency reached statistical significance only for the tested table tennis game, as opposed to the shooter. This disparity between games may be attributed to different factors. The shooter game's apparent robustness to added latency can be attributed to the implementation of lag compensation techniques, its fast-paced gameplay, hyperrealistic adjustments to projectile behavior (particularly regarding projectile speed), and unidirectional projectile traversal. This is in contrast to the more sensitive table tennis game, in which network degradations were more noticeable as they interfered with the continuous motion of the shared object (ball), which was also subject to the more realistic implementation of in-game physics, possibly making any deviations from the expected behaviour more noticeable.

Significant levels of correlation were found between QoE scores and individual QoE features pertaining to the perceived quality of networking. Scenarios with different levels of added latency yielded significant differences in input smoothness scores and perceived performance degradation scores for both games, while additionally producing significant differences in perceived co-player delay and input responsiveness for the table tennis game. Presented findings indicate that, in VR, realistic sports games simulators may be more sensitive to latency compared to first person shooters.

RQ12: What is the impact of social context on QoE and its features for multiplayer VR gaming?

Participants in Study 8 (Chapter 7) were grouped in pairs based on their relationship with their co-player. Half of all participants were paired with a close friend, while the other half played against a complete stranger. While we expected this factor would strongly impact the levels of social interaction during gameplay, with the friends group scoring higher on all items, only a mild effect of this nature was noted for the shooter game, with co-presence being the only feature that showed a statistically significant difference between groups. The results for the table tennis game deviated even more from the initial expectations, as strangers, on average, reported experiencing higher levels of co-presence, social bonding and togetherness compared to friends, as well as reporting higher scores of the overall QoE.

The impact of social context on features pertaining to the interplayer involvement experience constituent has also been explored. Friends were shown to be more competitive, reporting a significantly stronger desire to win for the shooter game, while reporting significantly higher revengefulness for both games. While the reasons behind the impact of relationship between co-players on measured features were not fully identified, collected data indicates that factoring prior relationships between participants into the methodology design of multiplayer user studies may not always be necessary, although it could provide relevant context for certain user studies.

8.2 Limitations and Future Work

After listing the conclusions of this thesis, it is also necessary to reflect on its limitations. With regard to participant recruitment, all studies were limited in terms of sample size. A larger sample of participants would likely lead to more reliable and generalized insights into the user experience, and possibly provide the input for the development of statistical models of VR QoE. In addition to sample sizes, it is important to note the demographics of recruited participants. Even though opting for the homogeneous sampling approach was a deliberate decision made for Studies 1-5, including underrepresented demographics could have provided highly relevant insights into their preferences regarding different interaction mechanics, along with their experiences of VRISE and workload, which may have deviated from the findings obtained for the used samples. Moreover, the majority of participants were inexperienced with VR technology. Subject to novelty effects, the results of studies focused on novice VR users may not provide an accurate reflection of the behaviours, opinions, and preferences of regular VR gamers.

As discussed in Vlahovic et al. [30], the commercial titles used as test material in the conducted studies represent only a small fraction of VR games. There are many other genres to explore, as well as a variety of games that incorporate different tasks, interaction mechanics, and methods of in-game locomotion. Other potentially relevant aspects, such as aesthetics, and narrative, were not explored in this thesis. It is also important to note that the custom application used as test material in Studies 3-5 provides numerous additional parameters and parameter values, and collects a broad range of measures that have not been explored in this thesis. The inclusion of these additional parameters and measures into our methodology may provide useful insights into the user experience and will be considered in future work.

With regard to Study 8, it is necessary to reiterate its exploratory nature. Divergent in terms of study objectives and used measures, this study was designed to provide preliminary findings to be further explored in future studies. Specific impacts of examined factors should be addressed in a more strategic manner.

Another issue to be noted is the large number of questionnaires and individual items participants were asked to complete for the purpose of conducted studies. Using a single, validated, standardized questionnaire which examines multiple dimensions of user experience may be preferable to the approach taken in this thesis, especially with regard to participant fatigue. Unfortunately, while there are questionnaires that encompass multiple aspects of the user experience with VR applications, e.g., [107], as well as those that are geared more toward VR games, e.g., [106], the ones we considered did not incorporate a broad spectrum of features that was evaluated in this research.

With all this in mind, it is worth emphasizing that the objective of this research was to identify and shed light on understudied aspects of VR gaming QoE, such as certain types of VRISE, specific implementations of game-specific mechanics, and the implications of networked multiplayer setups on study results. Additional studies approaching the explored topics are necessary to corroborate the findings presented in this thesis.

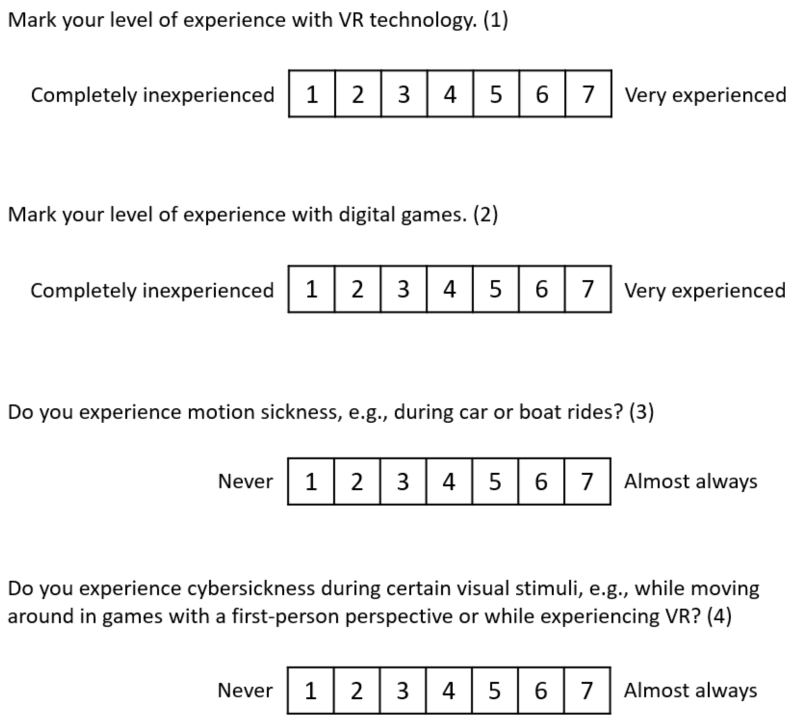
With regard to future work, one approach would be to address the identified limitations of this thesis. Firstly, valuable insights could be gained by conducting studies on larger samples, underrepresented demographics and more experienced VR users. Choosing different games as test material, or focusing on different parameters of the used material, could extend our current understanding of the discussed topics. For example, while it is not possible to draw definite conclusions regarding the appropriateness of different metrics of expertise based solely on the results of Study 8, they provide the incentive for more systematic research on the topic, perhaps also in the context of participant matchmaking.

Furthermore, considering that current measures of certain features may be lacking when it comes to addressing a diverse range of VR applications, the findings obtained in the scope of this research may be used to adapt or extend the commonly used subjective measures of different QoE features to include a broader range of items, perhaps those pertaining to overlooked VRISE. If appropriately validated, these newly developed tools could address the need for standardized measures of the VR gaming experience. Proposed models of VR gaming QoE could also be further developed, as additional research is necessary to validate the higher-level models and relationships that were not explored in the scope of presented research.

Lastly, over the course of this thesis, several themes were consistently underlined in one way or another: the multimodal interactivity, immersivity and obtrusiveness of the VR platform, the diversity of VR games with respect to game mechanics, and the diversity of user experiences and preferences with different games, especially when considering VRISE as a potential hazard to consumers, and workload as a complex ideal-point feature. These themes provided the initial inspiration behind the concept of the so-called VR **C**omfort **A**ccessibility and **S**afety (VR-CAS) rating system as an aid in the categorization of VR content according to the needs, preferences and limitations of individual users. The call for a joint industry and academic initiative to realize this concept has already been proposed in the recent conference paper by Vlahovic et al. [206], and further efforts in this direction are already underway.

Appendix A: Scales reported in Chapter 5

Pre-study questionnaire



Mark your level of experience with VR technology. (1)

Completely inexperienced	1	2	3	4	5	6	7	Very experienced

Mark your level of experience with digital games. (2)

Completely inexperienced	1	2	3	4	5	6	7	Very experienced

Do you experience motion sickness, e.g., during car or boat rides? (3)

Never	1	2	3	4	5	6	7	Almost always

Do you experience cybersickness during certain visual stimuli, e.g., while moving around in games with a first-person perspective or while experiencing VR? (4)

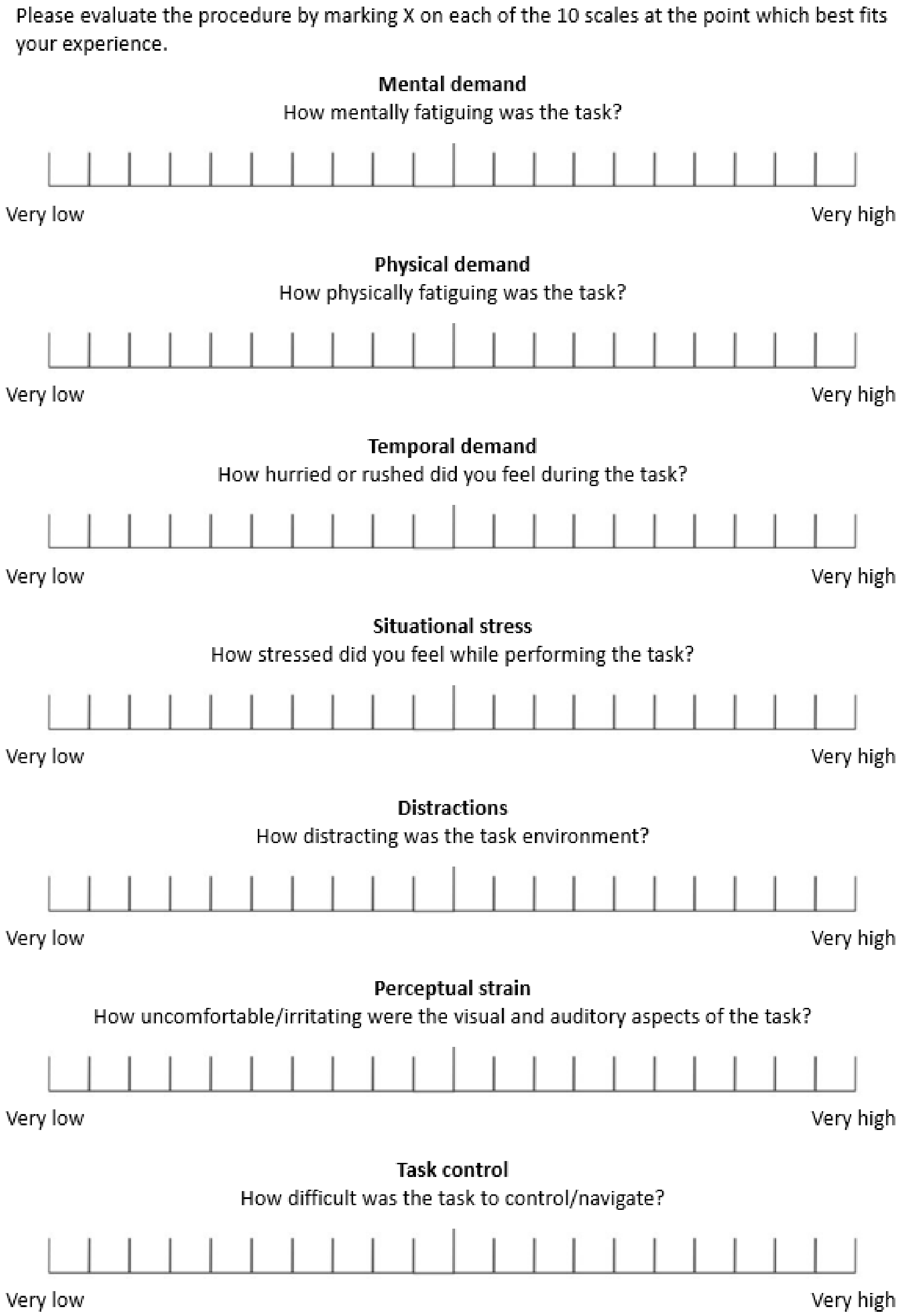
Never	1	2	3	4	5	6	7	Almost always

Based on the provided ratings, participants were categorized as follows.

Level of experience — items (1) and (2):
1 - Inexperienced
2, 3 - Beginner
4, 5 - Intermediate
6, 7 - Expert

Propensity toward motion sickness/cybersickness — items (3) and (4):
1 - None
2, 3 - Mild
4, 5 - Medium
6, 7 - Strong

Workload (Simulator Task Load Index — SIM-TLX [160])



Please evaluate the procedure by marking X on each of the 10 scales at the point which best fits your experience.

Mental demand

How mentally fatiguing was the task?

Very low Very high

Physical demand

How physically fatiguing was the task?

Very low Very high

Temporal demand

How hurried or rushed did you feel during the task?

Very low Very high

Situational stress

How stressed did you feel while performing the task?

Very low Very high

Distractions

How distracting was the task environment?

Very low Very high

Perceptual strain

How uncomfortable/irritating were the visual and auditory aspects of the task?

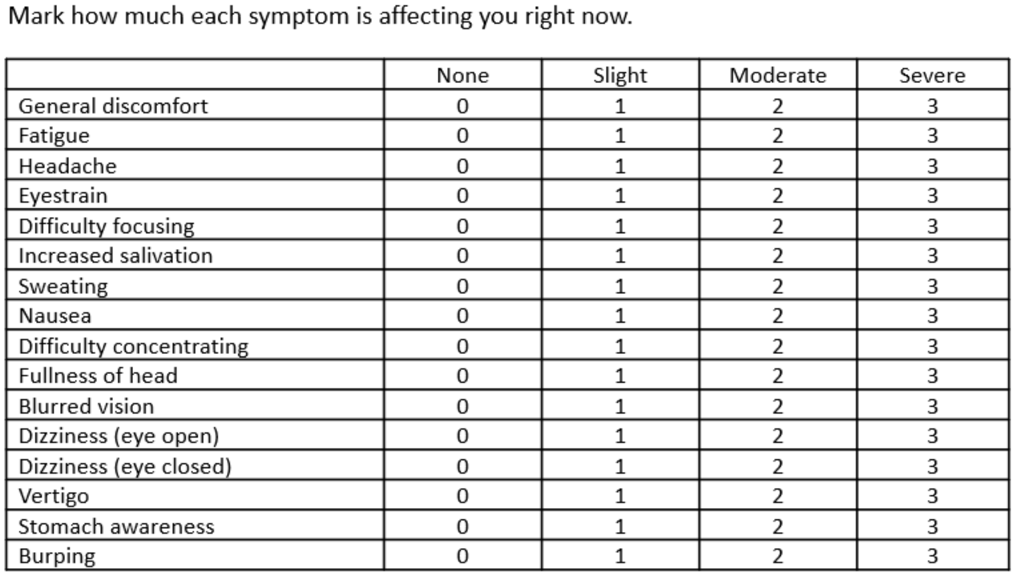
Very low Very high

Task control

How difficult was the task to control/navigate?

Very low Very high

Cybersickness (Simulator Sickness Questionnaire — SSQ [18])



Mark how much each symptom is affecting you right now.

	None	Slight	Moderate	Severe
General discomfort	0	1	2	3
Fatigue	0	1	2	3
Headache	0	1	2	3
Eyestrain	0	1	2	3
Difficulty focusing	0	1	2	3
Increased salivation	0	1	2	3
Sweating	0	1	2	3
Nausea	0	1	2	3
Difficulty concentrating	0	1	2	3
Fullness of head	0	1	2	3
Blurred vision	0	1	2	3
Dizziness (eye open)	0	1	2	3
Dizziness (eye closed)	0	1	2	3
Vertigo	0	1	2	3
Stomach awareness	0	1	2	3
Burping	0	1	2	3

Pain and muscle fatigue (using the Borg CR-10 scale [199])

What is the level of your arm pain?

0	Nothing at all
1	Very weak
2	Weak
3	Moderate
4	Somewhat strong
5	Strong
6	
7	Very strong
8	
9	
10	Extremely strong

What is the level of your arm fatigue?

0	Nothing at all
1	Very weak
2	Weak
3	Moderate
4	Somewhat strong
5	Strong
6	
7	Very strong
8	
9	
10	Extremely strong

What is the level of your neck pain?

0	Nothing at all
1	Very weak
2	Weak
3	Moderate
4	Somewhat strong
5	Strong
6	
7	Very strong
8	
9	
10	Extremely strong

What is the level of your neck fatigue?

0	Nothing at all
1	Very weak
2	Weak
3	Moderate
4	Somewhat strong
5	Strong
6	
7	Very strong
8	
9	
10	Extremely strong

What is the level of your upper back pain?

0	Nothing at all
1	Very weak
2	Weak
3	Moderate
4	Somewhat strong
5	Strong
6	
7	Very strong
8	
9	
10	Extremely strong

What is the level of your upper back fatigue?

0	Nothing at all
1	Very weak
2	Weak
3	Moderate
4	Somewhat strong
5	Strong
6	
7	Very strong
8	
9	
10	Extremely strong

What is the level of your lower back pain?

0	Nothing at all
1	Very weak
2	Weak
3	Moderate
4	Somewhat strong
5	Strong
6	
7	Very strong
8	
9	
10	Extremely strong

What is the level of your lower back fatigue?

0	Nothing at all
1	Very weak
2	Weak
3	Moderate
4	Somewhat strong
5	Strong
6	
7	Very strong
8	
9	
10	Extremely strong

HMD discomfort (using the Borg CR-10 scale [199])

What was your level of discomfort caused by VR HMD weight?

0	Nothing at all
1	Very weak
2	Weak
3	Moderate
4	Somewhat strong
5	Strong
6	
7	Very strong
8	
9	
10	Extremely strong

What was your level of discomfort caused by HMD fit in terms of tightness?

0	Nothing at all
1	Very weak
2	Weak
3	Moderate
4	Somewhat strong
5	Strong
6	
7	Very strong
8	
9	
10	Extremely strong

What was your level of discomfort caused by HMD fit in terms of looseness?

0	Nothing at all
1	Very weak
2	Weak
3	Moderate
4	Somewhat strong
5	Strong
6	
7	Very strong
8	
9	
10	Extremely strong

What was your level of discomfort caused by VR HMD temperature?

