2009

Presence-dependent Performance Differences Between Virtual Simulations And Miniature Worlds

Andre Huthmann
University of Central Florida

PRESENCE-DEPENDENT PERFORMANCE DIFFERENCES BETWEEN VIRTUAL SIMULATIONS AND MINIATURE WORLDS

by

ANDRE HUTHMANN
M.S. University of Central Florida, 2001

A dissertation submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy
in the Department of Industrial Engineering and Management Systems
in the College of Engineering and Computer Science
at the University of Central Florida,
Orlando, Florida

Spring Term 2009

Major Professors:
Florian Jentsch
Linda Malone

ABSTRACT

The purpose of simulation is to avoid reality-based constraints by the implementation of a synthetic model. Based on this advantage, interactive simulations have conquered all areas of applications from acquisition, and training, to research. Simulation results are transferred in many ways into reality and conclusions are drawn from the simulation to the application.

Many anecdotal observations on human-in-the-loop simulations have shown a significant difference in actor behavior between simulations and reality-based applications. It seems that the factors that makes simulation so attractive, namely the absence of constraints and especially of imminent danger for persons and equipment, influence the behavior and thereby the performance of the user. These differences between simulation and reality may lead to false conclusions based on simulation results.

The concept of perceiving a simulation as ‘real’ and of ‘being in’ the simulation is called ‘sense of presence’. This psychological construct can also be described as ‘level of disbelief’ towards the simulation. Hence, differences in behavior are based on such user’s assessment of a simulation and subsequently are supposed to be mediated by a difference in presence.

This research established significant differences in presence and performance between a simulation and a miniature-world teleoperation task. Presence and performance changed in identical tasks due to the application type and the connected danger to the robot. Also, the results supported a negative relationship between presence and performance: presence increased in the miniature-world and affected performance so

that performance decreased. The causal relationship of application type$\rightarrow$ presence$\rightarrow$ performance was established and demands the examination of simulation based results with respect to the perceived danger to equipment, before they are transferred into the real application.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

DES	Dissociative Experience Scale
EF	Effort (PQ)
EXPL	Exploration of VE (IPQ)
FR	Frustration Level (PQ)
HR	Average Heart Rate
INV	Involvement (IPQ)
IPQ	igroup presence questionnaire
ITQ	Immersive Tendencies Questionnaire
Maj/Errors	Falls off the table within time frame
MART	Malleable Attentional Resource Theory
MD	Mental Demand (PQ)
MiE/MiD	Miniature World easy/difficult task
NASA TLX	NASA Task Load Index
OP	Own Performance (PQ)
PD	Physical Demand (PQ)
PQ	Presence Questionnaire developed by Witmer & Singer
PRED	Predictability & Interaction (IPQ)
REAL	Realness (IPQ)
Runs	Number of laps within 5 minutes
SA	Situation Awareness
SAGAT	Situation Awareness Global Assessment Technique
SP	Spatial Presence (IPQ)
Sp/Speed	Number of laps within operational time
SSQ	Simulator Sickness Questionnaire
SUS	Presence Questionnaire developed by Slater et al.

SWAT	Subjective Workload Assessment Technique
TAS (TAQ)	Tellegen Absorption Scale (Questionnaire)
TD	Temporal Demand (PQ)
TPL	Team Performance Laboratory
UCF	University of Central Florida
UGV	Unmanned Ground Vehicle
VE	Virtual Environment
VEE/VED	Simulations easy/difficult task
WP	Workload Profile

CHAPTER 1: INTRODUCTION

1.1 Research Problem and Contextual Background

1.1.1 Performance-Affecting Factors in Virtual Environments

Differences in performance between human-in-the-loop simulations and their real counterparts are mostly undesired. In many cases, these differences are related to monetary and technical constraints resulting in a lowered fidelity of the simulation. But even in accepted high fidelity simulations, there seems to be a resistant performance difference:

- Pilots found simulations, which were faster than normal time (up to 1.75), more realistic than real time simulations (Kaber, Draper, & Usher, 2002). Despite the high level of simulation fidelity, they were able to perform faster in the simulation.
- The perceived workload of a teleoperation task differed significantly between simulation and live exercise; subsequently, the operators were faster in the simulation. The higher workload of the live exercise was connected to the perceived damage risk to the equipment (Kamsickas, 2003).
- The positive aspect of simulation, missing danger, can lead to false training: the trainee can forget this imminent danger in the real task (Rose, Attree, Brooks, Parslow, Penn, & Ambihaipahan, 2000).

These examples show that, despite the high fidelity of today's simulations, there seems to be an aspect which changes the behavior in comparison to reality. While these differences can be addressed in training and education, in simulation-based studies they can jeopardize the transfer of results from simulation to the real application.

The review of the literature below will show that the behavior of humans-in-the-loop, and hence their performance in simulations depends on three main factors:

- the user's cognition[1],
- the technical capabilities of the simulation, and
- the user's level of presence.

The user's cognition affects performance directly by enabling him to fulfill the given task and to perform within the virtual environment (VE). The simulation's technical ability to facilitate the user's task also directly affects performance.

Beside these direct effects, there is an also an indirect effect: the perceived sense of presence or short, 'presence', which describes the user's feeling of 'being in' the simulation (Sadowsky & Stanney, 2002). It is commonly assumed that the user needs some sense of being in the simulation to be able to perform within the VE, which in conclusion means that presence is necessary to experience VE. The necessary level of presence to perform satisfactory within the VE and the causal relationship between presence and performance are still subject of research and will be discussed later. But without doubt, the level of presence depends on the user's cognition (internal factors)

[1] cognition: action of knowing, including consciousness of things **and judgment** about them (Dictionary of Contemporary English, 1981)

and the technical design of the interface (external factors) (Sadowsky & Stanney, 2002). Figure 1 shows the relationships between performance and these three constructs.