0	Nothing at all
1	Very weak
2	Weak
3	Moderate
4	Somewhat strong
5	Strong
6	
7	Very strong
8	
9	
10	Extremely strong

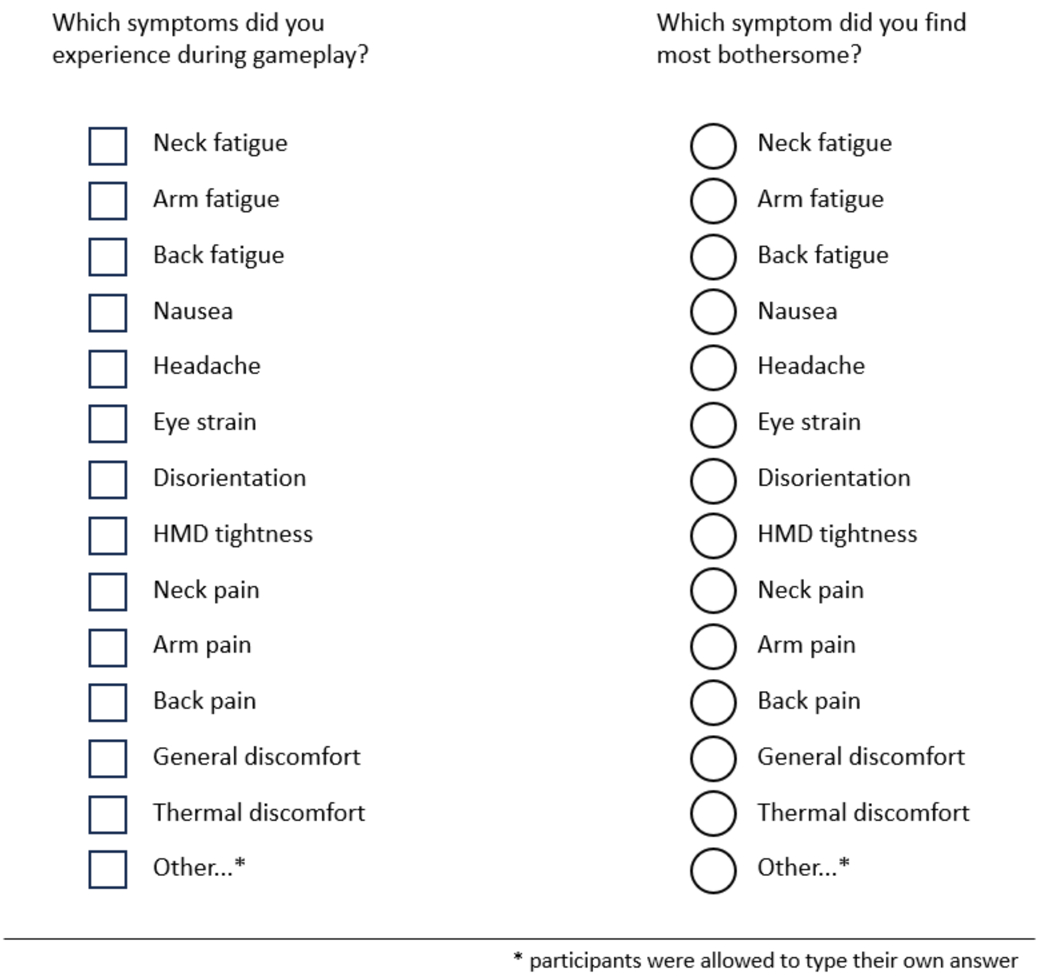
What was your level of discomfort caused by VR HMD display quality (e.g., resolution, etc.)?

0	Nothing at all
1	Very weak
2	Weak
3	Moderate
4	Somewhat strong
5	Strong
6	
7	Very strong
8	
9	
10	Extremely strong

What was your level of annoyance with the HMD cable?

0	Nothing at all
1	Very weak
2	Weak
3	Moderate
4	Somewhat strong
5	Strong
6	
7	Very strong
8	
9	
10	Extremely strong

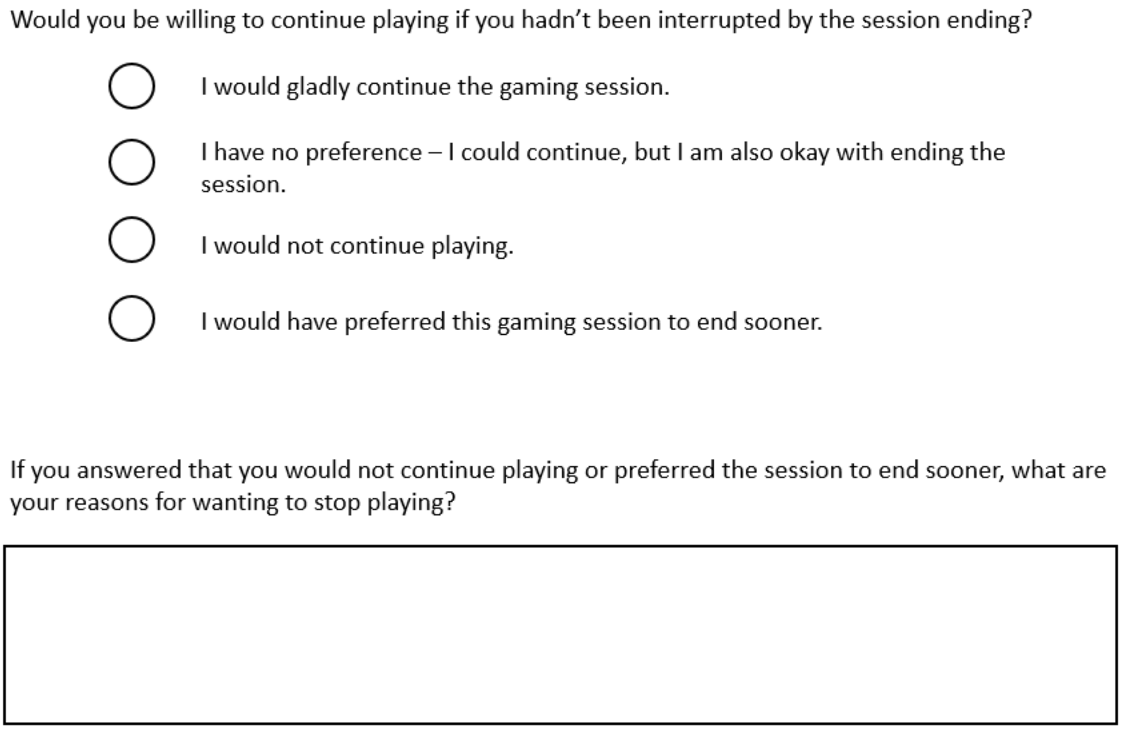
Overall prevalence and ranking of reported VRISE



Which symptoms did you experience during gameplay?	Which symptom did you find most bothersome?
☐ Neck fatigue	○ Neck fatigue
☐ Arm fatigue	○ Arm fatigue
☐ Back fatigue	○ Back fatigue
☐ Nausea	○ Nausea
☐ Headache	○ Headache
☐ Eye strain	○ Eye strain
☐ Disorientation	○ Disorientation
☐ HMD tightness	○ HMD tightness
☐ Neck pain	○ Neck pain
☐ Arm pain	○ Arm pain
☐ Back pain	○ Back pain
☐ General discomfort	○ General discomfort
☐ Thermal discomfort	○ Thermal discomfort
☐ Other...*	○ Other...*

* participants were allowed to type their own answer

Willingness to continue playing



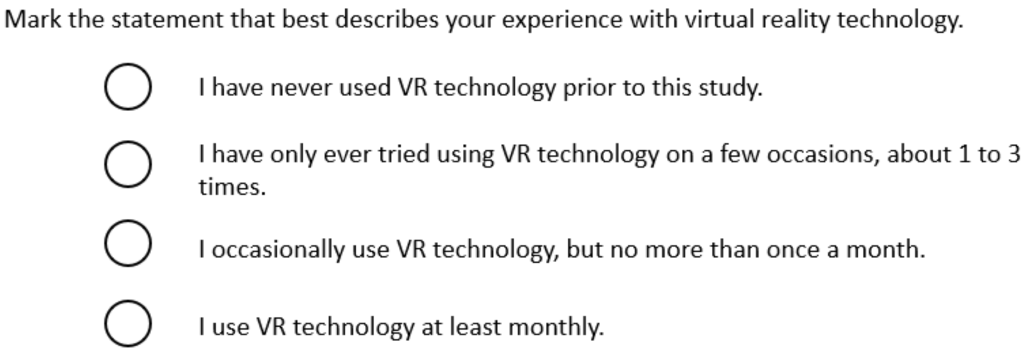
Would you be willing to continue playing if you hadn’t been interrupted by the session ending?

- ○ I would gladly continue the gaming session.
- ○ I have no preference – I could continue, but I am also okay with ending the session.
- ○ I would not continue playing.
- ○ I would have preferred this gaming session to end sooner.

If you answered that you would not continue playing or preferred the session to end sooner, what are your reasons for wanting to stop playing?

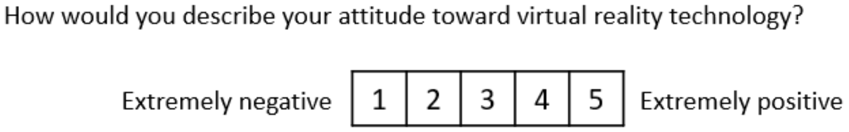
Appendix B: Scales reported in Chapter 6

Pre-study questionnaire



Mark the statement that best describes your experience with virtual reality technology.

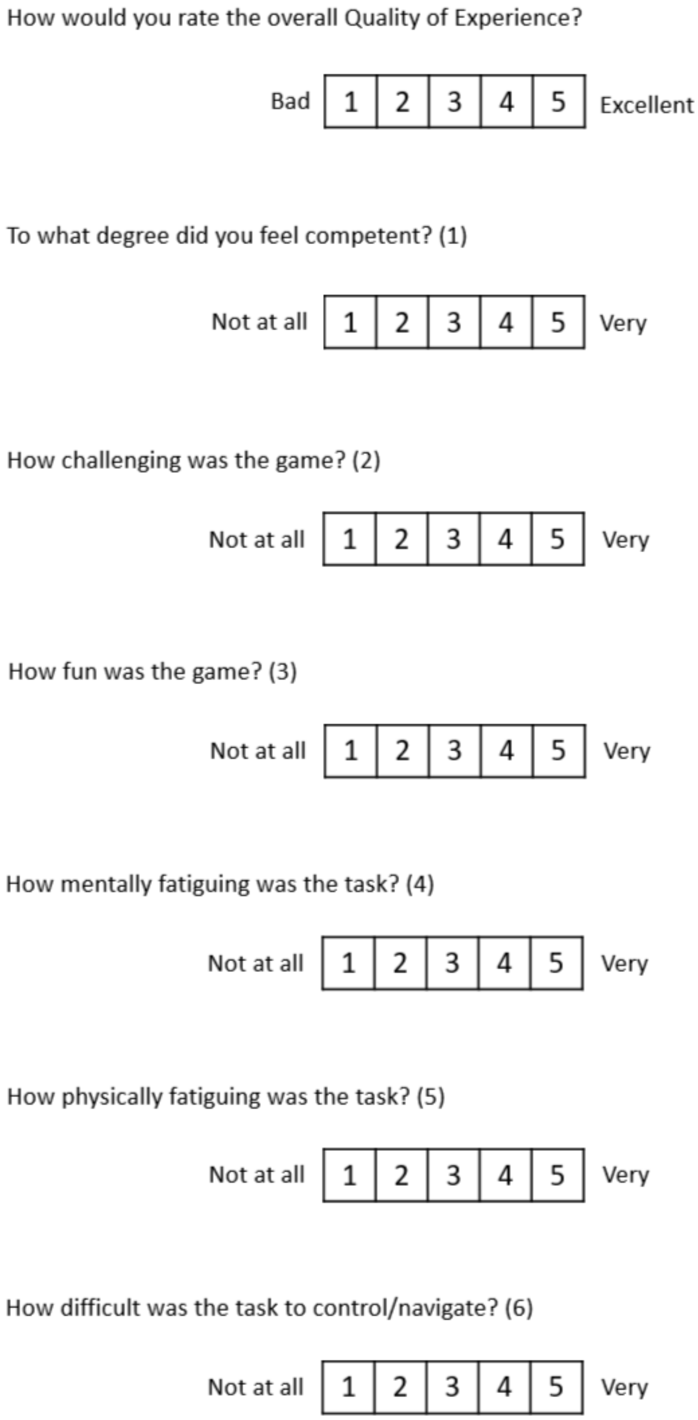
- ○ I have never used VR technology prior to this study.
- ○ I have only ever tried using VR technology on a few occasions, about 1 to 3 times.
- ○ I occasionally use VR technology, but no more than once a month.
- ○ I use VR technology at least monthly.



How would you describe your attitude toward virtual reality technology?

Extremely negative	1	2	3	4	5	Extremely positive

Post-scenario questionnaire



How would you rate the overall Quality of Experience?

Bad	1	2	3	4	5	Excellent

To what degree did you feel competent? (1)

Not at all	1	2	3	4	5	Very

How challenging was the game? (2)

Not at all	1	2	3	4	5	Very

How fun was the game? (3)

Not at all	1	2	3	4	5	Very

How mentally fatiguing was the task? (4)

Not at all	1	2	3	4	5	Very

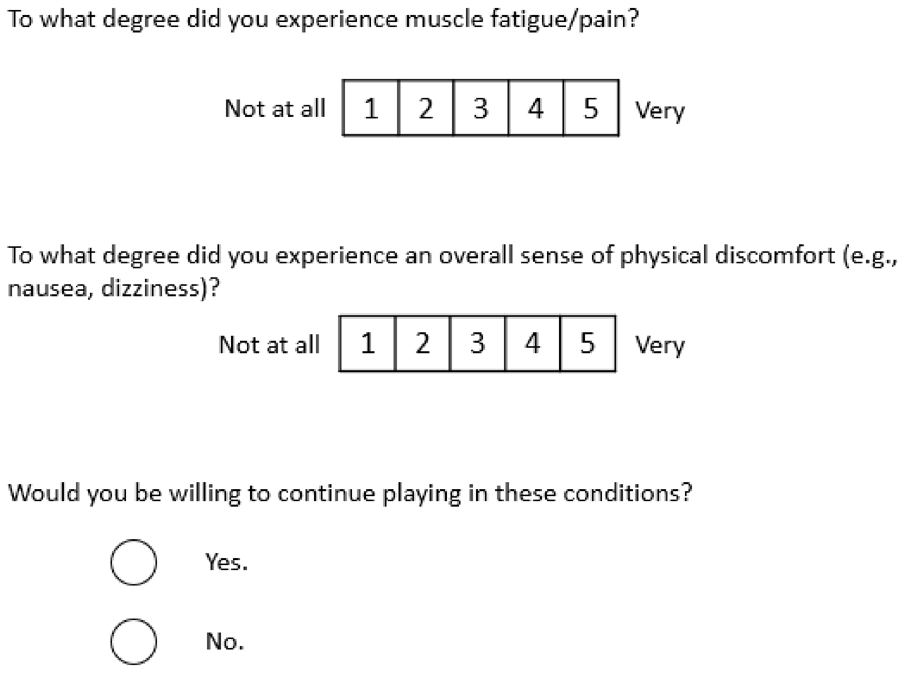
How physically fatiguing was the task? (5)

Not at all	1	2	3	4	5	Very

How difficult was the task to control/navigate? (6)

Not at all	1	2	3	4	5	Very

Items (1), (2), and (3) were inspired by the Game Experience Questionnaire (GEQ) [19, 20].
Items (4), (5), and (6) were adapted from the Simulator Task Load Index (SIM-TLX) questionnaire [160].



To what degree did you experience muscle fatigue/pain?

Not at all	1	2	3	4	5	Very

To what degree did you experience an overall sense of physical discomfort (e.g., nausea, dizziness)?

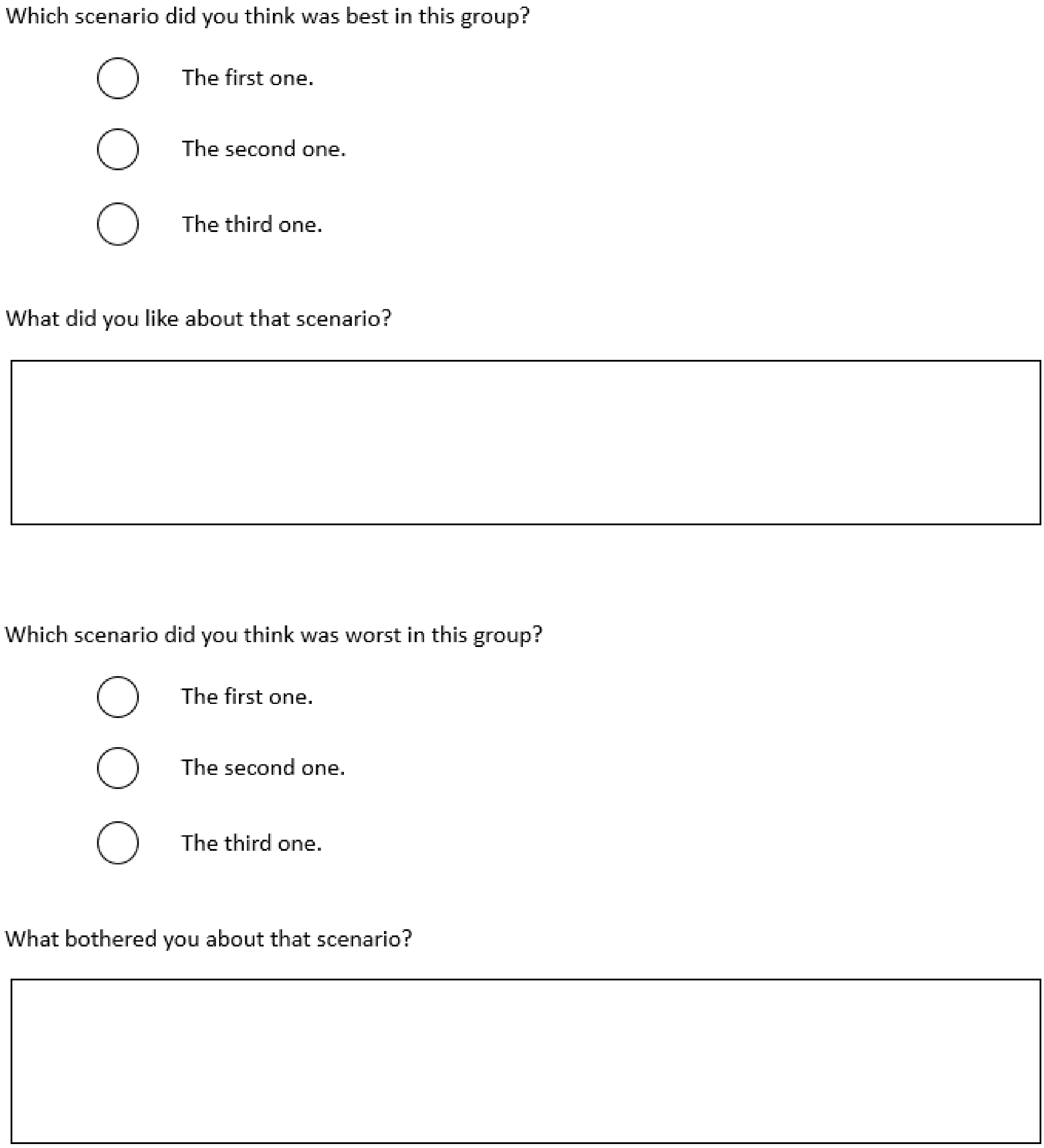
Not at all	1	2	3	4	5	Very

Would you be willing to continue playing in these conditions?

○ Yes.

○ No.

Post-group questionnaire



Which scenario did you think was best in this group?

- ○ The first one.
- ○ The second one.
- ○ The third one.

What did you like about that scenario?

Which scenario did you think was worst in this group?

- ○ The first one.
- ○ The second one.
- ○ The third one.

What bothered you about that scenario?

Appendix C: Results reported in Chapter 6

Presented here is a detailed overview of results reported in Chapter 6. Tables 1, 2, and 3 present the findings regarding participants' willingness to continue playing in given conditions.

Tables 4-14 present the analysis of subjective and objective measures collected during Study 3, Study 4, and Study 5. Results for the Slash IM (Study 3) are presented in Tables 4, 5, and 6. Results for the Pick-and-place IM (Study 4) are presented in Tables 7, 8, and 9. Results for the Shoot IM (Study 5) are presented in Tables 11, 12, 13, and 14. These tables include the following:

- mean values (M) and standard deviations (SD) for all measured subjective and objective measures for each scenario in a scenario group;
- the results of the Friedman test for each subjective and objective measure (except for Table 8, where the Wilcoxon Signed Rank test — abbreviated as Wilcoxon test in the table — was performed due to only two scenarios in the Remote grab scenario group);
- for cases where the Friedman test showed significant results, the results of the post hoc analysis with Wilcoxon Signed Rank tests (abbreviated as Wilcoxon test in the table) and Bonferroni correction were also reported.