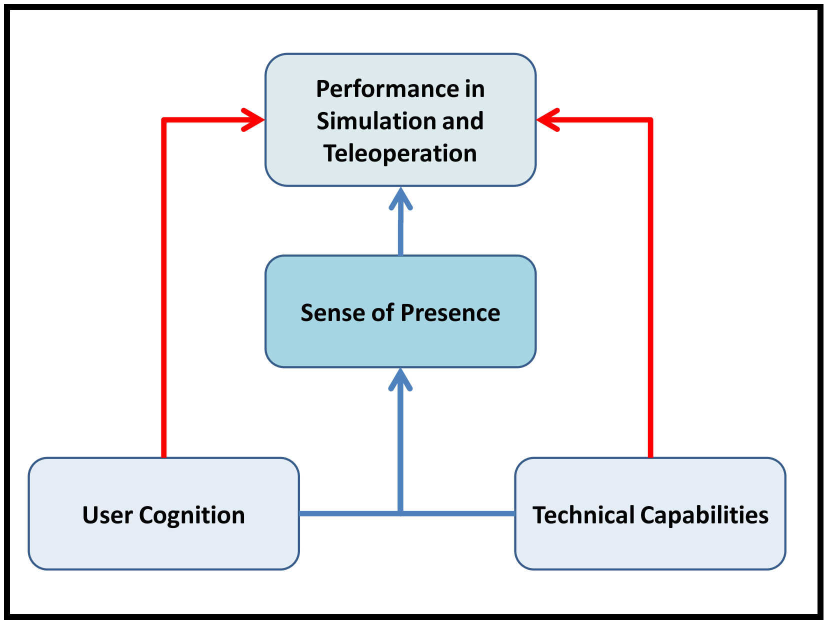


Figure 1: Performance in Teleoperation and Virtual Reality

1.1.2 Effects of User Cognition

The user's cognition has a *direct* impact on his/her performance simply by enabling the user to perform, but there are also *indirect* effects:

Jentsch and Bowers stated that the 'lack of real performance pressures' (1998, p. 247) in simulations cannot always be overcome by the user. In their experiments this fact was connected to missing motivation of some participants; the lack of motivation showed in non-task related talks and poor performance on the tasks itself.

Ford et al. (2008) found that some users applied prior experience of simulations to change their behavior in virtual environments: these users implemented a PC-based strategy of decoupling emotionally from the arousing scenario to ‘win’ the simulation. On the one hand the users reduced their presence to increase their performance; on the other hand they automatically connected simulation with ‘winning’ as their experience with gaming simulation had shown them.

In another study the subjective stress levels of participants in an experimental setting could be increased by the scenario, but did not reach high levels (Kingdon Hale, 2006). It could be concluded that within a simulation the level of stress induced by the scenario is somehow capped.

This foregoing is in line with observations by Regenbrecht et al. (1998) who noted that participants diminished their height anxiety during a virtual environment (VE) based experience by consciously decreasing their sense of presence. They reminded themselves continously that the scenario was a simulation.

The examples above show that those effects of the user’s cognition to the performance are connected to the level of presence. These *indirect* effects include prior experience (especially with games) and the assessment of the simulation as not real, which limits the sense of presence.

1.1.3 Effects of Technical Capabilities

Besides user cognition, technical capabilities of the simulation have an effect on presence and performance. The technical capabilities of the simulation *directly* affect the performance of the user: if the desired behavior cannot be performed within the technical constraints of the system, the simulation is flawed and lacks the needed level of fidelity. But additionally, the technical implementation of the simulation may have *indirect* effects:

- If a simulation requires workarounds (actions which do not correspond with reality) to overcome technical limitations, the mindset of the user changes (Woodman, 2006). The user's level of presence is limited by these workarounds; they are continuous reminders of the artificial nature of the simulation.
- Even if auditory clues in the simulation are not directly connected to the required task, they unconsciously support the level of immersion/presence (Biggs & Sriniwasan, 2002). Multimodality of the simulation's interface can increase presence, even if it is not directly related to the given task.
- Sherman and Craig (2003) stated that the technical design of the simulation's interface defines the level of immersion: the higher the level of immersion, the more intense the experience of the VE. The physical immersion is mainly defined by the design of the interface and has the aim to 'fool' (p. 382) the user's senses and to disconnect him from the real world. This sense of presence as the level of the user's disbelief is highly influenced by this level of physical immersion.

These *indirect* effects of the simulation's technical implementation to presence are significant and have to be acknowledged. How far this immersion induced level of presence really influences the performance is still subject to research.

1.2 Hypothesis

The user's cognition and the simulation's technical capabilities influence performance not only directly but also indirectly via presence. If two applications have identical direct effects on performance, a performance difference should be based on a different sense of presence. This means that if the application supports a behavior sufficiently similar to the real world task, and the user is capable of performing the task successfully, a difference in performance between simulation and reality must be related to the level of presence in the simulation. This conclusion is the basis for the hypothesis that the perceived level of presence is related to the type of application and affects performance.

My hypothesis is that the user's mental state depends on his assessment of the simulation as 'not real.' This limits his sense of presence and his performance is more careless and exempt from the fear of negative outcomes. I hypothesize that the dependency between sense of presence and performance is related to the type of implementation used (simulation vs. live exercise). The relationships of the constructs are shown in figure 2.