Table 1: Slash IM: willingness to continue playing in given conditions

Parameter	**Spawn angle**			**Weapon length**			**Force to destroy**		
Scenario label	**90_DEG**	**180_DEG**	**360_DEG**	**1U**	**0_7U**	**1_3U**	**0N**	**2N**	**6N**
Percentage of participants willing to continue	97%	90%	47%	87%	87%	93%	80%	93%	47%

Table 2: Shoot IM: willingness to continue playing in given conditions

Parameter	**Shoot force**			**Spawn angle**			**Visual aids**			
Scenario label	**20N**	**40N**	**80N**	**90_DEG**	**180_DEG**	**360_DEG**	**LAS_TRAJ**	**LAS**	**TRAJ**	**NO_AID**
Percentage of participants willing to continue	50%	80%	97%	90%	87%	77%	87%	87%	97%	93%

Table 3: Pick-and-place IM: willingness to continue playing in given conditions

Parameter	**Remote grab**		**Collider scale**			**Scale offset**			**Puzzle scale**		
Scenario label	**NO_GRAB**	**REM_GRAB**	**0_2U**	**0_5U**	**1U**	**0_8S**	**0_9S**	**1S**	**0_1U**	**0_4U**	**0_7U**
Percentage of participants willing to continue	97%	97%	83%	87%	87%	90%	90%	87%	93%	97%	87%

Table 4: Slash IM: analysis of subjective and objective measures for the target spawn angle scenario group (N = 30)

SUBJECTIVE MEASURES			Overall QoE	Player experience			Workload			VRISE	
				Competence	Challenge	Fun	Mental demand	Physical demand	Task control difficulty	Muscle pain/fatigue	Overall sense of physical discomfort
Scenario label	**90_DEG**	**M**	4.40	4.13	2.60	4.27	1.43	1.63	1.77	1.10	1.23
		SD	0.62	0.68	0.86	0.64	0.57	0.81	0.77	0.31	0.68
	180_DEG	**M**	3.97	3.53	3.30	4.03	1.90	1.77	2.40	1.30	1.40
		SD	0.96	1.01	1.02	0.89	0.92	0.77	0.97	0.60	0.86
	360_DEG	**M**	3.27	2.50	4.07	3.40	2.33	2.20	3.10	1.30	1.57
		SD	1.26	1.01	0.87	1.16	1.21	1.03	0.99	0.60	1.14
Friedman test		χ^2**(2)**	24.69	39.39	33.49	17.20	20.72	10.17	25.89	6.86	10.00
		p	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.006**	**0.000**	**0.032**	**0.007**
Wilcoxon test	**Wilcoxon (180_DEG-90_DEG)**	**Z**	-2.54	-3.11	-3.41	-1.81	-2.66	-1.00	-3.14	-1.86	-1.89
		p	**0.011**	**0.002**	**0.001**	0.071	**0.008**	0.317	**0.002**	0.063	0.059
	Wilcoxon (360_DEG-90_DEG)	**Z**	-3.81	-4.56	-4.38	-3.57	-3.45	-2.64	-4.24	-1.86	-2.27
		p	**0.000**	**0.000**	**0.000**	**0.000**	**0.001**	**0.008**	**0.000**	0.063	0.023
	Wilcoxon (360_DEG-180_DEG)	**Z**	-3.31	-4.00	-3.63	-2.78	-2.81	-2.35	-2.79	0.00	-1.89
		p	**0.001**	**0.000**	**0.000**	**0.005**	**0.005**	0.019	**0.005**	1.000	0.059

OBJECTIVE MEASURES			Average force	Accuracy
Scenario label	**90_DEG**	**M**	3.46	0.69
		SD	0.64	0.10
	180_DEG	**M**	3.34	0.56
		SD	0.68	0.11
	360_DEG	**M**	3.28	0.33
		SD	0.77	0.10
Friedman test		χ^2**(2)**	2.74	50.54
		p	0.254	**0.000**
Wilcoxon test	**Wilcoxon (180_DEG-90_DEG)**	**Z**		-4.15
		p		**0.000**
	Wilcoxon (360_DEG-90_DEG)	**Z**		-4.78
		p		**0.000**
	Wilcoxon (360_DEG-180_DEG)	**Z**		-4.76
		p		**0.000**

Table 5: Slash IM: analysis of subjective and objective measures for the force to destroy scenario group (N = 30)

SUBJECTIVE MEASURES			Overall QoE	Player experience			Workload			VRISE	
				Competence	**Challenge**	**Fun**	**Mental demand**	**Physical demand**	**Task control difficulty**	**Muscle pain/fatigue**	**Overall sense of physical discomfort**
Scenario label	**0N**	**M**	4.20	4.40	2.20	3.93	1.40	1.37	1.47	1.17	1.20
		SD	0.85	0.86	1.16	1.08	0.93	0.56	0.73	0.38	0.61
	2N	**M**	4.30	4.00	2.80	4.20	1.33	1.60	1.77	1.27	1.27
		SD	0.75	0.79	0.92	0.89	0.48	0.62	0.68	0.52	0.69
	6N	**M**	3.17	2.53	4.17	3.30	1.67	2.77	2.73	1.80	1.30
		SD	1.12	1.07	0.75	1.15	0.96	1.19	1.28	0.92	0.79
Friedman test		χ^2**(2)**	25.40	36.98	42.14	19.43	4.22	41.95	28.42	19.02	1.63
		p	**0.000**	**0.000**	**0.000**	**0.000**	0.121	**0.000**	**0.000**	**0.000**	0.444
Wilcoxon test	**Wilcoxon (2N-0N)**	**Z**	-0.55	-2.29	-2.68	-0.93		-2.33	-2.50	-1.13	
		p	0.581	0.022	**0.007**	0.352		0.020	**0.013**	0.257	
	Wilcoxon (6N-0N)	**Z**	-3.64	-4.53	-4.57	-2.56		-4.26	-3.76	-3.35	
		p	**0.000**	**0.000**	**0.000**	**0.011**		**0.000**	**0.000**	**0.001**	
	Wilcoxon (6N-2N)	**Z**	-3.78	-4.21	-4.46	-3.67		-4.32	-3.67	-3.09	
		p	**0.000**	**0.000**	**0.000**	**0.000**		**0.000**	**0.000**	**0.002**	

OBJECTIVE MEASURES			Average force	Accuracy
Scenario label	**0N**	**M**	2.91	0.81
		SD	0.84	0.07
	2N	**M**	3.53	0.71
		SD	0.76	0.10
	6N	**M**	5.05	0.34
		SD	0.76	0.12
Friedman test		χ^2**(2)**	48.42	57.22
		p	**0.000**	**0.000**
Wilcoxon test	**Wilcoxon (2N-0N)**	**Z**	-4.14	-4.57
		p	**0.000**	**0.000**
	Wilcoxon (6N-0N)	**Z**	-4.78	-4.78
		p	**0.000**	**0.000**
	Wilcoxon (6N-2N)	**Z**	-4.72	-4.78
		p	**0.000**	**0.000**

Table 6: Slash IM: analysis of subjective and objective measures for the weapon length scenario group (N = 30)

SUBJECTIVE MEASURES			Overall QoE	Player experience			Workload			VRISE	
				Competence	**Challenge**	**Fun**	**Mental demand**	**Physical demand**	**Task control difficulty**	**Muscle pain/fatigue**	**Overall sense of physical discomfort**
Scenario label	**1U**	**M**	4.23	4.10	2.67	4.07	1.33	1.47	1.87	1.23	1.27
		SD	0.86	0.84	1.03	0.91	0.55	0.63	0.90	0.43	0.58
	0_7U	**M**	4.10	3.90	3.10	3.90	1.50	1.80	2.00	1.33	1.37
		SD	0.76	0.88	1.09	0.84	0.78	0.89	0.95	0.55	0.81
	1_3U	**M**	4.23	4.17	2.73	4.00	1.40	1.53	1.83	1.30	1.27
		SD	0.68	0.75	0.98	0.79	0.62	0.63	0.79	0.47	0.58
Friedman test		χ^2**(2)**	2.39	6.17	8.27	0.38	4.75	7.14	0.62	1.37	2.92
		p	0.302	**0.046**	**0.016**	0.828	0.093	**0.028**	0.735	0.504	0.232
Wilcoxon test	**Wilcoxon (0_7U-1U)**	**Z**		-1.54	-2.01			-2.24			
		p		0.124	0.044			0.025			
	Wilcoxon (1_3U-1U)	**Z**		-0.53	-0.50			-0.82			
		p		0.593	0.617			0.414			
	Wilcoxon (1_3U-0_7U)	**Z**		-2.31	-2.30			-2.00			
		p		0.021	0.022			0.046			

OBJECTIVE MEASURES			Average force	Accuracy
Scenario label	**1U**	**M**	3.47	0.71
		SD	0.67	0.11
	0_7U	**M**	3.65	0.68
		SD	0.69	0.08
	1_3U	**M**	3.34	0.71
		SD	0.71	0.09
Friedman test		χ^2**(2)**	3.98	8.02
		p	0.136	**0.018**
Wilcoxon test	**Wilcoxon (0_7U-1U)**	**Z**		-2.70
		p		**0.007**
	Wilcoxon (1_3U-1U)	**Z**		-0.67
		p		0.501
	Wilcoxon (1_3U-0_7U)	**Z**		-2.06
		p		0.040

Table 7: Pick-and-place IM: analysis of subjective and objective measures for the collider scale scenario group (N = 30)

SUBJECTIVE MEASURES			Overall QoE	Player experience			Workload			VRISE	
				Competence	**Challenge**	**Fun**	**Mental demand**	**Physical demand**	**Task control difficulty**	**Muscle pain/fatigue**	**Overall sense of physical discomfort**
Scenario label	**0_2U**	**M**	3.60	3.13	3.37	3.73	2.10	1.33	3.03	1.03	1.07
		SD	1.28	1.46	1.07	1.17	1.06	0.61	1.33	0.18	0.25
	0_5U	**M**	4.40	4.30	2.50	4.10	1.50	1.20	1.67	1.00	1.03
		SD	0.72	0.92	1.04	0.88	0.63	0.48	0.92	0.00	0.18
	1U	**M**	4.30	4.60	2.33	4.03	1.50	1.17	1.63	1.03	1.07
		SD	0.75	0.77	1.12	0.96	0.73	0.46	0.96	0.18	0.25
Friedman test		χ^2**(2)**	15.49	29.85	17.34	3.60	12.52	4.57	31.14	2.00	2.00
		p	**0.000**	**0.000**	**0.000**	0.165	**0.002**	0.102	**0.000**	0.368	0.368
Wilcoxon test	**Wilcoxon(05U-02U)**	**Z**	-2.78	-3.44	-3.08		-2.95		-4.11		
		p	**0.005**	**0.001**	**0.002**		**0.003**		**0.000**		
	Wilcoxon(1U-02U)	**Z**	-2.76	-4.06	-3.05		-2.82		-4.02		
		p	**0.006**	**0.000**	**0.002**		**0.005**		**0.000**		
	Wilcoxon(1U-05U)	**Z**	-0.68	-1.67	-0.90		0.00		-0.25		
		p	0.499	0.095	0.369		1.000		0.805		

OBJECTIVE MEASURES			Percentage of puzzle pieces placed	Duration [s]	Duration of successful [s]	Percentage of successful participants
Scenario label	**0_2U**	**M**	76%	83.66	74.14	40%
		SD	28%	10.98	12.39	
	0_5U	**M**	95%	63.80	59.77	87%
		SD	15%	16.02	13.08	
	1U	**M**	98%	54.21	51.65	93%
		SD	8%	18.60	16.43	
Friedman		χ^2**(2)**	22.06	30.74	7.80	
		p	**0.000**	**0.000**	**0.020**	
Wilcoxon test	**Wilcoxon(05U-02U)**	**Z**	-2.76	-4.44	-2.85	
		p	**0.006**	**0.000**	**0.004**	
	Wilcoxon(1U-02U)	**Z**	-3.24	-4.57	-2.67	
		p	**0.001**	**0.000**	**0.008**	
	Wilcoxon(1U-05U)	**Z**	-1.06	-2.36	-2.43	
		p	0.288	0.019	**0.015**	

Table 8: Pick-and-place IM: analysis of subjective and objective measures for the remote grab scenario group (N = 30)

SUBJECTIVE MEASURES			Overall QoE	Player experience			Workload			VRISE	
				Competence	**Challenge**	**Fun**	**Mental demand**	**Physical demand**	**Task control diffi-culty**	**Muscle pain/fatigue**	**Overall sense of physical discom-fort**
Scenario label	**NO_GRAB**	**M**	4.50	4.50	2.47	4.03	1.43	1.17	1.47	1.00	1.07
		SD	0.63	0.73	1.04	0.85	0.57	0.38	0.63	0.00	0.25
	REM_GRAB	**M**	4.50	4.20	3.13	4.50	1.67	1.17	2.03	1.00	1.07
		SD	0.68	0.85	1.07	0.73	0.80	0.38	1.07	0.00	0.25
Wilcoxon test	**Wilcoxon (REM_GRAB-NO_GRAB)**	**Z**	-0.02	-1.89	-3.41	-2.56	-2.11	0.00	8.07	0.00	0.00
		p	0.983	0.059	**0.001**	**0.010**	**0.035**	1.000	**0.005**	1.000	1.000

OBJECTIVE MEASURES			Percentage of puzzle pieces placed	Duration [s]	Duration of successful [s]	Percentage of successful participants
Scenario label	**NO_GRAB**	**M**	100%	55.71	55.71	100%
		SD	0%	13.32	13.32	
	REM_GRAB	**M**	90%	71.08	62.98	70%
		SD	19%	17.20	14.10	
Wilcoxon	**Wilcoxon (REM_GRAB-NO_GRAB)**	**Z**	-2.68	-3.79	-2.38	
		p	**0.007**	**0.000**	**0.017**	

Table 9: Pick-and-place IM: analysis of subjective and objective measures for the scale offset scenario group (N = 30)

SUBJECTIVE MEASURES			Overall QoE	Player experience			Workload			VRISE	
				Competence	Challenge	Fun	Mental demand	Physical demand	Task control difficulty	Muscle pain/fatigue	Overall sense of physical discomfort
Scenario label	**0_1U**	**M**	4.43	4.43	2.43	4.00	1.37	1.13	1.47	1.07	1.07
		SD	0.77	0.82	1.04	0.79	0.61	0.35	0.73	0.25	0.25
	0_4U	**M**	4.33	4.43	2.30	3.93	1.53	1.13	1.50	1.00	1.07
		SD	0.71	0.82	1.02	0.83	0.90	0.35	0.73	0.00	0.25
	0_7U	**M**	4.27	4.13	3.03	3.87	1.70	1.23	2.20	1.00	1.07
		SD	0.87	0.97	1.19	0.94	0.88	0.43	1.24	0.00	0.25
Friedman test		**χ^2(2)**	1.97	4.96	10.93	0.79	5.64	4.50	20.28	4.00	
		p	0.374	0.084	**0.004**	0.673	0.060	0.105	**0.000**	0.135	
Wilcoxon test	**Wilcoxon(04U-01U)**	**Z**			-0.83				-0.30		
		p			0.405				0.763		
	Wilcoxon(07U-01U)	**Z**			-2.38				-3.24		
		p			0.017				**0.001**		
	Wilcoxon(07U-04U)	**Z**			-3.25				-3.46		
		p			**0.001**				**0.001**		

OBJECTIVE MEASURES			Percentage of puzzle pieces placed	Duration [s]	Duration of successful [s]	Percentage of successful participants
Scenario label	**0_1U**	**M**	100%	55.35	55.35	100%
		SD	0%	13.36	13.36	
	0_4U	**M**	99%	52.79	50.14	93%
		SD	4%	14.89	11.32	
	0_7U	**M**	90%	70.27	58.85	63%
		SD	16%	19.29	14.98	
Friedman		**χ^2(2)**	18.22	10.07	3.00	
		p	**0.000**	**0.007**	0.223	
Wilcoxon test	**Wilcoxon(04U-01U)**	**Z**	-1.41	-1.02		
		p	0.157	0.309		
	Wilcoxon(07U-01U)	**Z**	-2.99	-3.34		
		p	**0.003**	**0.001**		
	Wilcoxon(07U-04U)	**Z**	-2.80	-3.63		
		p	**0.005**	**0.000**		

Table 10: Pick-and-place IM: analysis of subjective and objective measures for the puzzle scale scenario group (N = 30)

SUBJECTIVE MEASURES			Overall QoE	Player experience			Workload			VRISE	
				Competence	Challenge	Fun	Mental demand	Physical demand	Task control difficulty	Muscle pain/fatigue	Overall sense of physical discomfort
Scenario label	**0_8S**	**M**	4.23	3.73	3.23	4.13	1.87	1.30	2.30	1.00	1.07
		SD	0.86	1.05	1.17	1.07	0.82	0.47	1.34	0.00	0.25
	0_9S	**M**	4.50	4.23	2.60	4.27	1.73	1.27	2.13	1.00	1.13
		SD	0.68	0.94	1.19	0.64	0.64	0.45	0.97	0.00	0.43
	1S	**M**	4.00	3.17	3.40	4.07	1.83	1.53	2.83	1.03	1.10
		SD	0.83	1.23	1.30	0.87	0.87	0.82	1.21	0.18	0.40
Friedman test		χ^2**(2)**	7.71	13.01	10.29	2.20	1.55	4.69	10.83	2.00	1.00
		p	**0.021**	**0.001**	**0.006**	0.333	0.461	0.096	**0.004**	0.368	0.607
Wilcoxon test	**Wilcoxon(0_9S-0_8S)**	**Z**	-1.37	-2.33	-2.39				-0.90		
		p	0.171	0.020	**0.017**				0.366		
	Wilcoxon(1S-0_8S)	**Z**	-1.26	-2.55	-0.91				-1.95		
		p	0.207	**0.011**	0.363				0.052		
	Wilcoxon(1S-0_9S)	**Z**	-2.51	-3.65	-3.10				-3.38		
		p	**0.012**	**0.000**	**0.002**				**0.001**		

OBJECTIVE MEASURES			Percentage of puzzle pieces placed	Duration [s]	Duration of successful [s]	Percentage of successful participants
Scenario label	**0_1U**	**M**	86%	78.04	67.57	53%
		SD	19%	13.78	10.82	
	0_4U	**M**	95%	68.84	63.55	80%
		SD	10%	15.34	12.28	
	0_7U	**M**	71%	84.09	67.91	27%
		SD	23%	12.73	16.23	
Friedman		χ^2**(2)**	27.30	17.37	7.14	
		p	**0.000**	**0.000**	**0.028**	
Wilcoxon test	**Wilcoxon(04U-01U)**	**Z**	-2.72	-3.16	-1.81	
		p	**0.006**	**0.002**	0.070	
	Wilcoxon(07U-01U)	**Z**	-2.98	-2.35	-1.18	
		p	**0.003**	0.019	0.237	
	Wilcoxon(07U-04U)	**Z**	-3.79	-3.53	-1.52	
		p	**0.000**	**0.000**	0.128	

Table 11: Shoot IM: analysis of subjective and objective measures for the target spawn angle scenario group (N = 30)

SUBJECTIVE MEASURES			**Overall QoE**	**Player experience**			**Workload**			**VRISE**	
				Competence	**Challenge**	**Fun**	**Mental demand**	**Physical demand**	**Task control difficulty**	**Muscle pain/fatigue**	**Overall sense of physical discomfort**
Scenario label	**90_DEG**	**M**	4.43	4.00	2.60	4.17	1.60	1.77	1.60	1.80	1.13
		SD	0.82	0.69	0.93	0.95	0.77	0.94	0.77	1.03	0.35
	180_DEG	**M**	4.70	3.97	2.93	4.53	1.80	1.90	1.77	1.70	1.17
		SD	0.60	0.72	0.94	0.73	0.85	1.06	0.82	0.88	0.46
	360_DEG	**M**	4.20	3.80	3.43	4.23	2.00	2.07	2.03	1.63	1.33
		SD	0.96	1.03	0.94	0.90	0.98	0.98	0.96	0.81	0.71
Friedman test		χ^2**(2)**	15.80	0.64	27.83	6.74	11.41	6.40	12.25	1.00	3.89
		p	**0.000**	0.727	**0.000**	**0.034**	**0.003**	**0.041**	**0.002**	0.607	0.143
Wilcoxon test	**Wilcoxon (180_DEG-90_DEG)**	**Z**	-2.11		-2.67	-2.50	-1.90	-1.41	-1.25		
		p	0.035		**0.008**	**0.012**	0.058	0.157	0.212		
	Wilcoxon (360_DEG-90_DEG)	**Z**	-1.73		-4.18	-0.23	-3.05	-2.18	-2.98		
		p	0.084		**0.000**	0.817	**0.002**	0.029	**0.003**		
	Wilcoxon (360_DEG-180_DEG)	**Z**	-3.42		-3.44	-2.31	-1.51	-1.51	-2.14		
		p	**0.001**		**0.001**	0.021	0.132	0.132	0.033		

OBJECTIVE MEASURES			**Number of shots fired**	**Accuracy**
Scenario label	**90_DEG**	**M**	174.17	0.65
		SD	98.39	0.19
	180_DEG	**M**	143.33	0.62
		SD	87.50	0.17
	360_DEG	**M**	103.93	0.61
		SD	46.95	0.17
Friedman test		χ^2**(2)**	50.03	10.85
		p	**0.000**	**0.004**
Wilcoxon test	**Wilcoxon (180_DEG-90_DEG)**	**Z**	-4.68	-2.86
		p	**0.000**	**0.004**
	Wilcoxon (360_DEG-90_DEG)	**Z**	-4.76	-2.69
		p	**0.000**	**0.007**
	Wilcoxon (360_DEG-180_DEG)	**Z**	-4.57	-0.61
		p	**0.000**	0.540

Table 12: Shoot IM: analysis of subjective measures for the visual aiming aids scenario group (N = 30)

SUBJECTIVE MEASURES			**Overall QoE**	**Player experience**			**Workload**			**VRISE**	
				Competence	**Challenge**	**Fun**	**Mental demand**	**Physical demand**	**Task control difficulty**	**Muscle pain/fatigue**	**Overall sense of physical discomfort**
Scenario label	**LAS_TRAJ**	**M**	4.40	4.03	2.33	4.07	1.63	1.43	1.67	1.30	1.03
		SD	0.81	0.93	0.88	1.17	0.89	0.77	1.03	0.70	0.18
	LAS	**M**	4.47	4.53	2.07	4.40	1.47	1.50	1.40	1.30	1.07
		SD	0.90	0.86	0.91	0.97	0.78	0.86	0.67	0.75	0.25
	TRAJ	**M**	4.17	3.57	2.90	4.33	1.83	1.87	1.90	1.57	1.03
		SD	0.91	0.90	0.80	0.88	0.91	0.97	0.96	0.90	0.18
	NO_AID	**M**	4.43	3.70	3.37	4.43	2.20	2.17	2.50	1.97	1.07
		SD	0.77	0.95	1.16	0.86	0.96	1.18	1.33	1.19	0.25
Friedman test		χ^2**(3)**	4.21	30.07	34.15	3.18	21.42	17.84	27.97	19.68	1.20
		p	0.240	**0.000**	**0.000**	0.365	**0.000**	**0.000**	**0.000**	**0.000**	0.753
Wilcoxon test	**Wilcoxon (LAS-LAS_TRAJ)**	**Z**		-3.42	-1.61		-0.88	-0.63	-2.13	0.00	
		p		**0.001**	0.106		0.376	0.527	0.033	1.000	
	Wilcoxon (TRAJ-LAS_TRAJ)	**Z**		-2.29	-2.32		-0.95	-2.59	-1.46	-1.64	
		p		0.022	0.020		0.341	0.010	0.144	0.101	
	Wilcoxon (TRAJ-LAS)	**Z**		-3.73	-3.60		-2.02	-2.08	-2.42	-1.79	
		p		**0.000**	**0.000**		0.043	0.038	0.015	0.074	
	Wilcoxon (NO_AID-LAS_TRAJ)	**Z**		-1.67	-3.59		-2.78	-3.35	-3.34	-3.09	
		p		0.095	**0.000**		**0.005**	**0.001**	**0.001**	**0.002**	
	Wilcoxon (NO_AID-LAS)	**Z**		-2.91	-4.03		-3.33	-2.99	-3.78	-2.96	
		p		**0.004**	**0.000**		**0.001**	**0.003**	**0.000**	**0.003**	
	Wilcoxon (NO_AID-TRAJ)	**Z**		-0.68	-2.14		-2.61	-1.70	-2.30	-2.97	
		p		0.499	0.033		0.009	0.089	0.022	**0.003**	

Table 13: Shoot IM: analysis of objective measures for the visual aiming aids scenario group (N = 30)

OBJECTIVE MEASURES			**Number of shots fired**	**Accuracy**
Scenario label	**LAS_TRAJ**	**M**	169.70	0.72
		SD	86.09	0.17
	LAS	**M**	195.53	0.80
		SD	64.51	0.14
	TRAJ	**M**	156.37	0.61
		SD	71.59	0.13
	NO_AID	**M**	168.10	0.51
		SD	71.86	0.23
Friedman test		χ^2**(3)**	33.19	60.68
		p	**0.000**	**0.000**
Wilcoxon test	**Wilcoxon (LAS-LAS_TRAJ)**	**Z**	-3.28	-4.22
		p	**0.001**	**0.000**
	Wilcoxon (TRAJ-LAS_TRAJ)	**Z**	-0.96	-3.72
		p	0.336	**0.000**
	Wilcoxon (TRAJ-LAS)	**Z**	-4.19	-4.51
		p	**0.000**	**0.000**
	Wilcoxon (NO_AID-LAS_TRAJ)	**Z**	-0.25	-4.30
		p	0.805	**0.000**
	Wilcoxon (NO_AID-LAS)	**Z**	-3.78	-4.78
		p	**0.000**	**0.000**
	Wilcoxon (NO_AID-TRAJ)	**Z**	-2.00	-2.48
		p	0.045	0.013

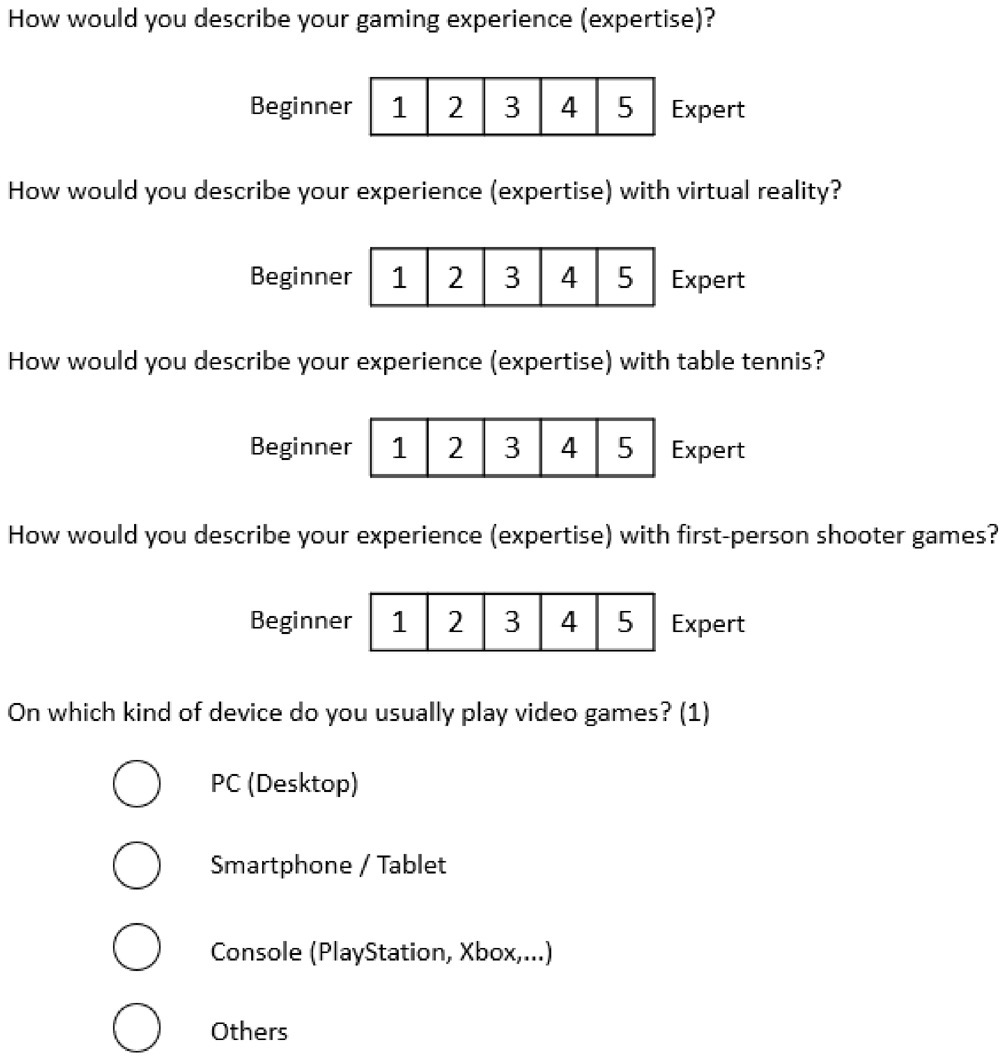
Table 14: Shoot IM: analysis of subjective and objective measures for the shoot force scenario group (N = 30)

SUBJECTIVE MEASURES			**Overall QoE**	**Player experience**			**Workload**			**VRISE**	
				Competence	**Challenge**	**Fun**	**Mental demand**	**Physical demand**	**Task control difficulty**	**Muscle pain/fatigue**	**Overall sense of physical discomfort**
Scenario label	**20N**	**M**	3.63	3.23	3.33	3.43	2.13	1.87	2.23	1.73	1.23
		SD	1.00	0.94	1.12	1.28	0.86	1.07	1.10	0.91	0.50
	40N	**M**	4.23	4.03	2.37	4.00	1.40	1.63	1.63	1.63	1.30
		SD	0.77	0.93	0.93	1.02	0.67	0.89	0.81	0.89	0.65
	80N	**M**	4.47	4.40	2.20	4.37	1.37	1.63	1.43	1.67	1.23
		SD	0.68	0.72	1.00	0.85	0.72	0.89	0.68	0.88	0.57
Friedman test		χ^2**(2)**	20.82	27.59	22.71	16.00	26.79	6.00	24.11	1.56	0.11
		p	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.050**	**0.000**	0.459	0.949
Wilcoxon test	**Wilcoxon (40N-20N)**	**Z**	-3.14	-3.82	-3.91	-2.27	-3.82	-2.11	-3.14		
		p	**0.002**	**0.000**	**0.000**	0.023	**0.000**	0.035	**0.002**		
	Wilcoxon (80N-20N)	**Z**	-3.58	-3.79	-3.65	-3.50	-3.57	-1.94	-3.61		
		p	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	0.052	**0.000**		
	Wilcoxon (80N-40N)	**Z**	-2.11	-1.97	-1.29	-2.18	-0.38	0.00	-2.45		
		p	0.035	0.049	0.197	0.029	0.705	1.000	**0.014**		

OBJECTIVE MEASURES			**Number of shots fired**	**Accuracy**
Scenario label	**20N**	**M**	171.97	0.52
		SD	107.73	0.18
	40N	**M**	186.27	0.63
		SD	96.94	0.17
	80N	**M**	192.20	0.65
		SD	76.59	0.14
Friedman test		χ^2**(2)**	20.47	23.85
		p	**0.000**	**0.000**
Wilcoxon test	**Wilcoxon (40N-20N)**	**Z**	-2.73	-4.56
		p	**0.006**	**0.000**
	Wilcoxon (80N-20N)	**Z**	-2.85	-4.19
		p	**0.004**	**0.000**
	Wilcoxon (80N-40N)	**Z**	-2.24	-1.34
		p	0.025	0.180

Appendix D: Scales reported in Chapter 7

Pre-study questionnaire



How would you describe your gaming experience (expertise)?

Beginner	1	2	3	4	5	Expert

How would you describe your experience (expertise) with virtual reality?

Beginner	1	2	3	4	5	Expert

How would you describe your experience (expertise) with table tennis?

Beginner	1	2	3	4	5	Expert

How would you describe your experience (expertise) with first-person shooter games?

Beginner	1	2	3	4	5	Expert

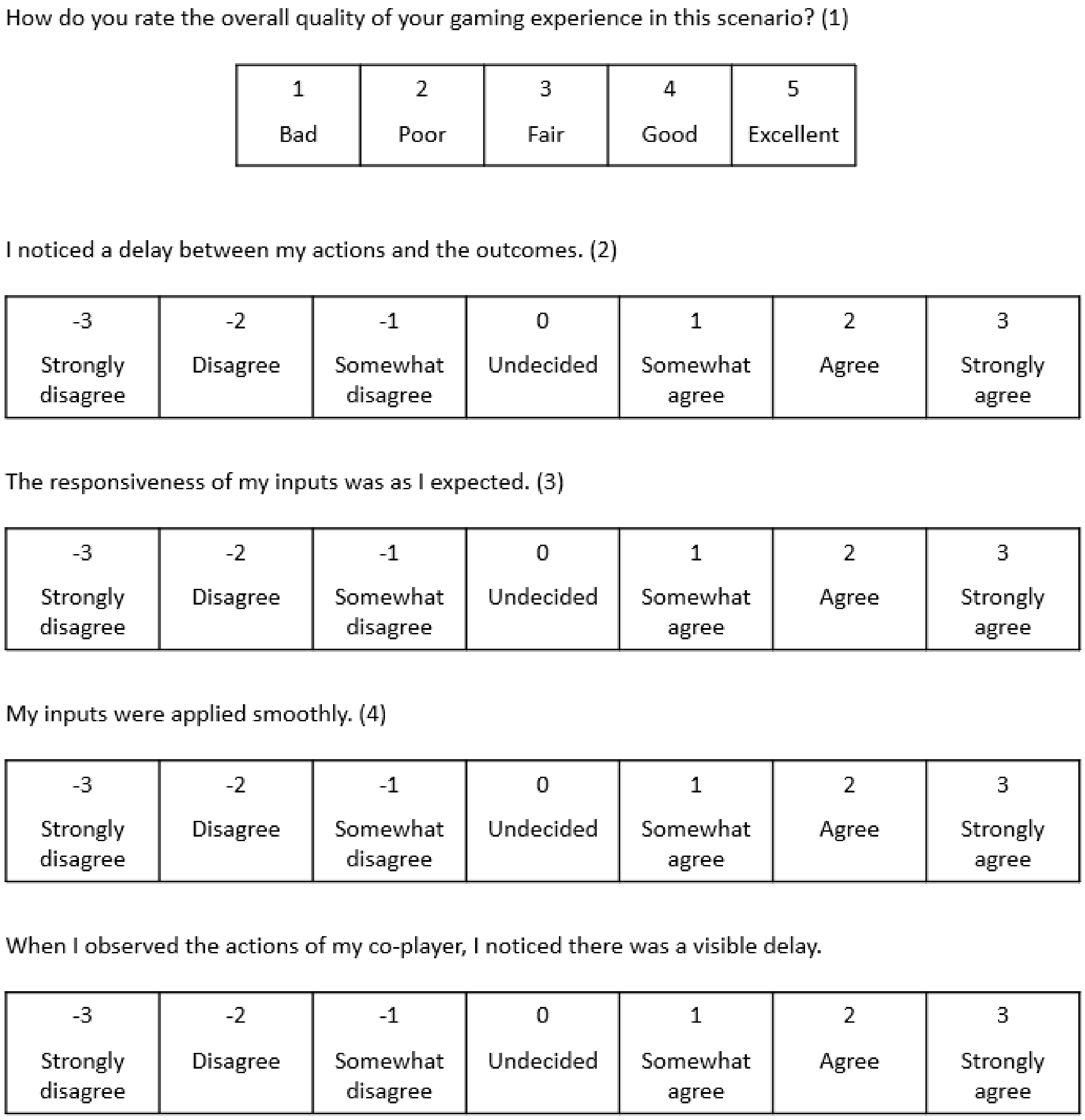
On which kind of device do you usually play video games? (1)

- ○ PC (Desktop)
- ○ Smartphone / Tablet
- ○ Console (PlayStation, Xbox,...)
- ○ Others

Item (1) was taken from [254].

Post-scenario questionnaire

After each scenario



How do you rate the overall quality of your gaming experience in this scenario? (1)

1 Bad	2 Poor	3 Fair	4 Good	5 Excellent

I noticed a delay between my actions and the outcomes. (2)

-3 Strongly disagree	-2 Disagree	-1 Somewhat disagree	0 Undecided	1 Somewhat agree	2 Agree	3 Strongly agree

The responsiveness of my inputs was as I expected. (3)

-3 Strongly disagree	-2 Disagree	-1 Somewhat disagree	0 Undecided	1 Somewhat agree	2 Agree	3 Strongly agree

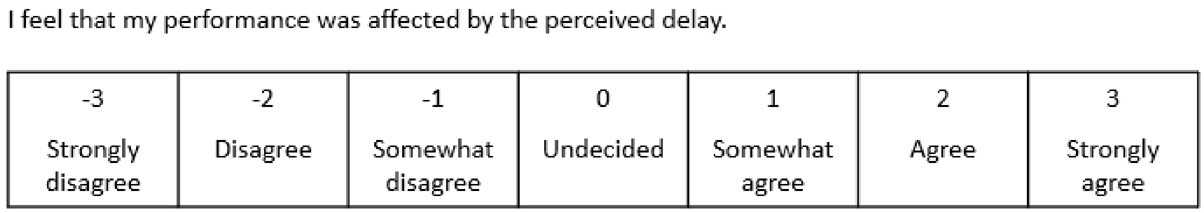
My inputs were applied smoothly. (4)

-3 Strongly disagree	-2 Disagree	-1 Somewhat disagree	0 Undecided	1 Somewhat agree	2 Agree	3 Strongly agree

When I observed the actions of my co-player, I noticed there was a visible delay.

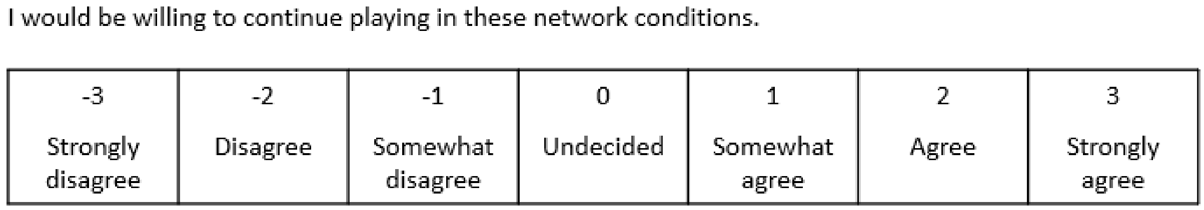
-3 Strongly disagree	-2 Disagree	-1 Somewhat disagree	0 Undecided	1 Somewhat agree	2 Agree	3 Strongly agree

Items (1), (2), (3), and (4) were adapted from the Gaming Input Quality Scale (GIPS) [254].



I feel that my performance was affected by the perceived delay.

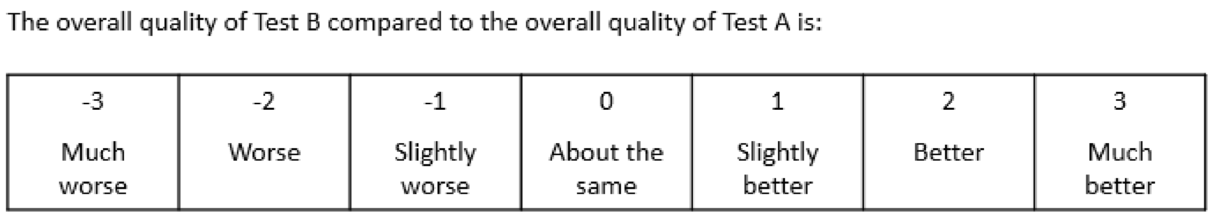
-3	-2	-1	0	1	2	3
Strongly disagree	Disagree	Somewhat disagree	Undecided	Somewhat agree	Agree	Strongly agree



I would be willing to continue playing in these network conditions.