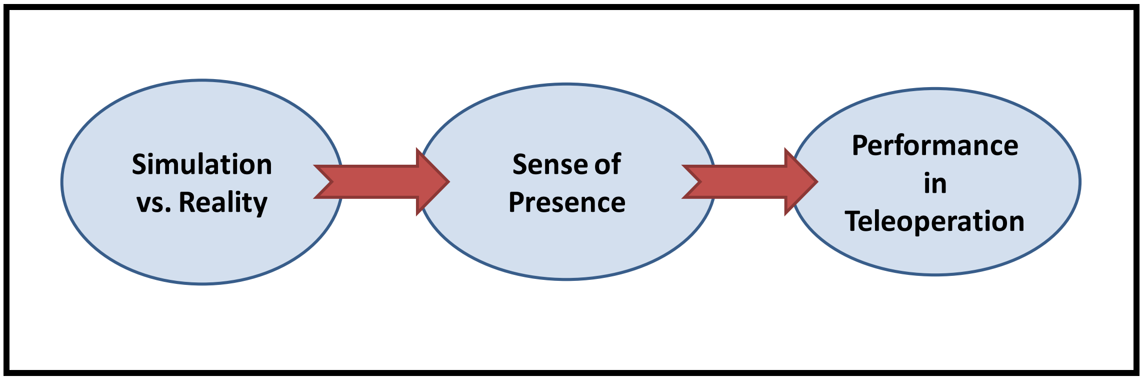


Figure 2: Proposed Relationship

1.3 The Hypothesis in the Context of Teleoperation

In teleoperation related training, education, and research, the robot is often replaced by a simulation while keeping the real control elements. The special case of teleoperation with an already reduced sense of reality due to its limitations in operational fidelity is predestined for such an approach. Teleoperation already puts the operator in a decoupled position from the robot and replacing it with a virtual simulation seems to be logical, effective, and efficient. Consequently most research concerning teleoperation is conducted by use of virtual simulations because of financial limits, availability of equipment, time limitations, and safety constraints.

Based on the hypothesis that the user behaves and thereby performs differently in simulations, these differences could jeopardize the generalization of such VE-based experiments. To be able to draw reasonable conclusions from simulations to reality, the previously described differences in behavior have to be assessed and quantified.

The main hypothesis stated that the level of presence and thereby performance is related to the type of application. This implies for a teleoperation task that with higher sense of presence the performance in measures of time to complete the task will decrease, while performance in measures of quality will increase. At first glance this contradicts the common assumption that higher presence leads to better performance. But performance is more often measured on quality than in time to completion.

1.4 Relevance of Research

The purpose of simulation is to replace reality by use of technical means to overcome limitations like safety constraints, time constraints, or availability problems. Simulations are used in a wide range of applications, from training to experimental research. The transfer of simulation results and the generalization of simulation-based experiments are customary, while fidelity of the simulation is the main concern in the evaluation of such results and their transferability. The behavioral differences are mostly - and in many cases justifiably - assumed to be based on the technical mismatch between the simulation and reality. Consistently, current simulation based experiments focus on the simulation’s fidelity to substantiate the generalization of their findings. This assessment of fidelity is often based on the quantifiable/objective technical implementation of the simulation and to a lesser extent on the qualitative/subjective perceived level of presence.

Teleoperation, as a special case, already has high limitations on fidelity. Therefore, the difference between a real and a simulated teleoperation is small and controllable. In an ideal case, the difference between a simulation and real teleoperation should only be the virtual or real function of the robot.

The previous discussion has shown that the observed behavioral differences between simulation and reality - despite general technical commonality - must point to another differentiating factor between simulation and reality: an individually different sense of presence. The observation and possible quantification of such presence-based off-set between the two applications could allow for the correction or validation of simulation based results.

1.5 Research Statement

The aim of this research was to analyze and quantify the assumed behavioral differences in a teleoperation task: on one side as a simulation and on the other as a real task in a miniature environment. The behavioral changes were supposed to surface in performance- and presence-differences, which were measured and analyzed. The research also hoped to develop a better understanding of the underlying construct of presence.

As a psychological factor, the sense of presence as a level of 'disbelief' is influenced by the user's assessment of the simulation itself. This assessment is often based on the advantage of simulations: the absence of danger or other material consequences. This missing danger has been shown to influence the user's behavior by relieving him/her from the consequences of his/her actions. Beside changes in performance during the simulation, the described behavioral difference can also endanger the generalization of simulation-based results.

It was therefore the aim of this research to establish a correlation between performance, sense of presence, and the application type to clarify the possible impact of simulations on performance. To eliminate technical influences, a comparison between a real miniature-based and a simulated teleoperation seemed to be a valid approach. The assumed correlation between sense of presence and performance is not undisputed and had to be established in the given research. Nevertheless, several studies already supported the approach that in spatial-related simulated tasks the sense of presence is correlated to performance.

While simulations are not only applied in training and education, research in teleoperation also heavily utilizes them to minimize costs and efforts. This research wanted to establish a better understanding of the transferability of such simulation-based results in teleoperation.

The underlying constructs of the hypothesis are (a) sense of presence in simulation and teleoperation and (b) the relation of presence and performance. The following review of the literature will evaluate past research of these constructs and their relationships. Additionally, possible measures for presence and performance will be examined.

CHAPTER 2: LITERATURE REVIEW

2.1 Introduction

The hypothesis stated that there is a difference in presence and, subsequently, performance between simulation and reality. This was based on the theory that the user's knowledge of the simulation as not real limits his sense of presence.

The psychological construct of presence in simulation and teleoperation is often discussed in the literature, but seldom defined or sufficiently comprehended. The following review of literature focuses (a) on the definitions of presence in teleoperation and simulations, (b) the effects of user's cognition and technical implications on presence, and (c) the relation of presence and performance. Furthermore the review will define possible measures for the above constructs.

2.2 Presence in Simulation and Teleoperation

The definitions of presence or telepresence and especially their distinction from the concept of immersion are not conclusive. This can be a reason for wrong conclusions about their influence on performance. A short survey of definitions is necessary to establish a common understanding of presence/telepresence and its possible connections to other concepts.

When Slater and Wilbur proposed their framework for immersive environments (1997), they distinguished between presence and immersion. According to their definition, the construct immersion is related to the interface technology and it is a quantifiable aspect of the simulation's technical implementation. Based on this, the quantifiable extent of immersion is defined by the level of the modality and fidelity. Presence, on the other hand, is described as a subjective and objective state of consciousness and is related to a sense of being in a place. Based on this definition, presence can be an increasing function of immersion. Consequently, Blade and Padgett (2002) state that the higher the level of immersion is, often, the greater is the sense of presence. But it is important to note that Blade and Padgett used the word 'often' and not 'always'. Indeed, immersion is only one of several factors affecting presence.