-3	-2	-1	0	1	2	3
Strongly disagree	Disagree	Somewhat disagree	Undecided	Somewhat agree	Agree	Strongly agree

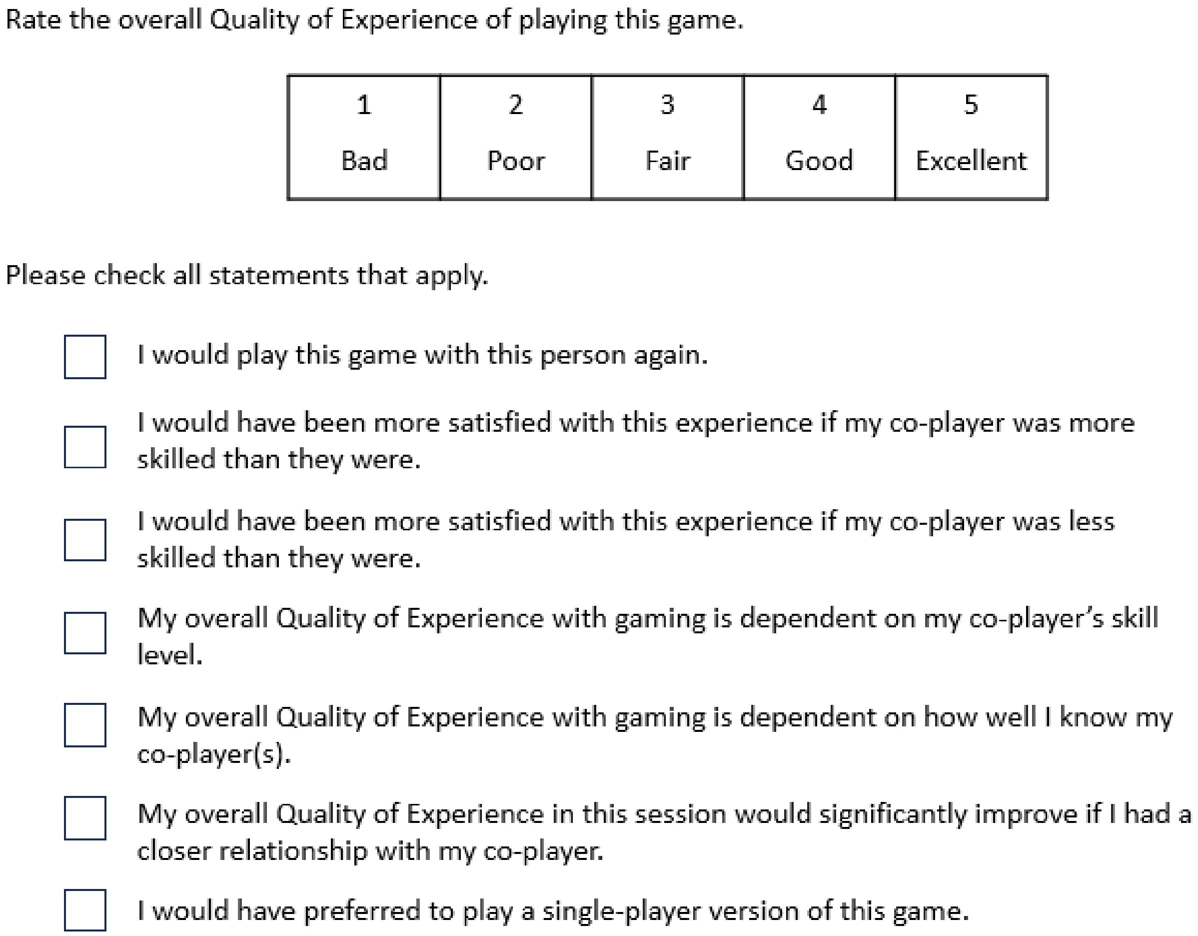
After each pair of scenarios (referred to as Test A and Test B)



The overall quality of Test B compared to the overall quality of Test A is:

-3	-2	-1	0	1	2	3
Much worse	Worse	Slightly worse	About the same	Slightly better	Better	Much better

Post-game questionnaire



Rate the overall Quality of Experience of playing this game.

1	2	3	4	5
Bad	Poor	Fair	Good	Excellent

Please check all statements that apply.

☐ I would play this game with this person again.

☐ I would have been more satisfied with this experience if my co-player was more skilled than they were.

☐ I would have been more satisfied with this experience if my co-player was less skilled than they were.

☐ My overall Quality of Experience with gaming is dependent on my co-player’s skill level.

☐ My overall Quality of Experience with gaming is dependent on how well I know my co-player(s).

☐ My overall Quality of Experience in this session would significantly improve if I had a closer relationship with my co-player.

☐ I would have preferred to play a single-player version of this game.

Game Experience Questionnaire (GEQ) [19, 20]
(Core Module: Competence)

I felt skillful.

0 Not at all	1 Slightly	2 Moderately	3 Fairly	4 Extremely

I felt competent.

0 Not at all	1 Slightly	2 Moderately	3 Fairly	4 Extremely

I was good at it.

0 Not at all	1 Slightly	2 Moderately	3 Fairly	4 Extremely

I felt successful.

0 Not at all	1 Slightly	2 Moderately	3 Fairly	4 Extremely

I was fast at reaching the game’s targets.

0 Not at all	1 Slightly	2 Moderately	3 Fairly	4 Extremely

Game Experience Questionnaire (GEQ) [19, 20]
(Social Presence Module: Negative feelings — chosen items)

I felt jealous about the other.

0 Not at all	1 Slightly	2 Moderately	3 Fairly	4 Extremely

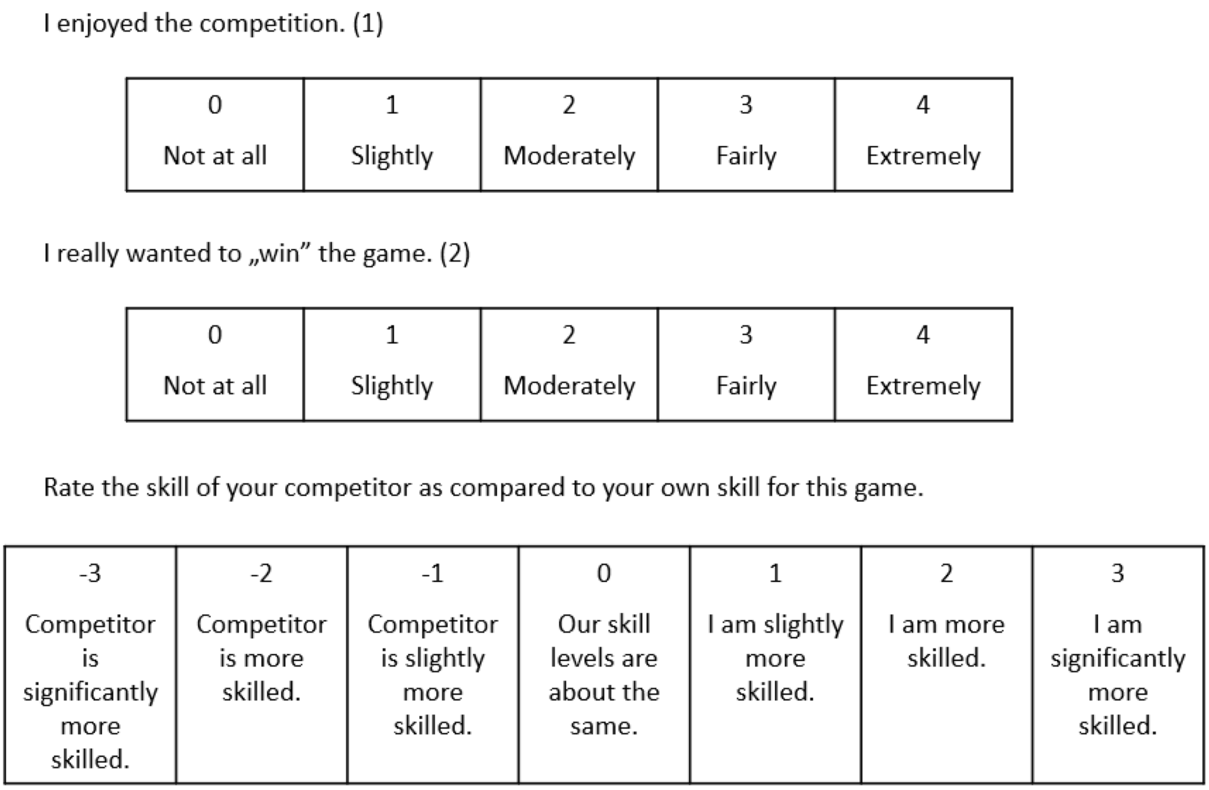
I felt revengeful.

0 Not at all	1 Slightly	2 Moderately	3 Fairly	4 Extremely

I felt schadenfreude (malicious delight).

0 Not at all	1 Slightly	2 Moderately	3 Fairly	4 Extremely

Competitive feelings and competitor skill



I enjoyed the competition. (1)

0 Not at all	1 Slightly	2 Moderately	3 Fairly	4 Extremely

I really wanted to „win" the game. (2)

0 Not at all	1 Slightly	2 Moderately	3 Fairly	4 Extremely

Rate the skill of your competitor as compared to your own skill for this game.

-3 Competitor is significantly more skilled.	-2 Competitor is more skilled.	-1 Competitor is slightly more skilled.	0 Our skill levels are about the same.	1 I am slightly more skilled.	2 I am more skilled.	3 I am significantly more skilled.

Item (1) was adapted from [257].
Item (2) was adapted from [113].

Social interaction

I communicated with the other participant. (1)

0	1	2	3	4
Not at all	Slightly	Moderately	Fairly	Extremely

I felt a strong sense of co-presence, as if we were in the same room.

0	1	2	3	4
Not at all	Slightly	Moderately	Fairly	Extremely

The game created some sort of social bonding between me and the other player. (2)

0	1	2	3	4
Not at all	Slightly	Moderately	Fairly	Extremely

I had the feeling we were doing something together. (3)

0	1	2	3	4
Not at all	Slightly	Moderately	Fairly	Extremely

Item (1) was adapted from [255].
Items (2) and (3) were adapted from [256].

Bibliography

[1]Perkis, A., Timmerer, C. *et al.*, “QUALINET White Paper on Definitions of Immersive Media Experience (IMEx)”, European Network on Quality of Experience in Multimedia Systems and Services, 14th QUALINET meeting (online), 2020.

[2]Sutherland, I. E., “A head-mounted three dimensional display”, in Proceedings of the December 9-11, 1968, Fall Joint Computer Conference, Part I, 1968, str. 757–764.

[3]Pallavicini, F., Ferrari, A., Zini, A., Garcea, G., Zanacchi, A., Barone, G., Mantovani, F., “What Distinguishes a Traditional Gaming Experience from One in Virtual Reality? An Exploratory Study”, in Int’l Conf. on App. Hum. Fact. and Ergon. Springer, 2017, str. 225–231.

[4]Shelstad, W. J., Smith, D. C., Chaparro, B. S., “Gaming on the Rift: How Virtual Reality Affects Game User Satisfaction”, in Proc. of the Hum. Fact. and Ergon. Soc. Ann. Meet., Vol. 61(1). SAGE Publ. Sage CA: Los Ang., CA, 2017, str. 2072–2076.

[5]Pallavicini, F., Pepe, A., Minissi, M. E., “Gaming in Virtual Reality: What Changes in Terms of Usability, Emotional Response and Sense of Presence Compared to Non-Immersive Video Games?”, Simul. & Gam., Vol. 50, No. 2, 2019, str. 136–159.

[6]“AR & VR - Worldwide | Statista Market Forecast”, dostupno na: https://www.statista.com/outlook/amo/ar-vr/worldwide [Online; accessed 31. Jul. 2023]. Jun. 2023.

[7]Biocca, F., Delaney, B., “Immersive virtual reality technology”, Communication in the Age of Virtual Reality, Vol. 15, No. 32, 1995, str. 10–5555.

[8]ITU-T Recomm. P.1320, “Quality of Experience Assessment of Extended Reality Meetings”, Int’l Telecomm. Union-Telecomm. Standard. Sect., 2022.

[9]Weech, S., Kenny, S., Barnett-Cowan, M., “Presence and Cybersickness in Virtual Reality are Negatively Related: A Review”, Front. in Psychol., Vol. 10, 2019, str. 158.

[10]Vlahovic, S., Suznjevic, M., Skorin-Kapov, L., “A framework for the classification and evaluation of game mechanics for virtual reality games”, Electronics, Vol. 11, No. 18, 2022, str. 2946.

[11]Malone, T. W., “Toward a theory of intrinsically motivating instruction”, Cognitive Science, Vol. 5, No. 4, 1981, str. 333–369.

[12]Vlahovic, S., Slivar, I., Silic, M., Skorin-Kapov, L., Suznjevic, M., “Exploring the facets of the multiplayer vr gaming experience”, ACM Transactions on Multimedia Computing, Communications, and Applications (accepted).

[13]ITU-T P.809, “Recommendation ITU-T P.809 - Subjective evaluation methods for gaming quality”, 2018.

[14]Int. Telecommunication Union, “ITU-T Recommendation P.1320 – QoE assessment of extended reality (XR) meetings”, 2022.

[15]Int. Telecommunication Union, “ITU-T Work Item: Interactive test methods for subjective assessment of extended reality communications, under study”, 2022.

[16]Pérez, P., González-Sosa, E., Gutiérrez, J., García, N., “Emerging Immersive Communication Systems: Overview, Taxonomy, and Good Practises for QoE Assessment”, arXiv preprint arXiv:2205.05953, 2022.

[17]Int. Telecommunication Union, “ITU-T Work Item: Subjective test method for interactive virtual reality applications, under study”, 2023.

[18]Kennedy, R., Hettinger, L., Lilienthal, M., “Simulator Sickness”, Mot. and Space Sick., 1988, str. 317–341.

[19]Poels, K., De Kort, Y., Ijsselsteijn, W., “"It is Always a Lot of Fun!" Exploring Dimensions of Digital Game Experience Using Focus Group Methodology”, in Proc. of the 2007 conf. on Fut. Play, 2007, str. 83–89.

[20]IJsselsteijn, W., De Kort, Y., Poels, K., “The Game Experience Questionnaire”, Eindhoven: Techn. Univ. Eindh., 2013, str. 3–9.

[21]Lin, J.-W., Duh, H. B.-L., Parker, D. E., Abi-Rached, H., Furness, T. A., “Effects of Field of View on Presence, Enjoyment, Memory, and Simulator Sickness in a Virtual Environment”, in Proc. of IEEE Virtual Reality 2002. IEEE, 2002, str. 164–171.

[22]Tan, C. T., Leong, T. W., Shen, S., Dubravs, C., Si, C., “Exploring Gameplay Experiences on the Oculus Rift”, in Proceedings of the 2015 annual symposium on computer-human interaction in play, 2015, str. 253–263.

[23]Yoo, S., Carter, M., Kay, J., “VRmove: Design Framework for Balancing Enjoyment, Movement and Exertion in VR Games”, in Proc. of the 2018 Annual Symposium on Computer-Human Interaction in Play Companion Extended Abstracts, 2018, str. 295–307.

[24]Foxman, M., Leith, A. P., Beyea, D., Klebig, B., Chen, V. H. H., Ratan, R., “Virtual reality genres: Comparing preferences in immersive experiences and games”, in Extended Abstracts of the 2020 Annual Symposium on Computer-Human Interaction in Play, 2020, str. 237–241.

[25]Hirzle, T., Cordts, M., Rukzio, E., Gugenheimer, J., Bulling, A., “A critical assessment of the use of ssq as a measure of general discomfort in vr head-mounted displays”, in Proceedings of the 2021 ACM CHI Virtual Conference on Human Factors in Computing Systems-CHI, Yokohama, Japan, 2021, str. 8–13.

[26]Souchet, A. D., Lourdeaux, D., Pagani, A., Rebenitsch, L., “A narrative review of immersive virtual reality’s ergonomics and risks at the workplace: cybersickness, visual fatigue, muscular fatigue, acute stress, and mental overload”, Virtual Reality, 2022, str. 1–32.

[27]Raake, A., Egger, S., “Quality and Quality of Experience”, in Qual. of exp. Springer, 2014, str. 11–33.

[28]Vlahovic, S., Suznjevic, M., Skorin-Kapov, L., “A Survey of Challenges and Methods for Quality of Experience Assessment of Interactive vr Applications”, Journal on Multimodal User Interfaces, Vol. 16, No. 3, 2022, str. 257–291.

[29]Vlahovi ć, S., “Evaluacija iskustvene kvalitete višekorisničkih igara temeljenih na virtualnoj stvarnosti”, Doktorski rad, University of Zagreb. Faculty of Electrical Engineering and Computing, 2018.

[30]Vlahovic, S., Skorin-Kapov, L., Suznjevic, M., Pavlin-Bernardic, N., “Not just cybersickness: Short-term effects of popular vr game mechanics on physical discomfort and reaction time”, Virtual Reality (submitted, accepted subject to revision).

[31]Stanney, K. M., Kingdon, K. S., Graeber, D., Kennedy, R. S., “Human Performance in Immersive Virtual Environments: Effects of Exposure Duration, User Control, and Scene Complexity”, Hum. Perf., Vol. 15, No. 4, 2002, str. 339–366.

[32]Brooks, J. O., Goodenough, R. R., Crisler, M. C., Klein, N. D., Alley, R. L., Koon, B. L., Logan Jr, W. C., Ogle, J. H., Tyrrell, R. A., Wills, R. F., “Simulator sickness during

driving simulation studies", Accident Analysis & Prevention, Vol. 42, No. 3, 2010, str. 788–796.

[33]Butler, A. A., Menant, J. C., Tiedemann, A. C., Lord, S. R., "Age and Gender Differences in Seven Tests of Functional Mobility", Journal of neuroengineering and rehabilitation, Vol. 6, 2009, str. 1–9.

[34]Biocca, F., "Will Simulation Sickness Slow Down the Diffusion of Virtual Environment Technology?", Presence: Teleop. & Virt. Envir., Vol. 1, No. 3, 1992, str. 334–343.

[35]Munafo, J., Diedrick, M., Stoffregen, T. A., "The Virtual Reality Head-Mounted Display Oculus Rift Induces Motion Sickness and is Sexist in its Effects", Exp. brain res., Vol. 235, No. 3, 2017, str. 889–901.

[36]Hildebrandt, J., Schmitz, P., Valdez, A. C., Kobbelt, L., Ziefle, M., "Get well soon! human factors' influence on cybersickness after redirected walking exposure in virtual reality", in International Conference on Virtual, Augmented and Mixed Reality. Springer, 2018, str. 82–101.

[37]ITU-T Recomm. G.1035, "Influencing Factors on Quality of Experience for Virtual Reality Services", 2020.

[38]Le Callet, P., Möller, S., Perkis, A. *et al.*, "Qualinet White Paper on Definitions of Quality of Experience", Eur. net. on qual. of exp. in multim. syst. and serv., Vol. 3, No. 2012, 2012.

[39]Chihara, T., Seo, A., "Evaluation of Physical Workload Affected by Mass and Center of Mass of Head-Mounted Display", Applied ergonomics, Vol. 68, 2018, str. 204–212.

[40]Heaney, D., "Data Suggests Oculus Rift S IPD Range 'Best' For Just Half Of Adults", UploadVR, Mar. 2023, dostupno na: https://www.uploadvr.com/data-suggests-oculus-rift-s-ipd-range-best-for-around-half-of-adults/?fbclid=IwAR3gZyO607l7hoC3QsZZdFooKA2LCOIvrZS6YgTvzlU-9tF6LsZgvZh02Ek

[41]Laffont, P.-Y., Martin, T., Gross, M., Tan, W. D., Lim, C., Au, A., Wong, R., "Rectifeye: A Vision-Correcting System for Virtual Reality", in SIGGRAPH ASIA 2016 VR Showcase. Association for Computing Machinery, 2016, str. 1–2.

[42]Suznjevic, M., Mandurov, M., Matijasevic, M., "Performance and qoe assessment of htc vive and oculus rift for pick-and-place tasks in vr", in 2017 Ninth international conference on quality of multimedia experience (QoMEX). IEEE, 2017, str. 1–3.

[43]Gonçalves, G., Monteiro, P., Melo, M., Vasconcelos-Raposo, J., Bessa, M., “A comparative study between wired and wireless virtual reality setups”, IEEE Access, Vol. 8, 2020, str. 29 249–29 258.

[44]Doumanoglou, A., Griffin, D., Serrano, J., Zioulis, N., Phan, T. K., Jiménez, D., Zarpalas, D., Alvarez, F., Rio, M., Daras, P., “Quality of experience for 3-d immersive media streaming”, IEEE Transactions on Broadcasting, Vol. 64, No. 2, 2018, str. 379–391.

[45]Vlahovic, S., Suznjevic, M., Skorin-Kapov, L., “Challenges in Assessing Network Latency Impact on QoE and In-Game Performance in VR First Person Shooter Games”, in ConTEL 2019). IEEE, 2019, str. 1–8.

[46]Concannon, D., Flynn, R., Murray, N., “A quality of experience evaluation system and research challenges for networked virtual reality-based teleoperation applications”, in Proceedings of the 11th ACM Workshop on Immersive Mixed and Virtual Environment Systems, 2019, str. 10–12.

[47]Fan, C.-L., Yen, S.-C., Huang, C.-Y., Hsu, C.-H., “On the optimal encoding ladder of tiled 360 videos for head-mounted virtual reality”, IEEE Transactions on Circuits and Systems for Video Technology, Vol. 31, No. 4, 2020, str. 1632–1647.

[48]Zou, W., Feng, S., Mao, X., Yang, F., Ma, Z., “Enhancing quality of experience for cloud virtual reality gaming: An object-aware video encoding”, in 2021 IEEE International Conference on Multimedia & Expo Workshops (ICMEW). IEEE, 2021, str. 1–6.

[49]Roth, D., Lugrin, J.-L., Galakhov, D., Hofmann, A., Bente, G., Latoschik, M. E., Fuhrmann, A., “Avatar realism and social interaction quality in virtual reality”, in 2016 IEEE Virtual Reality (VR). IEEE, 2016, str. 277–278.

[50]Garau, M., Slater, M., Vinayagamoorthy, V., Brogni, A., Steed, A., Sasse, M. A., “The impact of avatar realism and eye gaze control on perceived quality of communication in a shared immersive virtual environment”, in Proceedings of the SIGCHI conference on Human factors in computing systems, 2003, str. 529–536.

[51]Hvass, J., Larsen, O., Vendelbo, K., Nilsson, N., Nordahl, R., Serafin, S., “Visual realism and presence in a virtual reality game”, in 2017 3DTV Conference: The True Vision-Capture, Transmission and Display of 3D Video (3DTV-CON). IEEE, 2017, str. 1–4.

[52]Vlahovi ć, S., Suznjevic, M., Skorin-Kapov, L., “Subjective assessment of different locomotion techniques in virtual reality environments”, in 2018 Tenth International Conference on Quality of Multimedia Experience (QoMEX). IEEE, 2018, str. 1–3.

[53]Boletsis, C., Cedergren, J. E., “Vr locomotion in the new era of virtual reality: An empirical comparison of prevalent techniques”, Advances in Human-Computer Interaction, Vol. 2019, 2019.

[54]Hameed, A., Perkis, A., “Spatial storytelling: Finding interdisciplinary immersion”, in International Conference on Interactive Digital Storytelling. Springer, 2018, str. 323–332.

[55]Irshad, S., Perkis, A., “Increasing user engagement in virtual reality: The role of interactive digital narratives to trigger emotional responses”, in Proceedings of the 11th Nordic Conference on Human-Computer Interaction: Shaping Experiences, Shaping Society, 2020, str. 1–4.

[56]Paes, D., Irizarry, J., “A usability study of an immersive virtual reality platform for building design review: Considerations on human factors and user interface”, in Construction Research Congress, Vol. 2018, 2018.

[57]Wang, Y., Hu, Y., Chen, Y., “An experimental investigation of menu selection for immersive virtual environments: Fixed versus handheld menus”, Virtual Reality, Vol. 25, No. 2, 2021, str. 409–419.

[58]Hupont, I., Gracia, J., Sanagustin, L., Gracia, M. A., “How do New Visual Immersive Systems Influence Gaming QoE? A Use Case of Serious Gaming with Oculus Rift”, in 2015 7th Int. Work. on Qual. of Multim. Exp. IEEE, 2015, str. 1–6.

[59]Christensen, J. V., Mathiesen, M., Poulsen, J. H., Ustrup, E. E., Kraus, M., “Player Experience in a VR and Non-VR Multiplayer Game”, in Proc. of the Virt. Real. Int. Conf.-Laval Virt. ACM, 2018, str. 10.

[60]Jekosch, U., Voice and Speech Quality Perception: Assessment and Evaluation. Springer Science & Business Media, 2006.

[61]Möller, S., Wältermann, M., Garcia, M.-N., “Features of quality of experience”, in Quality of Experience. Springer, 2014, str. 73–84.

[62]ITU-T Recomm. P.809, “Subjective evaluation Methods for Gaming Quality”, Int’l Telecomm. Union-Telecomm. Standard. Sect., 2018.

[63]Möller, S., Schmidt, S., Beyer, J., “Gaming Taxonomy: An Overview of Concepts and Evaluation Methods for Computer Gaming QoE”, in 2013 5th Int’l Workshop on Qual. of Mult. Exper. (QoMEX). IEEE, 2013, str. 236–241.

[64]Urvoy, M., Barkowsky, M., Le Callet, P., “How visual fatigue and discomfort impact 3d-tv quality of experience: A comprehensive review of technological, psychophysical, and psychological factors”, Annals of Telecommunications-Annales des Télécommunications, Vol. 68, No. 11, 2013, str. 641–655.

[65]Sagnier, C., Loup-Escande, E., Lourdeaux, D., Thouvenin, I., Valléry, G., “User Acceptance of Virtual Reality: An Extended Technology Acceptance Model”, Int’l Jour. of Human–Comp. Interact., 2020, str. 1–15.

[66]Davis, F. D., “Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology”, MIS quart., 1989, str. 319–340.

[67]Hassenzahl, M., “User Experience (UX) Towards an Experiential Perspective on Product Quality”, in Proc. of the 20th Conf. on l’Interact. Homme-Mach., 2008, str. 11–15.

[68]Manis, K. T., Choi, D., “The Virtual Reality Hardware Acceptance Model (VR-HAM): Extending and Individuating the Technology Acceptance Model (TAM) for Virtual Reality Hardware”, Jour. of Bus. Res,, Vol. 100, 2019, str. 503–513.

[69]Kari, T., Kosa, M., “Acceptance and use of virtual reality games: An extension of hmsam”, Virtual Reality, 2023, str. 1–21.

[70]Lowry, P. B., Gaskin, J., Twyman, N., Hammer, B., Roberts, T., “Taking ‘Fun and Games’ Seriously: Proposing the Hedonic-Motivation System Adoption Model (HMSAM)”, Journal of the Association for Information Systems, Vol. 14, No. 11, 2012, str. 617–671.

[71]Shafer, D. M., Carbonara, C. P., Korpi, M. F., “Factors Affecting Enjoyment of Virtual Reality Games: A Comparison Involving Consumer-Grade Virtual Reality Technology”, Games for health journal, Vol. 8, No. 1, 2019, str. 15–23.

[72]Shafer, D. M., Carbonara, C. P., Popova, L., “Spatial Presence and Perceived Reality as Predictors of Motion-Based Video Game Enjoyment”, Presence: Teleoperators and Virtual Environments, Vol. 20, No. 6, 2011, str. 591–619.

[73]Shafer, D. M., Carbonara, C. P., Popova, L., “Controller Required? The Impact of Natural Mapping on Interactivity, Realism, Presence, and Enjoyment in Motion-Based Video Games”, Presence: Teleoperators and Virtual Environments, Vol. 23, No. 3, 2014, str. 267–286.

[74]Shafer, D. M., Carbonara, C. P., “Examining Enjoyment of Casual Videogames”, Games for health journal, Vol. 4, No. 6, 2015, str. 452–459.

[75]Tcha-Tokey, K., Christmann, O., Loup-Escande, E., Loup, G., Richir, S., "Towards a Model of User Experience in Immersive Virtual Environments", Advances in Human-Computer Interaction, Vol. 2018, 2018.

[76]Drachen, A., Mirza-Babaei, P., Nacke, L. E., Games User Research. Oxford Univ. Press, 2018.

[77]Schatz, R., Regal, G., Schwarz, S., Suettc, S., Kempf, M., "Assessing the qoe impact of 3d rendering style in the context of vr-based training", in 2018 Tenth international conference on quality of multimedia experience (QoMEX). IEEE, 2018, str. 1–6.

[78]Steinicke, F., Bruder, G., "A Self-Experimentation Report About Long-Term Use of Fully-Immersive Technology", in Proc. of the 2nd ACM symp. on Spat. user interact., 2014, str. 66–69.

[79]de Gortari, A. B. O., Aronsson, K., Griffiths, M., "Game Transfer Phenomena in Video Game Playing: A Qualitative Interview Study", Int'l Jour. of Cyber Beh., Psych. and Learn., Vol. 1, No. 3, 2011, str. 15–33.

[80]Cantor, J., "Why Horror Doesn't Die: The Enduring and Paradoxical Effects of Frightening Entertainment", in Psych. of entert. Routledge, 2013, str. 333–346.

[81]Lin, J.-H. T., "Fear in Virtual Reality (VR): Fear Elements, Coping Reactions, Immediate and Next-Day Fright Responses Toward a Survival Horror Zombie Virtual Reality Game", Comp. in Hum. Behav., Vol. 72, 2017, str. 350–361.

[82]Heilemann, F., Zimmermann, G., Münster, P., "Accessibility Guidelines for VR games-A Comparison and Synthesis of a Comprehensive Set", Frontiers in Virtual Reality, Vol. 2, 2021, str. 697504.

[83]Oculus, "Oculus, Informative Guides to Help You Design, Develop, and Distribute Your VR App", https://developer.oculus.com/resources/, accessed: 2021-11-05. 2021.

[84]Rolnick, A., Lubow, R., "Why is the driver rarely motion sick? the role of controllability in motion sickness", Ergonomics, Vol. 34, No. 7, 1991, str. 867–879.

[85]Porcino, T. M., Clua, E., Trevisan, D., Vasconcelos, C. N., Valente, L., "Minimizing Cyber Sickness in Head Mounted Display Systems: Design Guidelines and Applications", in 2017 IEEE 5th int. conf. on serious games and appl. for health. IEEE, 2017, str. 1–6.

[86]Jerald, J., The VR Book: Human-Centered Design for Virtual Reality. Morgan & Claypool, 2015.

[87]So, R. H., Lo, W., “Cybersickness with virtual reality training applications: A claustrophobia phenomenon with headmounted displays?”, in First World Congress on Ergonomics For Global Quality and Productivity, Hong Kong, 1998.

[88]Lo, W., So, R. H., “Cybersickness in the Presence of Scene Rotational Movements Along Different Axes”, Applied ergonomics, Vol. 32, No. 1, 2001, str. 1–14.

[89]Terenzi, L., Zaal, P., “Rotational and Translational Velocity and Acceleration Thresholds for the Onset of Cybersickness in Virtual Reality”, in AIAA Scitech 2020 Forum, 2020, str. 0171.

[90]Bonato, F., Bubka, A., Palmisano, S., Phillip, D., Moreno, G., “Vection Change Exacerbates Simulator Sickness in Virtual Environments”, Presence: Teleop. and Virt. Envir., Vol. 17, No. 3, 2008, str. 283–292.

[91]Farmani, Y., Teather, R. J., “Viewpoint Snapping to Reduce Cybersickness in Virtual Reality”, in Proceedings of the 44th Graphics Interface Conference, 2018, str. 168–175.

[92]Draper, M. H., Viirre, E. S., Furness, T. A., Gawron, V. J., “Effects of Image Scale and System Time Delay on Simulator Sickness within Head-Coupled Virtual Environments”, Human factors, Vol. 43, No. 1, 2001, str. 129–146.

[93]Davis, S., Nesbitt, K., Nalivaiko, E., “Comparing the Onset of Cybersickness Using the Oculus Rift and Two Virtual Roller Coasters”, in Proc. of the 11th Australas. Conf. on Interact. Entert. (IE 2015), Vol. 27, 2015, str. 30.

[94]Pouke, M., Tiiro, A., LaValle, S. M., Ojala, T., “Effects of visual realism and moving detail on cybersickness”, in 2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR). IEEE, 2018, str. 665–666.

[95]Rebenitsch, L., “Managing Cybersickness in Virtual Reality”, XRDS: Crossroads, The ACM Magazine for Students, Vol. 22, No. 1, 2015, str. 46–51.

[96]Steed, A., Frlston, S., Lopez, M. M., Drummond, J., Pan, Y., Swapp, D., “An ‘In the Wild’Experiment on Presence and Embodiment Using Consumer Virtual Reality Equipment”, IEEE transact. on visual. and comp. graph., Vol. 22, No. 4, 2016, str. 1406–1414.

[97]Parger, M., Mueller, J. H., Schmalstieg, D., Steinberger, M., “Human upper-body inverse kinematics for increased embodiment in consumer-grade virtual reality”, in Proceedings of the 24th ACM Symposium on Virtual Reality Software and Technology, 2018, str. 1–10.

[98]Erickson, A., Kim, K., Bruder, G., Welch, G. F., “Effects of dark mode graphics on visual acuity and fatigue with virtual reality head-mounted displays”, in 2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR). IEEE, 2020, str. 434–442.

[99]Jaeger, B. K., Mourant, R. R., “Comparison of Simulator Sickness Using Static and Dynamic Walking Simulators”, in Proc. of the Hum. Fact. and Ergon. Soc. Ann. Meet., Vol. 45.27. SAGE Publications Sage CA: Los Angeles, CA, 2001, str. 1896–1900.

[100]Hettinger, L. J., Riccio, G. E., “Visually induced motion sickness in virtual environments”, Presence: Teleoperators & Virtual Environments, Vol. 1, No. 3, 1992, str. 306–310.

[101]Carnegie, K., Rhee, T., “Reducing Visual Discomfort with HMDs Using Dynamic Depth of Field”, IEEE computer graphics and applications, Vol. 35, No. 5, 2015, str. 34–41.

[102]ITU-T Work Item P.IntVR (SG 12), “Subjective Test Methods for Interactive Virtual Reality Applications”, https://www.itu.int/itu-t/workprog/wp_item.aspx?isn=17045, accessed: 2021-11-05. 2021.

[103]Poels, K., de Kort, Y. A., IJsselsteijn, W. A., “D3. 3: Game experience questionnaire: Development of a self-report measure to assess the psychological impact of digital games”, 2007.

[104]Abeele, V. V., Nacke, L. E., Mekler, E. D., Johnson, D., “Design and preliminary validation of the player experience inventory”, in Proceedings of the 2016 Annual Symposium on Computer-Human Interaction in Play Companion Extended Abstracts, 2016, str. 335–341.

[105]Abeele, V. V., Spiel, K., Nacke, L., Johnson, D., Gerling, K., “Development and validation of the player experience inventory: A scale to measure player experiences at the level of functional and psychosocial consequences”, International Journal of Human-Computer Studies, Vol. 135, 2020, str. 102370.

[106]Kourtesis, P., Collina, S., Doumas, L. A., MacPherson, S. E., “Validation of the virtual reality neuroscience questionnaire: Maximum duration of immersive virtual reality sessions without the presence of pertinent adverse symptomatology”, Frontiers in human neuroscience, Vol. 13, 2019, str. 417.

[107]Tcha-Tokey, K., Christmann, O., Loup-Escande, E., Richir, S., “Proposition and validation of a questionnaire to measure the user experience in immersive virtual environments”, The International Journal of Virtual Reality, Vol. 16, 10 2016, str. 33-48.

[108]Witmer, B. G., Singer, M. J., “Measuring Presence in Virtual Environments: A Presence Questionnaire”, Presence, Vol. 7, No. 3, 1998, str. 225–240.

[109]Witmer, B. G., Jerome, C. J., Singer, M. J., “The Factor Structure of the Presence Questionnaire”, Presence: Teleop. & Virt. Envir., Vol. 14, No. 3, 2005, str. 298–312.

[110]Schubert, T., Friedmann, F., Regenbrecht, H., “Decomposing the Sense of Presence: Factor Analytic Insights”, in 2nd Int. Work. on Pres., Vol. 1999, 1999.

[111]Slater, M., McCarthy, J., Maringelli, F., “The influence of body movement on subjective presence in virtual environments”, Human Factors, Vol. 40, No. 3, 1998, str. 469–477.

[112]Usoh, M., Arthur, K., Whitton, M. C., Bastos, R., Steed, A., Slater, M., Brooks Jr, F. P., “Walking > walking-in-place> flying, in virtual environments”, in Proceedings of the 26th Annual Conference on Computer Graphics and Interactive Techniques, 1999, str. 359–364.

[113]Jennett, C., Cox, A. L., Cairns, P., Dhoparee, S., Epps, A., Tijs, T., Walton, A., “Measuring and Defining the Experience of Immersion in Games”, Int’l jour. of hum.-comp. stud., Vol. 66, No. 9, 2008, str. 641–661.

[114]Kennedy, R. S., Lane, N. E., Berbaum, K. S., Lilienthal, M. G., “Simulator sickness questionnaire: An enhanced method for quantifying simulator sickness”, The international journal of aviation psychology, Vol. 3, No. 3, 1993, str. 203–220.

[115]Stone III, W. B., “Psychometric evaluation of the simulator sickness questionnaire as a measure of cybersickness”, Doktorski rad, Iowa State University, 2017.

[116]Kim, H. K., Park, J., Choi, Y., Choe, M., “Virtual Reality Sickness Questionnaire (VRSQ): Motion Sickness Measurement Index in a Virtual Reality Environment”, Appl. Ergon., Vol. 69, 2018, str. 66–73.

[117]Bowman, D. A., Gabbard, J. L., Hix, D., “A Survey of Usability Evaluation in Virtual Environments: Classification and Comparison of Methods”, Presence: Teleop. & Virt. Environ., Vol. 11, No. 4, 2002, str. 404–424.

[118]Regal, G., Schatz, R., Schrammel, J., Suette, S., “Vrate: A unity3d asset for integrating subjective assessment questionnaires in virtual environments”, in 2018 Tenth international conference on quality of multimedia experience (QoMEX). IEEE, 2018, str. 1–3.

[119]Timmerer, C., Ebrahimi, T., Pereira, F., “Toward a New Assessment of Quality”, Computer, Vol. 48, No. 3, 2015, str. 108–110.

[120]Skorin-Kapov, L., Varela, M., Hoßfeld, T., Chen, K.-T., “A Survey of Emerging Concepts and Challenges for QoE Management of Multimedia Services”, ACM Transact. on Multim. Comp., Comm. and Appli., Vol. 14, No. 2s, 2018, str. 29.

[121]Gardlo, B., Egger, S., Hossfeld, T., “Do Scale-Design and Training Matter for Video QoE Assessments Through Crowdsourcing?”, in Proc. of the 4th Int. Work. on Crowdsourc. for Multim. ACM, 2015, str. 15–20.

[122]Clark, L. A., Watson, D., “Constructing Validity: Basic Issues in Objective Scale Development”, Psychol. Assess., Vol. 7, No. 3, 1995, str. 309.

[123]Wilson, G. M., Sasse, M. A., “Do Users Always Know What’s Good for Them? Utilising Physiological Responses to Assess Media Quality”, in Peop. and comp. XIV—Usab. or else! Springer, 2000, str. 327–339.