Unlike to the distinction between the objective technical aspect of immersion and the subjective sense of presence, Witmer and Singer (1998) saw immersion as the perceptional and subjective part of presence. They as well described presence as 'the subjective experience of being in one place or environment, even if the one is physically situated in another' (Witmer & Singer, 1998, p. 225), but also defined two parts of presence: involvement and immersion. While involvement describes the psychological level of attention the user puts on the VE stimuli, immersion is the user's *perception* of being 'enveloped by, included in, and interacting with' the VE (Witmer & Singer, 1998, p. 227). Witmer and Singer's definition of immersion as subjective perception makes the construct less quantifiable. This subjective definition of immersion is also reflected in Witmer and Singer's presence questionnaire.

The different distinctions between presence and immersion led to diverse conclusions about the constructs. Sadowsky and Stanney (2002) described these two approaches as schools, one in which immersion is a psychological effect and the other that sees immersion as technology related.

Despite the different general definitions of immersion and the term's distinction from presence, there are common observations of the construct presence. Stanney described presence as the subjective experience of being in one place or environment even when one is physically located in another (1998) which concurs both with the definitions of Slater et al. and Witmer and Singer.

In their article 'A Review of Presence and Performance in Virtual Environments', Nash et al. (2000) found that presence is related to the user's perception of the physical environment independent of the actual physical setting. They described presence as individual, a mental state, and thereby subjective. Bystrom et al. (1999) hypothesized that some sense of presence is necessary to perform in the VE: if the level of immersion is sufficient and the user allocates adequate attentional resources to the simulation, the factor of 'disbelief' is overcome and the sensation of presence is developed. It can be summarized that the construct of presence is a subjective experience and essential for experiencing a VE.

In teleoperation, similar to the virtual simulation, there is the distinction between the physical actual environment of the operator and the remote location of the robot. 'Telepresence is the perception that one is at a different location, created by sensory data transmitted from that location and possibly interaction with the environment at that location through telemanipulators' (Blade & Padgett, 2002), which makes telepresence comparable to presence in VE. With both telepresence and presence as psychological user centered construct with basically the same definition, Lee (2004) concluded that it is not meaningful to distinct between the two. He stated that telepresence and presence describe the same psychological construct and can be used interchangeably. This concurs with Draper et al. (1998), who also disregarded the different environments and saw telepresence and presence as the same user-centered psychological construct.

Despite the still ongoing discussion about the definition of presence and especially its distinction from the concept of immersion, there seem to be some common developments concerning presence and telepresence:

- Telepresence and presence describe the same psychological construct to have the feeling of being in a remote/virtual environment despite the actual physical surroundings.
- Presence is a *psychological* construct, therefore subjective, and difficult to capture.
- The *technical* implementation of the interface can be defined as the level of immersion, which is a necessity for presence. By this, immersion is not interchangeable with presence but is a prerequisite for presence.

This research follows these commonalities and defines presence as the overall subjective and psychological state of the user to be in the remote/displayed location with the exclusion of the physical environment's stimuli. The technical and quantifiable level of the interface's modality and fidelity is seen as level of immersion. Presence and telepresence, however, can be used interchangeably.

2.3 Performance, Presence, and Their Relation to User Cognition

The user's cognition (knowledge and mental state) is both essential for the performance of the task itself and the sense of presence. Kaber et al. (2002) described the following individual factors affecting performance and presence in VE: personality traits, user experience, and psychophysical factors.