[124]Bouchard, S., St-Jacques, J., Robillard, G., Renaud, P., “Anxiety increases the feeling of presence in virtual reality”, Presence: Teleoperators and Virtual Environments, Vol. 17, No. 4, 2008, str. 376–391.

[125]Murphy, D., Higgins, C., “Secondary Inputs for Measuring User Engagement in Immersive VR Education Environments”, arXiv preprint arXiv:1910.01586, 2019.

[126]Sheridan, T. B., “Further Musings on the Psychophysics of Presence”, Presence: Teleop. & Virt. Env., Vol. 5, No. 2, 1996, str. 241–246.

[127]Bailenson, J. N., Aharoni, E., Beall, A. C., Guadagno, R. E., Dimov, A., Blascovich, J., “Comparing Behavioral and Self-Report Measures of Embodied Agents’ Social Presence in Immersive Virtual Environments”, in Proc. of the 7th Annual Int’l Work. on Pres., 2004, str. 1864–1105.

[128]Möller, S., Raake, A., Quality of Experience: Advanced Concepts, Applications and Methods. Springer, 2014.

[129]Department of Health, E. *et al.*, “The Belmont Report. Ethical Principles and Guidelines for the Protection of Human Subjects of Research.”, The Jour. of the Amer. Coll. of Dent., Vol. 81, No. 3, 2014, str. 4.

[130]LaViola Jr, J. J., “A Discussion of Cybersickness in Virtual Environments”, ACM Sigchi Bull., Vol. 32, No. 1, 2000, str. 47–56.

[131]Pohlmeyer, A. E., Hecht, M., Blessing, L., “User Experience Lifecycle Model ContinUE [Continuous User Experience]”, Der Mensch im Mittep. techn. Syst. Fortschr.-Berichte VDI Reihe, Vol. 22, 2009, str. 314–317.

[132]Roto, V., Law, E., Vermeeren, A., Hoonhout, J., “User Experience White Paper: Bringing Clarity to the Concept of User Experience”, in Dagstuhl Sem. on Demarcat. User Exper., 2011, str. 12.

[133]Weiss, B., Guse, D., Möller, S., Raake, A., Borowiak, A., Reiter, U., “Temporal Development of Quality of Experience”, in Quality of experience. Springer, 2014, str. 133–147.

[134]Slater, M., Steed, A., “A virtual presence counter”, Presence, Vol. 9, No. 5, 2000, str. 413–434.

[135]Barreda-Ángeles, M., Redondo-Tejedor, R., Pereda-Baños, A., “Psychophysiological methods for quality of experience research in virtual reality systems and applications”, IEEE COMSOC MMTC Commun Front, Vol. 4, No. 1, 2018, str. 14–20.

[136]Stanney, K. M., Hale, K. S., Nahmens, I., Kennedy, R. S., “What to Expect from Immersive Virtual Environment Exposure: Influences of Gender, Body Mass Index, and Past Experience”, Hum. Fact., Vol. 45, No. 3, 2003, str. 504–520.

[137]Karapanos, E., Martens, J.-B., Hassenzahl, M., “Reconstructing Experiences Through Sketching”, arXiv preprint arXiv:0912.5343, 2009.

[138]Wilson, G., McGill, M., “Violent Video Games in Virtual Reality: Re-Evaluating the Impact and Rating of Interactive Experiences”, in Proc. of the 2018 Annual Symposium on Computer-Human Interaction in Play, 2018, str. 535–548.

[139]Bockelman, P., Lingum, D., “Factors of Cybersickness”, in Int. Conf. on Hum.-Comp. Interact. Springer, 2017, str. 3–8.

[140]Leow, F.-T., Ch’ng, E., Zhang, T., Cai, S., See, S., “‘In-the-Wild’ Observation and Evaluation of a Chinese Heritage VR Environment with HTC VIVE”, in Int’l Conf. on Virt. Sys. and Multim., 31 Oct-2 Nov, 2017.

[141]Moustafa, F., Steed, A., “A Longitudinal Study of Small Group Interaction in Social Virtual Reality”, in Proc. of the 24th ACM Symp. on Virt. Real. Softw. and Tech., 2018, str. 1–10.

[142]Mai, C., Wiltzius, T., Alt, F., Hußmann, H., “Feeling alone in public: Investigating the influence of spatial layout on users’ vr experience”, in Proceedings of the 10th Nordic Conference on Human-Computer Interaction, 2018, str. 286–298.

[143]Eghbali, P., Väänänen, K., Jokela, T., “Social acceptability of virtual reality in public spaces: Experiential factors and design recommendations”, in Proceedings of the 18th International Conference on Mobile and Ubiquitous Multimedia, 2019, str. 1–11.

[144]Hossfeld, T., Keimel, C., Timmerer, C., “Crowdsourcing quality-of-experience assessments”, Computer, Vol. 47, No. 9, 2014, str. 98–102.

[145]Madary, M., Metzinger, T. K., “Real Virtuality: A Code of Ethical Conduct. Recommendations for Good Scientific Practice and the Consumers of VR-Technology”, Front. in Robot. and AI, Vol. 3, 2016, str. 3.

[146]Behr, K.-M., Nosper, A., Klimmt, C., Hartmann, T., “Some Practical Considerations of Ethical Issues in VR Research”, Presence, Vol. 14, No. 6, 2005, str. 668–676.

[147]Lewis, C. H., Griffin, M. J., “Human Factors Consideration in Clinical Applications of Virtual Reality”, Studies in health technology and informatics, 1997, str. 35–58.

[148]Fairchild, K. M., Lee, B. H., Loo, J., Ng, H., Serra, L., “The Heaven and Earth Virtual Reality: Designing Applications for Novice Users”, in Proc. of IEEE virt. real. ann. int. symp. IEEE, 1993, str. 47–53.

[149]Henrich, J., Heine, S. J., Norenzayan, A., “The weirdest people in the world?”, Behavioral and Brain Sciences, Vol. 33, No. 2-3, 2010, str. 61–83.

[150]Plechatá, A., Sahula, V., Fayette, D., Fajnerová, I., “Age-related differences with immersive and non-immersive virtual reality in memory assessment”, Frontiers in psychology, Vol. 10, 2019, str. 1330.

[151]Liu, Q., Wang, Y., Tang, Q., Liu, Z., “Do you feel the same as i do? differences in virtual reality technology experience and acceptance between elderly adults and college students”, Frontiers in Psychology, Vol. 11, 2020, str. 2555.

[152]Serino, S., Scarpina, F., Dakanalis, A., Keizer, A., Pedroli, E., Castelnuovo, G., Chirico, A., Catallo, V., Di Lernia, D., Riva, G., “The role of age on multisensory bodily experience: An experimental study with a virtual reality full-body illusion”, Cyberpsychology, Behavior, and Social Networking, Vol. 21, No. 5, 2018, str. 304–310.

[153]Ausburn, L. J., Martens, J., Baukal Jr, C. E., Agnew, I., Dionne, R., Ausburn, F. B., “User characteristics, trait vs. state immersion, and presence in a first-person virtual world”, Journal For Virtual Worlds Research, Vol. 12, No. 3, 2019.

[154]Chiarovano, E., Wang, W., Rogers, S. J., MacDougall, H. G., Curthoys, I. S., De Waele, C., “Balance in virtual reality: Effect of age and bilateral vestibular loss”, Frontiers in neurology, Vol. 8, 2017, str. 5.

[155]Felnhofer, A., Kothgassner, O. D., Beutl, L., Hlavacs, H., Kryspin-Exner, I., “Is Virtual Reality Made for Men Only? Exploring Gender Differences in the Sense of Presence”, Proc. of the Int’l Society on presence research, 2012, str. 103–112.

[156]ITU-T Recomm. G.1035, “Influencing Factors on Quality of Experience for Virtual Reality Services”, Int’l Telecomm. Union-Telecomm. Standard. Sect., 2021.

[157]Nacke, L., Drachen, A., “Towards a Framework of Player Experience Research”, in Proceedings of the second international workshop on evaluating player experience in games at FDG, Vol. 11, 2011.

[158]Engl, S., Nacke, L. E., “Contextual Influences on Mobile Player Experience–A Game User Experience Model”, Entertainment Computing, Vol. 4, No. 1, 2013, str. 83–91.

[159]Hart, S. G., Staveland, L. E., “Development of nasa-tlx (task load index): Results of empirical and theoretical research”, in Advances in Psychology. Elsevier, 1988, Vol. 52, str. 139–183.

[160]Harris, D., Wilson, M., Vine, S., “Development and validation of a simulation workload measure: The simulation task load index (sim-tlx)”, Virtual Reality, Vol. 24, No. 4, 2020, str. 557–566.

[161]Cobb, S. V., Nichols, S., Ramsey, A., Wilson, J. R., “Virtual reality-induced symptoms and effects (vrise)”, Presence: Teleoperators & Virtual Environments, Vol. 8, No. 2, 1999, str. 169–186.

[162]Carroll, J. D., “Individual Differences and Multidimensional Scaling”, Multidimensional Scaling: Theory and Applications in the Behavioral Sciences, Vol. 1, 1972, str. 105–155.

[163]Järvinen, A., Games without frontiers: Theories and methods for game studies and design. Tampere University Press, 2008.

[164]Sicart, M., “Defining game mechanics”, Game Studies, Vol. 8, No. 2, 2008, str. 1–14.

[165]Fabricatore, C., “Gameplay and game mechanics design: a key to quality in videogames”, in OECD-CERI Expert Meeting on Videogames and Education, 2007.

[166]LaViola Jr, J. J., Kruijff, E., McMahan, R. P., Bowman, D., Poupyrev, I. P., 3D User Interfaces: Theory and Practice. Addison-Wesley Professional, 2017.

[167]Smith, R. B., “Experiences with the alternate reality kit: An example of the tension between literalism and magic”, ACM SIGCHI Bulletin, Vol. 17, No. SI, 1986, str. 61–67.

[168]Hale, K. S., Stanney, K. M., Handbook of Virtual Environments: Design, Implementation, and Applications. CRC Press, 2014.

[169]Bowman, D. A., McMahan, R. P., Ragan, E. D., “Questioning naturalism in 3d user interfaces”, Communications of the ACM, Vol. 55, No. 9, 2012, str. 78–88.

[170]Guiard, Y., “Asymmetric division of labor in human skilled bimanual action: The kinematic chain as a model”, Journal of Motor Behavior, Vol. 19, No. 4, 1987, str. 486–517.

[171]Canzoneri, E., Ubaldi, S., Rastelli, V., Finisguerra, A., Bassolino, M., Serino, A., “Tool-use reshapes the boundaries of body and peripersonal space representations”, Experimental brain research, Vol. 228, No. 1, 2013, str. 25–42.

[172]Bufacchi, R. J., Iannetti, G. D., “An action field theory of peripersonal space”, Trends in cognitive sciences, Vol. 22, No. 12, 2018, str. 1076–1090.

[173]Holmes, N. P., “Does tool use extend peripersonal space? a review and re-analysis”, Experimental brain research, Vol. 218, No. 2, 2012, str. 273–282.

[174]Bunn, J. W., Scientific Principles of Coaching. Prentice Hall, 1972.

[175]Ellenbecker, T. S., Aoki, R., “Step by step guide to understanding the kinetic chain concept in the overhead athlete”, Current Reviews in Musculoskeletal Medicine, Vol. 13, No. 2, 2020, str. 155–163.

[176]Fletcher, E., “Sword Mechanics for VR”, dostupno na: https://evanfletcher42.com/2018/12/29/sword-mechanics-for-vr [Online; accessed 24. Feb. 2022]. 12 2018.

[177]Zindulka, T., Bachynskyi, M., Müller, J., “Performance and experience of throwing in virtual reality”, in Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems, 2020, str. 1–8.

[178]Vlahovic, S., Suznjevic, M., Pavlin-Bernardic, N., Skorin-Kapov, L., “The effect of vr gaming on discomfort, cybersickness, and reaction time”, in 2021 13th International Conference on Quality of Multimedia Experience (QoMEX). IEEE, 2021, str. 163–168.

[179]Reason, J. T., Brand, J. J., Motion Sickness. Acad. Press, 1975.

[180]Rebenitsch, L., Owen, C., “Review on Cybersickness in Applications and Visual Displays”, Virt. Real., Vol. 20, No. 2, 2016, str. 101–125.

[181]Chang, E., Kim, H. T., Yoo, B., “Virtual Reality Sickness: A Review of Causes and Measurements”, Int’l Journal of Human–Computer Interaction, Vol. 36, No. 17, 2020, str. 1658–1682.

[182]Stanney, K. M., Kennedy, R. S., Drexler, J. M., “Cybersickness is not simulator sickness”, Proceedings of the Human Factors and Ergonomics Society annual meeting, Vol. 41, No. 2, 1997, str. 1138–1142.

[183]Sevinc, V., Berkman, M. I., “Psychometric evaluation of simulator sickness questionnaire and its variants as a measure of cybersickness in consumer virtual environments”, Applied ergonomics, Vol. 82, 2020, str. 102958.

[184]Boring, S., Jurmu, M., Butz, A., “Scroll, tilt or move it: Using mobile phones to continuously control pointers on large public displays”, in Proceedings of the 21st Annual Conference of the Australian Computer-Human Interaction Special Interest Group: Design: Open 24/7, 2009, str. 161–168.

[185]Hincapié-Ramos, J. D., Guo, X., Moghadasian, P., Irani, P., “Consumed endurance: A metric to quantify arm fatigue of mid-air interactions”, in Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, 2014, str. 1063–1072.

[186]Hansberger, J. T., Peng, C., Mathis, S. L., Areyur Shanthakumar, V., Meacham, S. C., Cao, L., Blakely, V. R., “Dispelling the gorilla arm syndrome: The viability of prolonged gesture interactions”, in International conference on virtual, augmented and mixed reality. Springer, 2017, str. 505–520.

[187]Penumudi, S. A., Kuppam, V. A., Kim, J. H., Hwang, J., “The effects of target location on musculoskeletal load, task performance, and subjective discomfort during virtual reality interactions”, Applied ergonomics, Vol. 84, 2020, str. 103010.

[188]Evans, E., Naugle, K. E., Kaleth, A. S., Arnold, B., Naugle, K. M., “Physical activity intensity, perceived exertion, and enjoyment during head-mounted display virtual reality games”, Games for Health Journal, Vol. 10, No. 5, 2021, str. 314–320.

[189]Knight, J. F., Baber, C., “Effect of head-mounted displays on posture”, Human Factors, Vol. 49, No. 5, 2007, str. 797–807.

[190]Baur, D., Pfeifle, C., Heyde, C., “Cervical spine injury after virtual reality gaming: A case report”, Journal of Medical Case Reports, Vol. 15, No. 1, 2021, str. 1–4.

[191]Nalivaiko, E., Davis, S. L., Blackmore, K. L., Vakulin, A., Nesbitt, K. V., “Cybersickness provoked by head-mounted display affects cutaneous vascular tone, heart rate and reaction time”, Physiology & Behavior, Vol. 151, 2015, str. 583–590.

[192]Nesbitt, K., Davis, S., Blackmore, K., Nalivaiko, E., “Correlating reaction time and nausea measures with traditional measures of cybersickness”, Displays, Vol. 48, 2017, str. 1–8.

[193]Mittelstaedt, J. M., Wacker, J., Stelling, D., “Vr aftereffect and the relation of cybersickness and cognitive performance”, Virtual Reality, Vol. 23, No. 2, 2019, str. 143–154.

[194]Fernandez-Ruiz, J., Wong, W., Armstrong, I. T., Flanagan, J. R., “Relation between reaction time and reach errors during visuomotor adaptation”, Behavioural brain research, Vol. 219, No. 1, 2011, str. 8–14.

[195]Szpak, A., Michalski, S. C., Saredakis, D., Chen, C. S., Loetscher, T., “Beyond feeling sick: The visual and cognitive aftereffects of virtual reality”, IEEE Access, Vol. 7, 2019, str. 130 883–130 892.

[196]Szpak, A., Michalski, S. C., Loetscher, T. *et al.*, “Exergaming with beat saber: An investigation of virtual reality aftereffects”, Journal of Medical Internet Research, Vol. 22, No. 10, 2020, str. e19840.

[197]Meta, “View comfort ratings for Meta Quest store content”, dostupno na: https://www.meta.com/help/quest/articles/accounts/purchasing-apps/oculus-store-comfort-rating [Online; accessed 18. Feb. 2023]. Feb. 2023.

[198]The VR Shop, “Comfort Rating Explained - Green, Yellow, Red”, dostupno na: https://www.virtual-reality-shop.co.uk/comfort-rating-explained [Online; accessed 18. Feb. 2023]. Feb. 2023.

[199]Borg, G., Borg’s Perceived exertion and pain scales. Human Kinetics, 1998.

[200]Deary, I. J., Liewald, D., Nissan, J., “A free, easy-to-use, computer-based simple and four-choice reaction time programme: The deary-liewald reaction time task”, Behavior Research Methods, Vol. 43, No. 1, 2011, str. 258–268.

[201]Hart, S. G., “NASA-task load index (NASA-TLX); 20 years later”, in Proc. of the human factors and ergonomics society annual meeting, Vol. 50, No. 9. Sage CA: CA, 2006, str. 904–908.

[202]Jager, J., Putnick, D. L., Bornstein, M. H., “Ii. more than just convenient: The scientific merits of homogeneous convenience samples”, Monographs of the Society for Research in Child Development, Vol. 82, No. 2, 2017, str. 13–30.

[203]Prabaswari, A. D., Basumerda, C., Utomo, B. W., “The mental workload analysis of staff in study program of private educational organization”, in IOP Conference Series: Materials Science and Engineering, Vol. 528, No. 1. IOP Publishing, 2019, str. 012018.

[204]Triberti, S., Repetto, C., Riva, G., “Psychological factors influencing the effectiveness of virtual reality–based analgesia: A systematic review”, Cyberpsychology, Behavior, and Social Networking, Vol. 17, No. 6, 2014, str. 335–345.

[205]Evans, E., Naugle, K. E., Ovispo, A., Kaleth, A. S., Arnold, B., Naugle, K. M., “Active virtual reality games reduce pain sensitivity in young, healthy adults”, Frontiers in Virtual Reality, 2021, str. 150.

[206]Vlahovic, S., Skorin-Kapov, L., Car, Z., “An initiative toward an enhanced industry-reported comfort, accessibility, and safety rating system for vr applications”, in 2023 15th International Conference on Quality of Multimedia Experience (QoMEX). IEEE, 2023, str. 95–98.

[207]Bimberg, P., Weissker, T., Kulik, A., “On the usage of the simulator sickness questionnaire for virtual reality research”, in 2020 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW). IEEE, 2020, str. 464–467.

[208]Gianaros, P. J., Muth, E. R., Mordkoff, J. T., Levine, M. E., Stern, R. M., “A Questionnaire for the Assessment of the Multiple Dimensions of Motion Sickness”, Aviation, space, and environmental medicine, Vol. 72, No. 2, 2001, str. 115.

[209]Szpak, A., Richards, A., Michalski, S. C., Loetscher, T., “Getting the most out of virtual reality: Evaluating short breaks to reduce cybersickness and cognitive aftereffects”, in 2022 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW). IEEE, 2022, str. 533–537.

[210]Dye, M. W., Green, C. S., Bavelier, D., “Increasing speed of processing with action video games”, Current directions in psychological science, Vol. 18, No. 6, 2009, str. 321–326.

[211]Audiffren, M., Tomporowski, P. D., Zagrodnik, J., “Acute aerobic exercise and information processing: Energizing motor processes during a choice reaction time task”, Acta Psychologica, Vol. 129, No. 3, 2008, str. 410–419.

[212]Wickens, C. D., Helleberg, J., Goh, J., Xu, X., Horrey, W. J., “Pilot task management: Testing an attentional expected value model of visual scanning”, Savoy, IL, UIUC Institute of Aviation Technical Report, 2001.