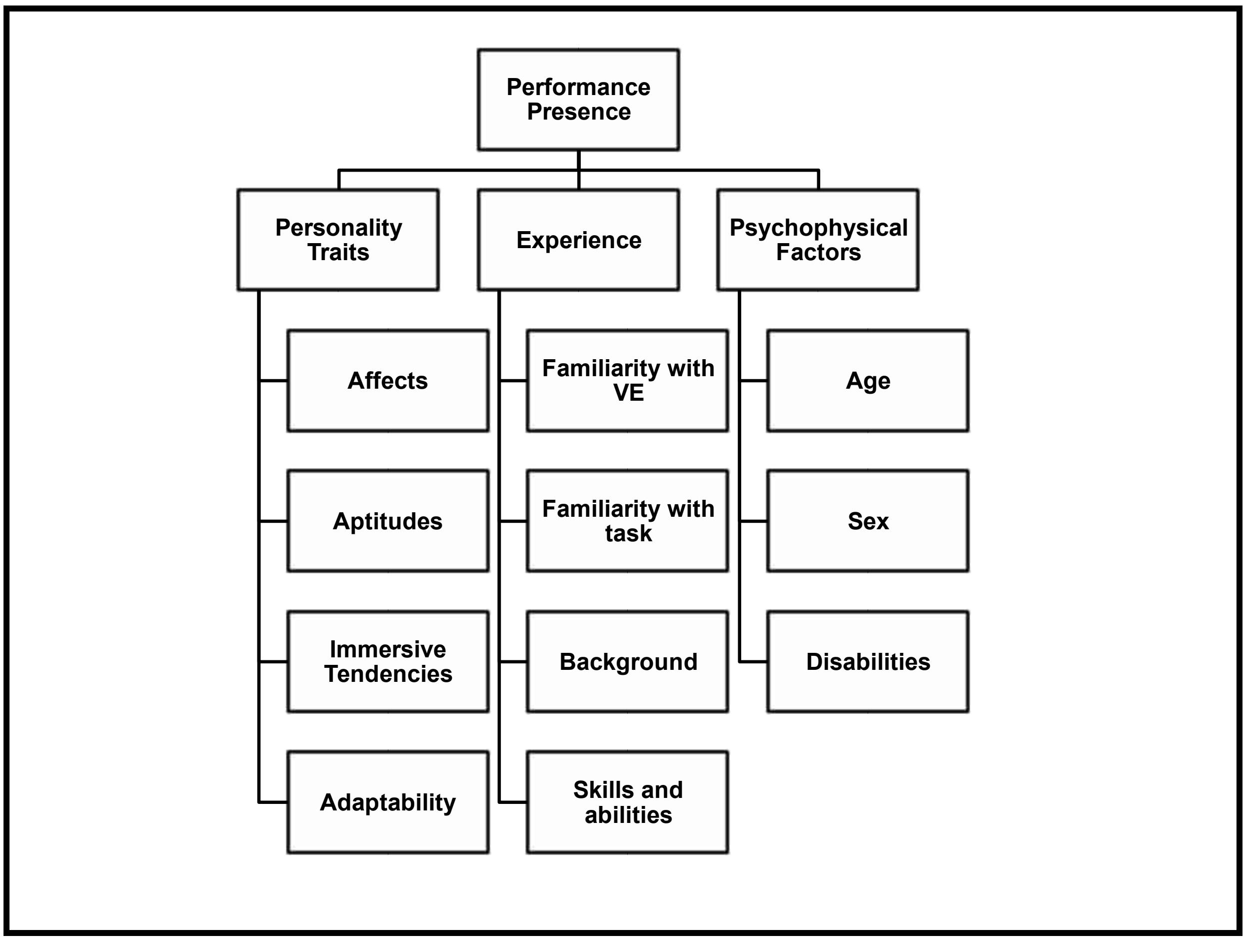


Figure 3: Factors Affecting Performance and Presence

Kaber et al. also observed that the factors are often confounded with each other and are difficult to isolate. The aim of the following discussion is the description of the individual differences affecting presence and performance, and the evaluation of their influence. The discussion will follow the factors depicted in Figure 3.

2.3.1 Individual Factors Affecting Presence and Performance

2.3.1.1 Personality Traits

Tennyson and Breuer (2002) included '*affects*' in their model of the user's cognitive system. These affects (motivations, feeling, attitudes, emotions, anxiety, and values) have a decisive impact on the user's behavior. Nash et al. (2000) stated that the user's motivation[2] has a positive correlation to presence and not only to performance. Ford Morie et al. (2008) successfully used priming (staging of the simulation as serious vs. gaming) in their study to increase the emotional connection to the given scenario and thereby increased the perceived sense of presence. This supports that affects influence both the individual performance and sense of presence. But, despite this impact, the assessment and measurement of affects was beyond the scope of this research. A randomized experimental setting had to assure that affects did not significantly influence the measurements.

Kaber et al. categorized the user's *aptitudes* into spatial, reasoning, and verbal. The proposed egocentric teleoperation tasks will only need marginal reasoning and no

[2] Refers to the willingness of a user to interact and accept the VE.

verbal skills, but extensive spatial skills. Spatial ability is defined as ‘the capacity to perform tasks requiring the mental manipulation of spatial relationships, such as mental rotation, mirror drawing, map-reading, or finding one's way around an unfamiliar environment’ (Colman, 2001). Witmer et al. (2000) found in their VE-based experiment that the individual difference in spatial abilities influenced performance significantly. These findings are supported by Rehfeld (2006), Sloan (2005), and Lathan and Tracey (2002) who all found a linear positive relationship between spatial abilities and performance in a teleoperation task. Chen et al. (2005) conducted two studies concerning unmanned ground vehicle (UGV) operators and also found the strong relationship between their spatial abilities and speed and accuracy in performance (Chen & Joyner, 2006). Additionally, Rehfeld (2006) found that spatial abilities can be trained by use of simulation, which concurs with Finkelstein (1999) who observed that gaming experienced users had higher spatial abilities scores. It can be concluded that spatial ability as an individual factor has a high impact on the performance in teleoperation tasks and VE. To distinguish between presence- and spatial ability-related performance differences, the experimental design accounted for the participant’s spatial ability.

Another personality trait is *immersive tendencies*: Witmer and Singer (1998) and Kaber et al. (2002) found a relationship between the user’s individual tendencies for immersion and the perceived sense of presence in VE. The strength of this relationship led to the established method to introduce a pre-test of immersive tendencies into presence related experiments.

In their review, Nash et al. found a strong negative relationship between *adaptability* and presence. The faster the user adapted to the new environment the more he/she got distracted by the real environment. This increased perception of the actual environment contrary to the desired perception of the VE decreases presence. But they concluded that more research was necessary to support further conclusions (Nash, Edwards, Thompson, & Barfield, 2000).

2.3.1.2 Experience

Performance highly depends on prior experience or training. With a higher level of experience in the respective tasks, teleoperation in general, and virtual environments we expect better performance. Sherman and Craig (2003) observed that memory, abilities, past experience, emotional state, and cultural background also influence the VE experience which is connected to presence. Additionally, the simulation's domain (task, environment and interface) should match the user's domain knowledge to maximize performance. One important aspect of the user's experience is prior gaming practice: Lee and Perez observed that '… a gaming environment may produce different results from straight simulation[3] because participants are asked to perform with cognitive goals (winning) added' (Lee & Perez, 2008, p. 172). Over time, this imprint of winning before

[3] Gaming simulations are distinguished from serious simulation by their intent of entertainment, as opposed to education, training, and research.

learning is applied during all simulations, independent of their intention. Prior experience of the operator and the training effects during the experiment had to be observed.

Besides these direct effects of experience and knowledge on performance, user experience also influences the perceived sense of presence. Nash et al. (2000) stated in their review of the literature that increased experience and practice (familiarity) with VE were associated with higher presence. This concurs with observations by Finkelstein (1999) and Lee et al. (2004). On the other hand, Ford Morie et al. (2008) observed that users utilized a gaming-strategy to perform better in emotional demanding scenarios: they 'decoupled' themselves emotionally from the simulation, and hence limited their presence. Gaming experience also had to be observed in the experiment to test for significant effects on the measurements of performance.

Kaber et al. (2002) found in their review of studies that user experience with VEs and with the required task influenced the perceived cognitive task load and the development of situation awareness during the simulation. Also, as mentioned, Kamsickas (2003) observed that the perceived workload of a simulation was less than that of the live task, which supports the notion that knowledge of VE influences performance. Kamsickas also observed that a lower workload resulted in a better performance. With respect to workload, Rehfeld used the MART (Malleable Attentional Resource Theory) to find and analyze an optimal level of mental workload in a teleoperation task. The result was the theory of a U-shaped relationship between performance and workload. This means that the workload has to 'fit' the participant's knowledge to optimize performance. The perceived workload had to be measured and adjusted to prevent negative effects on performance in the experiment.

2.3.1.3 Psychophysical Factors

Additional individual factors are sex, age, and disabilities, but whether sex influences spatial tasks is an ongoing discussion. While Peters (2005) found significant differences between sexes in performance on mental rotation tests, Chen et al. (2005) found no significant sex differences in performance of a teleoperation task. It was beyond the scope of this study to establish sex differences, but a possible effect on performance was avoided by stratifying sex across experimental groups. The factors age and disabilities in relation with the proposed teleoperation task are highly confounded with the user's experience level and were randomized over the experimental groups. Additionally, limited data about age and visual limitations was collected in an initial questionnaire.

2.3.1.4 Summary of Individual Factors

The cited studies and literature reviews showed that prior life experience influences performance and presence. The main factors are:

- Affects like motivation and emotional connection.
- Familiarity with VE, especially gaming experience.
- Familiarity with the given task (mental model).
- Ability to perform in the given task (esp. spatial ability).

These factors are difficult to control and were randomized and monitored during the experiment. The experimental design ensured an even distribution among the participants to statistically eliminate this influence on performance.

Other individual effects playing a major role in performance and presence are immersive tendencies, spatial aptitudes, and a matching cognitive workload to prevent either under- or overload (perceived workload). These factors have established measures and will be discussed in the following.

2.3.2 Measures for Individual Cognitive Factors

2.3.2.1 Immersive Tendencies

The level of sensed presence is highly dependent on the user's immersive tendencies. Possible measures for immersive tendencies are the Immersive Tendencies Questionnaire (ITQ), the Tellegen Absorption Scale (TAS), and the Dissociative Experience Scale (DES).

The ITQ is the pre-test for the presence questionnaire (PQ) established by Witmer and Singer as test for the participant's predisposition to be present in a VE (AD-A286 183, 1994). In the current version, it consists of 18 questions with a scale from 0 to 7, which are summed without weighting (Witmer & Singer, 1998; Appendix C). Witmer and Singer found a small correlation (r=0.24, p<0.01) between the ITQ and the IPQ in data across several experiments (Witmer & Singer, 1998, p. 237).

Tellegen and Atkinson stated that 'absorption is interpreted as the disposition for having episodes of 'total' attention that fully engages one's representational (…) resources' (Tellegen & Atkinson, 1974, p. 268). The Tellegen Absorption Questionnaire or Scale (TAQ/TAS) consists of 34 questions to assess the openness of a person to such absorbing experiences, which are similar to immersive tendencies. The questions cover 6 factors, which are (a) responsiveness to engaging stimuli, (b) synesthesia[4], (c) enhanced cognition, (d) oblivious/dissociative involvement, (e) vivid reminiscence, and (f) enhanced awareness (Kihlstrom, 2006). Each question is answered 'true' or 'false' and the positive answers are summed.

Defining dissociative experiences as a 'discontinuity in awareness' (Carlson, Waller, & Putnam, 1996, p. 300), the Dissociative Experience Scale (DES) measures the degree of dissociation from reality which exists to a higher or lesser extend in every person from normal (e.g. daydreaming) to even pathological. The 28 DES questions cover three factors of dissociative experiences: amnesia for the dissociative experience, absorption and imaginative involvement, and derealization/depersonalization. Each question has a score from 0 to 100 and the final DES score is the average of all questions.

Wiederhold et al. (2001) found a high correlation between TAS and DES and also between Witmer and Singer's PQ and DES, but did not test the ITQ. Nevertheless, the established positive relationship between the ITQ and possible presence measures supported the use of the ITQ in this study.

[4] synesthesia: a concomitant sensation; especially : a subjective sensation or image of a sense (as of color) other than the one (as of sound) being stimulated (synesthesia)

2.3.2.2 Spatial Abilities

To distinguish between the performances differences based on implementation type and cognition-based performance differences, the user's spatial ability had to be measured and observed. As possible means, the Guilford and Zimmerman tests divide spatial abilities into two factors: spatial orientation and spatial visualization (Guilford & Zimmerman, 1948). While spatial orientation is described as the ability to realize spatial relations with reference to one's own body, spatial visualization is the capability for processing and imaging movements or other changes in visual objects. The two tests were administered prior to the operational tasks.

2.3.2.3 Perceived Workload

Like spatial abilities, the user's perceived workload is also connected to performance. Hart and Staveland defined workload as 'the cost incurred by a human operator to achieve a particular level of performance' (1988, p. 240). Rehfeld (2006) confirmed the U-shaped relationship between performance and workload, which states that both too high and too low mental workload negatively affect the performance. The mental workloads of the given tasks in this study were observed to analyze possible influences of the tasks' difficulty level on performance, with the aim to ensure a correct' level of task difficulty. Potential measures for workload are the NASA Task Load Index (NASA TLX), the Subjective Workload Assessment Technique (SWAT), or the Workload Profile (WP).

The NASA TLX (Hart & Staveland, 1988) uses six subscales to represent the sources of workload: mental demand (MD), physical demand (PD), temporal demand (TD), own performance (OP), effort (EF), and frustration (FR) level. The subscales cover the task-related (PD, MD, and TD), behavior-related (EF, OP), and subject-related (FR) factors of workload. In a first questionnaire, the six sources are rated by the participant each on a scale from 0 to 100. In the second part, the sources are pair-wise weighted by the participant to derive the sources with the highest effect on the overall workload. The rates of the first part are then multiplied by the evaluated weights.

The Subjective Workload Assessment Technique (SWAT) uses three factors: time load, mental effort load, and psychological stress load (Reid & Nygren, 1988). Each load has three levels, thus making 27 possible combinations. In a first step, the participants have to sort a subjective ranking set of all combinations to develop a weight scheme. This weight scheme can later be used to assign a workload index based on the assessment of the task with regard to the load (e.g. high time load, low mental effort, and medium stress). The SWAT allows for the workload comparison of unequal tasks. Hart and Staveland (1988) commented on the SWAT that it gives no information about the source of workload, has a low sensibility and reveals thereby less information than the NASA TLX.