[213]Wickens, C. D., Goh, J., Helleberg, J., Horrey, W. J., Talleur, D. A., “Attentional models of multitask pilot performance using advanced display technology”, Human Factors, Vol. 45, No. 3, 2003, str. 360–380.

[214]Wickens, C. D., Noticing Events in the Visual Workplace: The SEEV and NSEEV Models, ser. Cambridge Handbooks in Psychology, 2015, str. 749–768.

[215]Wickens, C. D., Helton, W. S., Hollands, J. G., Banbury, S., Engineering Psychology and Human Performance. Routledge, 2021.

[216]Fitts, P. M., "The information capacity of the human motor system in controlling the amplitude of movement.", Journal of Experimental Psychology, Vol. 47, No. 6, 1954, str. 381.

[217]Murata, A., Iwase, H., "Extending fitts' law to a three-dimensional pointing task", Human Movement Science, Vol. 20, No. 6, 2001, str. 791–805.

[218]Clark, L. D., Bhagat, A. B., Riggs, S. L., "Extending fitts' law in three-dimensional virtual environments with current low-cost virtual reality technology", International Journal of Human-Computer Studies, Vol. 139, 2020, str. 102413.

[219]Poupyrev, I., Weghorst, S., Billinghurst, M., Ichikawa, T., "A framework and testbed for studying manipulation techniques for immersive vr", in Proceedings of the ACM Symposium on Virtual reality Software and Technology, 1997, str. 21–28.

[220]Sermarini, J., Kider Jr, J. T., LaViola Jr, J. J., McConnell, D. S., "A kinematic evaluation of linear and parabolic pointing in virtual reality", in Proceedings of the Human Factors and Ergonomics Society Annual Meeting, Vol. 65, No. 1. SAGE Publications Sage CA: Los Angeles, CA, 2021, str. 86–90.

[221]Vicencio-Moreira, R., Mandryk, R. L., Gutwin, C., Bateman, S., "The effectiveness (or lack thereof) of aim-assist techniques in first-person shooter games", in Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, 2014, str. 937–946.

[222]Kocur, M., Dechant, M. J., Lankes, M., Wolff, C., Mandryk, R., "Eye caramba: Gaze-based assistance for virtual reality aiming and throwing tasks in games", in ACM Symposium on Eye Tracking Research and Applications, 2020, str. 1–6.

[223]Feick, M., Kleer, N., Tang, A., Krüger, A., "The virtual reality questionnaire toolkit", in Adjunct Publication of the 33rd Annual ACM Symposium on User Interface Software and Technology, 2020, str. 68–69.

[224]Hameed, A., Perkis, A., Möller, S., "Evaluating hand-tracking interaction for performing motor-tasks in vr learning environments", in 2021 13th International Conference on Quality of Multimedia Experience (QoMEX). IEEE, 2021, str. 219–224.

[225]Matokanovi ć, M., "Quality of experience assessment of interaction mechanics in virtual reality", Doktorski rad, University of Zagreb. Faculty of Electrical Engineering and Computing, 2022.

[226]Vlahovic, S., Suznjevic, M., Skorin-Kapov, L., “The impact of network latency on gaming qoe for an fps vr game”, in 11th Intl. Conf. on Quality of Multimedia Experience (QoMEX). IEEE, 2019, str. 1–3.

[227]Slivar, I., Vlahovic, S., Silic, M., Skorin-Kapov, L., Suznjevic, M., “The impact of network and social context on quality of experience for competitive multiplayer virtual reality games”, in Proceedings of the 2nd Workshop on Games Systems, 2022, str. 16–21.

[228]Claypool, M., Claypool, K., “Latency and player actions in online games”, Communications of the ACM, Vol. 49, No. 11, 2006, str. 40–45.

[229]Quax, P., Monsieurs, P., Lamotte, W., De Vleeschauwer, D., Degrande, N., “Objective and subjective evaluation of the influence of small amounts of delay and jitter on a recent first person shooter game”, in Proceedings of 3rd ACM SIGCOMM workshop on Network and system support for games, 2004, str. 152–156.

[230]Beigbeder, T., Coughlan, R., Lusher, C., Plunkett, J., Agu, E., Claypool, M., “The Effects Of Loss and Latency on User Performance in Unreal Tournament 2003®”, in Proceedings of 3rd ACM SIGCOMM workshop on Network and system support for games. ACM, 2004, str. 144–151.

[231]Dick, M., Wellnitz, O., Wolf, L., “Analysis of factors affecting players’ performance and perception in multiplayer games”, in Proc. of 4th ACM SIGCOMM workshop on Network and system support for games, ser. NetGames, 2005, str. 1–7.

[232]Claypool, M., Claypool, K., “Latency Can Kill: Precision and Deadline in Online Games”, in Proc. of the 1st Ann. ACM SIGMM Conf. on Multimedia Systems. ACM, 2010, str. 215–222.

[233]Raaen, K., Kjellmo, I., “Measuring latency in virtual reality systems”, in International Conference on Entertainment Computing. Springer, 2015, str. 457–462.

[234]Lincoln, P., Blate, A., Singh, M., Whitted, T., State, A., Lastra, A., Fuchs, H., “From motion to photons in 80 microseconds: Towards minimal latency for virtual and augmented reality”, IEEE transactions on visualization and computer graphics, Vol. 22, No. 4, 2016, str. 1367–1376.

[235]Kijima, R., Miyajima, K., “Measurement of head mounted display’s latency in rotation and side effect caused by lag compensation by simultaneous observation—an example result using oculus rift dk2”, in 2016 IEEE virtual reality (VR). IEEE, 2016, str. 203–204.

[236]Zhao, J., Allison, R. S., Vinnikov, M., Jennings, S., “Estimating the motion-to-photon latency in head mounted displays”, in 2017 IEEE Virtual Reality (VR). IEEE, 2017, str. 313–314.

[237]Choi, S.-W., Lee, S., Seo, M.-W., Kang, S.-J., “Time sequential motion-to-photon latency measurement system for virtual reality head-mounted displays”, Electronics, Vol. 7, No. 9, 2018, str. 171.

[238]Stauffert, J.-P., Niebling, F., Latoschik, M. E., “Simultaneous run-time measurement of motion-to-photon latency and latency jitter”, in 2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR). IEEE, 2020, str. 636–644.

[239]Kojic, T., Schmidt, S., Möller, S., Voigt-Antons, J.-N., “Influence of network delay in virtual reality multiplayer exergames: Who is actually delayed?”, in 2019 11th Intl. Conf. on Quality of Multimedia Experience (QoMEX). IEEE, 2019, str. 1–3.

[240]Vaghi, I., Greenhalgh, C., Benford, S., “Coping with inconsistency due to network delays in collaborative virtual environments”, in Proceedings of the ACM symposium on Virtual reality software and technology, 1999, str. 42–49.

[241]Liu, S., Xu, X., Claypool, M., “A survey and taxonomy of latency compensation techniques for network computer games”, ACM Computing Surveys (CSUR), 2022.

[242]Anderson-Hanley, C., Snyder, A. L., Nimon, J. P., Arciero, P. J., “Social facilitation in virtual reality-enhanced exercise: Competitiveness moderates exercise effort of older adults”, Clinical interventions in aging, 2011, str. 275–280.

[243]Ventura, R. B., Richmond, S., Nadini, M., Nakayama, S., Porfiri, M., “Does winning or losing change players’ engagement in competitive games? experiments in virtual reality”, IEEE Transactions on Games, Vol. 13, No. 1, 2019, str. 23–34.

[244]ITU-T Recomm. G.1032, “Influencing Factors on Gaming Quality of Experience”, Int’l Telecomm. Union-Telecomm. Standard. Sect., 2017.

[245]Wood, G., Wright, D. J., Harris, D., Pal, A., Franklin, Z. C., Vine, S. J., “Testing the Construct Validity of a Soccer-Specific Virtual Reality Simulator Using Novice, Academy, and Professional Soccer Players”, Virtual Reality, Vol. 25, 2021, str. 43–51.

[246]Michalski, S. C., Szpak, A., Saredakis, D., Ross, T. J., Billinghurst, M., Loetscher, T., “Getting your game on: Using virtual reality to improve real table tennis skills”, PloS one, Vol. 14, No. 9, 2019, str. e0222351.

[247]Schlickum, M. K., Hedman, L., Enochsson, L., Kjellin, A., Felländer-Tsai, L., "Systematic Video Game Training in Surgical Novices Improves Performance in Virtual Reality Endoscopic Surgical Simulators: A Prospective Randomized Study", World journal of surgery, Vol. 33, 2009, str. 2360–2367.

[248]Rivu, R., Zhou, Y., Welsch, R., Mäkelä, V., Alt, F., "When friends become strangers: Understanding the influence of avatar gender on interpersonal distance in virtual reality", in Human-Computer Interaction–INTERACT 2021: 18th IFIP TC 13 International Conference, Bari, Italy, August 30–September 3, 2021, Proceedings, Part V. Springer, 2021, str. 234–250.

[249]Vella, K., Klarkowski, M., Johnson, D., Hides, L., Wyeth, P., "The social context of video game play: Challenges and strategies", in Proceedings of the 2016 ACM Conference on Designing Interactive Systems, 2016, str. 761–772.

[250]Steed, A., Spante, M., Heldal, I., Axelsson, A.-S., Schroeder, R., "Strangers and friends in caves: an exploratory study of collaboration in networked ipt systems for extended periods of time", in Proceedings of the 2003 symposium on Interactive 3D graphics, 2003, str. 51–54.

[251]Koji ć, T., Voigt-Antons, J.-N., Schmidt, S., Tetzlaff, L., Kortowski, B., Sirotina, U., Möller, S., "Influence of virtual environments and conversations on user engagement during multiplayer exergames", in 2018 10th Intl. Conf. on Quality of Multimedia Experience (QoMEX). IEEE, 2018, str. 1–3.

[252]Liszio, S., Emmerich, K., Masuch, M., "The influence of social entities in virtual reality games on player experience and immersion", in Proceedings of the 12th international conference on the foundations of digital games, 2017, str. 1–10.

[253]Mayor, J., Raya, L., Sanchez, A., "A comparative study of virtual reality methods of interaction and locomotion based on presence, cybersickness and usability", IEEE Transactions on Emerging Topics in Computing, 2019.

[254]Schmidt, S., "Assessing the quality of experience of cloud gaming services", Doktorski rad, Technische Universität Berlin, 2021.

[255]Hudson, S., Matson-Barkat, S., Pallamin, N., Jegou, G., "With or without you? interaction and immersion in a virtual reality experience", Journal of Business Research, Vol. 100, 2019, str. 459–468.

[256]Mueller, F., Gibbs, M., Vetere, F., Agamanolis, S., “Design space of networked exertion games demonstrated by a three-way physical game based on table tennis”, Computers in Entertainment (CIE), Vol. 6, No. 3, 2008, str. 1–31.

[257]Cechanowicz, J. E., Gutwin, C., Bateman, S., Mandryk, R., Stavness, I., “Improving player balancing in racing games”, in Proc. of the first ACM SIGCHI annual symposium on Computer-human interaction in play, 2014, str. 47–56.

[258]Houston, J., Harris, P., McIntire, S., Francis, D., “Revising the competitiveness index using factor analysis”, Psychological Reports, Vol. 90, No. 1, 2002, str. 31–34.

[259]Donnellan, M. B., Oswald, F. L., Baird, B. M., Lucas, R. E., “The mini-ipip scales: tiny-yet-effective measures of the big five factors of personality.”, Psychological assessment, Vol. 18, No. 2, 2006, str. 192.

[260]Liu, S., Claypool, M., Kuwahara, A., Scovell, J., Sherman, J., “The effects of network latency on competitive first-person shooter game players”, in 2021 13th International Conference on Quality of Multimedia Experience (QoMEX). IEEE, 2021, str. 151–156.

[261]Moller, S., Schmidt, S., Zadtootaghaj, S., “New itu-t standards for gaming qoe evaluation and management”, in 2018 Tenth International Conference on Quality of Multimedia Experience (QoMEX). IEEE, 2018, str. 1–6.

[262]Xu, J., Wah, B. W., “Consistent synchronization of action order with least noticeable delays in fast-paced multiplayer online games”, ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM), Vol. 13, No. 1, 2016, str. 1–25.

[263]Deleuze, J., Christiaens, M., Nuyens, F., Billieux, J., “Shoot at first sight! first person shooter players display reduced reaction time and compromised inhibitory control in comparison to other video game players”, Computers in Human Behavior, Vol. 72, 2017, str. 570–576.

[264]Peng, W., Hsieh, G., “The influence of competition, cooperation, and player relationship in a motor performance centered computer game”, Computers in Human Behavior, Vol. 28, No. 6, 2012, str. 2100–2106.

List of Figures

List of Tables

Biography

Sara Vlahović, born in Zagreb, Croatia, earned her bachelor's degree in Computing and master's degree in Information and Communication Technology from the Faculty of Electrical Engineering and Computing (FER) at the University of Zagreb. Starting in September 2018, she has taken on the role of a research assistant at the Department of Telecommunications at FER. She participated in several research projects and serves as a teaching assistant for two faculty courses.

In 2018, she enrolled in the PhD program at the Faculty of Electrical Engineering and Computing, field of Computing, under the mentorship of Prof. Lea Skorin-Kapov. She completed her PhD qualifying exam in October 2019 and defended her doctoral dissertation topic in April 2021. Conducted as part of the activities at The Multimedia Quality of Experience Research Lab (MUEXlab), her research is centered on evaluating the Quality of Experience of Virtual Reality gaming.

She authored or co-authored three papers published in international journals, one paper currently under review in an international journal, and 10 papers in conference proceedings. She also served as a co-author of the QUALINET White Paper on Definitions of Immersive Media Experience (IMEx). Currently, she is actively involved as one of the contributors to the ITU-T work item P.IntVR entitled "Subjective Test Method for Interactive Virtual Reality Applications". She has reviewed papers for several scientific journals.

List of publications

Journal papers

Published journal papers

1. **Vlahović, S.**, Slivar, I., Šilić, M., Skorin-Kapov, L., Sužnjević, M., Exploring the Facets of the Multiplayer VR Gaming Experience, ACM Transactions on Multimedia Computing, Communications, and Applications (accepted).
2. **Vlahović, S.**, Sužnjević, M., Skorin-Kapov, L., A framework for the classification and evaluation of game mechanics for virtual reality games, Electronics, Vol. 11 (18), 2022,

2946.

3. **Vlahović, S.**, Sužnjević, M., Skorin-Kapov, L., A survey of challenges and methods for Quality of Experience assessment of interactive VR applications, Journal on Multimodal User Interfaces, Vol. 16 (3), 2022, pp. 257–291.

Journal papers under review

1. **Vlahović, S.**, Skorin-Kapov, L., Sužnjević, M., Pavlin-Bernardić, N., Not just cybersickness: Short-term effects of popular VR game mechanics on physical discomfort and reaction time, Virtual Reality (submitted, accepted subject to revision).

Conference papers

1. Moslavac, M., Brzica, L., Drozd, L., Kušurin, N., **Vlahović, S.**, Skorin-Kapov, L., Assessment of Varied User Representations and XR Environments in Consumer-Grade XR Telemeetings, 17th International Conference on Telecommunications (ConTEL), 2023.
2. **Vlahović, S.**, Skorin-Kapov, L., Car, Ž., An initiative toward an enhanced industry-reported comfort, accessibility, and safety rating system for VR applications, 15th International Conference on Quality of Multimedia Experience (QoMEX), pp. 95–98, 2023.
3. Slivar, I., **Vlahović, S.**, Silić, M., Skorin-Kapov, L., Sužnjević, M., The impact of network and social context on Quality of Experience for competitive multiplayer virtual reality games, Proceedings of the 2nd Workshop on Games Systems, in conjunction with MMSys, 16–21, 2022.
4. **Vlahović, S.**, Skorin-Kapov, L., QoE Assessment for Virtual Reality Gaming, PhD Forum–Book of Abstracts, 2021.
5. **Vlahović, S.**, Sužnjević, M., Pavlin-Bernardić, N.,Skorin-Kapov, L., The effect of VR gaming on discomfort, cybersickness, and reaction time, 13th International Conference on Quality of Multimedia Experience (QoMEX), pp. 163–168, 2021.
6. **Vlahović, S.**, Manđurov, M., Sužnjević, M., Skorin-Kapov, L., Usability assessment of a wearable video-communication system, 12th International Conference on Quality of Multimedia Experience (QoMEX), 2020.
7. **Vlahović, S.**, Skorin-Kapov, L., QoE Assessment for Interactive Immersive AR/VR Applications, SoftCOM 2019 PhD Forum, Split, pp. 14–15, 2019.
8. **Vlahović, S.**, Sužnjević, M., Skorin-Kapov, L., Challenges in assessing network latency impact on QoE and in-game performance in VR first person shooter games, 15th International Conference on Telecommunications (ConTEL), 1–8, 2019.
9. **Vlahović, S.**, Sužnjević, M., Skorin-Kapov, L., The impact of network latency on gaming QoE for an FPS VR game, 11th international Conference on Quality of Multimedia

Experience (QoMEX), 2019.

10. **Vlahović, S.**, Sužnjević, M., Skorin-Kapov, L., Subjective assessment of different locomotion techniques in virtual reality environments, 10th International Conference on Quality of Multimedia Experience (QoMEX), pp. 177–179, 2018.

QUALINET White Paper

1. Timmerer, C., Barakovi ć, S., Baraković Husić, J., Bech, S., Bosse, S., Botev, J., Brunnström, K., Cruz, L., De Moor, K., de Polo Saibanti, A., Durnez, W., Egger-Lampl, S., Engelke, U., Falk, T.H., Hameed, A., Hines, A., Kojić, T., Kukolj, D., Liotou, E., Milovanović, D., Möller, S., Murray, N., Naderi, B., Pereira, M., Perry, S., Pinheiro, A., Pinilla, A., Raake, A., Agrawal, SR., Reiter, U., Rodrigues, R., Schatz, R., Schelkens, P., Schmidt, S., Sabet, S.S., Singla, A., Skorin-Kapov, L., Sužnjević, M., Uhrig, S., **Vlahović, S.**, Voigt-Antons, J.-N., Zadtootaghaj, S., QUALINET White Paper on Definitions of Immersive Media Experience (IMEx), European Network on Quality of Experience in Multimedia Systems and Services, 14th QUALINET meeting (online), May 25, 2020. Online: https://arxiv.org/abs/2007.070322020

Životopis

Sara Vlahović rođena je u Zagrebu, gdje je završila preddiplomski studij Računarstvo i diplomski studij Informacijske i komunikacijske tehnologije na Fakultetu elektrotehnike i računarstva (FER) Sveučilišta u Zagrebu. Od rujna 2018. je zaposlena kao asistentica na Zavodu za telekomunikacije FER-a. Sudjelovala je na nekoliko projekata te drži vježbe na jednom preddiplomskom i jednom diplomskom kolegiju.

2018. godine upisala je doktorski studij na Fakultetu elektrotehnike i računarstva, smjer Računarstvo, pod mentorstvom prof. dr. sc. Lee Skorin-Kapov. Položila je kvalifikacijski doktorski ispit u listopadu 2019. i obranila temu doktorske disertacije u travnju 2021. godine. U sklopu aktivnosti Laboratorija za istraživanje multimedijske kvalitete iskustva (MUEXlab) provodi istraživanja usredotočena na procjenu kvalitete iskustva igranja igara u virtualnoj stvarnosti.

Autorica ili koautorica je tri znanstvena rada objavljena u međunarodnim časopisima, jednog znanstvenog rada koji je trenutno u postupku recenzije u međunarodnom časopisu te deset radova u zbornicima skupova. Sudjelovala je na pisanju dokumenta QUALINET White Paper on Definitions of Immersive Media Experience (IMEx), a trenutno aktivno sudjeluje na izradi radne točke ITU-T P.IntVR pod nazivom "Subjective Test Method for Interactive Virtual Reality Applications". Recenzirala je radove za više znanstvenih časopisa.