The Workload Profile (WP) defines eight workload dimensions based on the multiple resource model (Wickens, 2002), which are perceptual/central processing, response selection and execution, spatial processing, verbal processing, visual processing, auditory processing, manual output, and speech output (Tsang & Velazquez, 1996). The participants then assign their perceived level of attention required by the task for each dimension, ranging from 0 (0%, no attention) to 1 (100%, maximum attention). A summation of the percentages across the dimension gives an overall workload rating.

Rubio et al. (2004) compared the three multidimensional subjective workload measures. Their research found significant use and validation both of the NASA TLX and the SWAT. Rubio et al. also found all three measures equally intrusive to task performance, considered WP slightly more sensitive, noted a high convergent validity[5] between all measures, and found a slightly higher correlation to performance for the NASA-TLX. Further, SWAT and WP need significant effort during their measurement and can be time consuming as reported by Hart and Staveland (1988), Reid and Nygren (1988), and Tsang and Velazquez (1996). Consequently, the NASA TLX was assessed as the best measure of workload in the given context to establish a sufficient high workload on the participant to ensure optimal performance.

2.4 Presence and the Simulation's Capabilities

The technical capabilities of the simulation can affect both presence and performance. In particular, the technical influence of the simulation on presence is summa-

[5] Convergent validity is the degree to which an operation is similar to (converges on) other operations that it theoretically should also be similar to (Wikipedia).

rized under the construct of immersion, which is mainly determined by the interface design. This is supported by Sherman and Craig who stated that the more diversified and complex the modality of interaction is, the higher is the level of immersion, and thereby the experience of the virtual reality (Sherman & Craig, 2003). They also stated that, to enhance this experience, the interface should allow for seamless information flow between the user's real world and the virtual world.

Sherman and Craig further distinguished between the hardware and software components of the interface, both determining the level of immersion. *Input* via hardware (e.g. body tracking, voice/sound recognition, physical controllers) on one side defines the way of communication from user to the simulation. The hardware *output* with different and complimentary modes (visual, aural, haptic, and olfactory), on the other side, defines the communication from simulation to the user.

For one of those complimentary modes, May and Badcock (2002) stated that visual display of motion in VE is technically problematic since the technology is not able to provide a natural visual display, which contradicts the desired seamless information flow. Visual factors like resolution and update rates are technically contradicting, but both influence the perception of VE. May and Badcock also found that a mismatch between visual and vestibular motion cues lead to simulator sickness. In fact, the display of self motion in egocentric VEs (vection) is a key underlying element for presence, but it is also correlated to simulation sickness (Hettinger, 2002).

Although visual stimuli have a greater influence on the perception of the VE, there is also a strong spatial coupling between auditory and visual senses (Storms, 2002). Shilling and Shinn-Cunningham (2002) for example found that complementing auditory clues are essential for the environmental realism of the VE which eventually leads to a higher sense of presence.

In addition to the described visual and auditory cues, haptic interfaces support the touch, feel, and manipulation of objects in VE. Beside the sometimes necessary provision of task-relevant cues, they also increase sense of presence (Biggs & Sriniwasan, 2002). While the other interface actions are one-way, haptic interaction is inherently bidirectional between user and VE.

The hardware design of the interface input side, even if it is not haptic, also influences performance directly and indirectly via presence. Control and sensory factors contribute to presence as found by Witmer and Singer (1994; 1998) and are covered in theirs and others presence questionnaires.

Bystrom (1999) stated that immersion is a quantifiable measure of the interface technology (following Singer) and can be determined by inclusiveness (exclusion of real world stimuli), extensiveness (number of interface modalities), surrounding (panoramic field of view), and vividness (display resolution). Similarly, Nash et al. (2000) found that important factors influencing performance in VE are: (a) field of view, (b) display rendering, (c) control devices, (d) haptic feedback, and (e) head tracking. In contrast, factors affecting presence in VE are (among others): breadth (sensory modality), depth of vision, resolution, motion, self-representation, speed (e.g. update rates), range (ability to change and modify VE), natural interaction techniques (e.g. head tracking), and seamless interaction (unobtrusive).

The literature showed a significant impact of the technical interface specifications/design not only on performance, but also on presence. The *direct* effect of technical capabilities on the user's performance was evident even without using the construct of immersion: if the system lacks the fidelity to support the task, performance is limited. The *indirect* effect of the technical interface design to performance via presence was basis of the hypothesis and will be observed in the experiment.

Contrary to the desired optimal interface design in many VE applications, the design of teleoperation interfaces has to follow strict restrictions and cannot solely focus on maximum immersion or an optimal sense of presence. By design, teleoperation systems already have a limited capability to induce presence and also limit the operator's possibility to interact with the environment. Figure 4 shows as example an operator stand of an UGV which especially shows the limitations on visual cues (displays) and control interfaces.

Figure 4: Example of technical limitations to teleoperation (Kamsickas, 2003)

Based on those direct effects of the simulation's capabilities, the experimental design for the study therefore had to eliminate or equalize any technical differences between the two applications that could affect performance and presence. The observable aspects were (a) visual and auditory interface design, (b) control devices and degree of control, and (c) VE implementation: speed of robot, feedback, and content/scenario. The hardware-based differences between the two interface designs were minimized by keeping interface and controls identical in both applications.

2.5 Relationship of Presence and Performance

Although a certain level of presence seems to be necessary to perform in a VE, the research on a linear positive and causal relationship between presence and performance has not been conclusive. Research on this relationship is described in the following.

In his experiment, Snow (1996) manipulated the level of immersion within a VE, and the users assessed their sense of presence after the virtual task. The tasks were distance estimation, manipulation of an object, moving in the VE, searching, and target selection. Although he established a strong relationship between immersion and presence, only a weak relationship between presence and performance was found.

Similarly, Witmer and Singer (1998) could not establish a congruent relationship between presence and performance in their experiments. They stated that they believed the inconclusive findings were related to individual factors of the participants. But all experiments 'however' (p. 237) showed a positive relation between presence and performance. The measure used for presence was the PQ and the tasks involved perception, locomotion, and manipulation. Using a navigational and locomotion task, Finkelstein (1999) found a positive relationship between performance and presence (r_s=0.413, p=.021). Here as well the PQ was used as measure for presence.

The sometimes contradictory findings of studies concerning the relation of presence and performance led to several reviews of those experiments. As mentioned, Bystrom (1999) postulated that some sense of presence is necessary for performing in the

VE but Welch (1999) stated that technical improvements which increase performance and presence do not automatically lead to conclusions about the relationship of performance and presence itself. Draper et al. (1998) stated that some studies did not distinguish between increased performance based on increased technical capabilities of the interface or on an increased sense of presence. This was supported by Nash et al. (2000) who observed that the multi-factorial dependencies between presence and performance complicate the confirmation of a direct relation.

Beside the difficulty of distinguishing between the direct and indirect effects of the interface design on performance, Sadowsky and Stanney (2002) saw another problem of the research: typical measures for presence are generally questionnaires, which were not standardized and developed in large numbers. Another complicating aspect is the subjectivity of post-test questionnaires: the rating of presence depends on the user's prior experiences (Freeman, Avons, Pearson, & Ijsselstein, 1999), the user's memory of the event (Wiederhold, et al., 2001), and the perceived level of their own performance. The latter would mean that increased performance could lead to increased sense of presence measured by post-test questionnaires[6] (Slater, 1999). Sadowsky and Stanney stated, however, that the relation of presence to performance has face validity and that some studies indeed provided evidence for a positive relationship, although these were

[6] *Main aspect is the correlation between 'control' in the questionnaire and task performance. Better performance might lead to higher feeling of control of the VE and subsequently to higher presence.*

strongly task related. They concluded that further study was necessary. During the analysis of the relationships of task difficulty (workload), situating awareness, and telepresence, Riley (2001) found a positive relationship for presence and performance in a simulated mine-clearing task (r=-0.327, p=.001).

Although many experiments showed a positive relationship between performance and presence and there is some face validity to it, a general direct and causal relation could not be established. The use of different definitions of presence, especially their distinction from immersion, could have led to different assessments of increased performance based on higher presence. Additionally, the multiple relations between presence, user cognition, interface implementation, and performance were often confounded in the experimental settings, based on the fact that the necessary definition and experimental measurements of unconfounded variables to pinpoint the relationship between presence and performance are difficult. Following Slater's definition of immersion, many studies reveal a relationship between the technical defined immersion level and performance, which is different from the here hypothesized performance to presence relationship. The perceived sequence that higher (technical) immersion leads to higher presence and consequently better performance is too interrelated to allow a comprehensive conclusion. For example, does the wider field of view (higher immersion) only influence the presence related impact on performance or is it the direct effect of this technical improvement (improved feedback/interface)?

The relationship of presence and performance, as one of the main constructs of this study, had to be established during the experiment to support the hypothesis. Consequently, to establish a positive relationship between presence and performance with not confounded immersion, the level of immersion had to be equalized in the experimental settings. Additionally, the possible perceived performance dependency of post-test questionnaires for presence had also to be addressed in the experimental design. To observe and reject this performance dependency, the introduction of different levels of difficulty which will affect the performance, but not presence, can support the conclusion that performance differences are solely the result of changes in presence, and not vice versa. An additional application of 'online' tests to minimize the post-test disadvantages had to be observed.

With this approach, the experiment un-confounded presence, immersion, and performance. Additionally, the test of the proposed hypothesis further enhances the common understanding of the relationship between presence and performance.

2.6 Measures for Presence and Performance

The hypothesis stated that the type of implementation and the related sense of presence are influencing performance. This defines 'type of implementation' as the independent variable, with 'presence' as mediating variable, and 'performance' as dependent. Possible measures for the mediating/dependent variables will be discussed in the following.

2.6.1 Performance

The experimental use of an unmanned ground vehicle's (UGV) teleoperation allowed for the definition of quantitative measures of performance. These measures had to be independent from the implementation's technical abilities to establish the proposed relationship between implementation, presence, and performance.

Park (1998) used speed (time to complete task) and accuracy (counting errors) as performance measures of a simulated teleoperation task. He stated that speed alone was not sufficient as performance measure since it does not measure quality of performance. Nevertheless, time to finish task was also used in related studies by Williams (2001), Riley (2001), Chen et al. (2005), and Sloan (2005).

The aforementioned studies showed the common use of completion time for teleoperation tasks as measure for speed. In a similar approach, this study utilized the number of runs [runs] in a given timeframe as one measure for speed.

Additional to changes in speed, the quality of performance was expected to decrease and had to be included as measure. Aim was the distinction between very careful operators (slow, hence low performance) and more risky ones (fast, but more mistakes and hence also low performance). The number of operational errors like hitting obstacles was used a measure for quality [errors]. The later experimental design will show that number of runs as one measure for speed did not cover the mentioned distinction between risky and careful operators and an additional measure, average speed of the robot [runs/min.], was introduced (Figure 5).

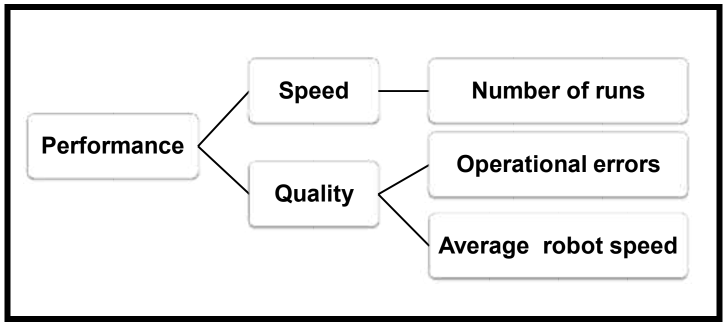


Figure 5: Performance Measures

2.6.2 Presence

As previously shown, presence is a construct which is not exactly defined and which has no direct quantitative/objective measure. Sherman and Craig (2003) described questionnaires and user observation as qualitative, while measurements of physical reactions are quantitative. Sadowsky and Stanney developed a more elaborate list of measures (Table 1), and an important difference to Sherman and Craig is the classification of user observations (reflexive motor acts) as objective measures.

Table 1: Presence Measures (Sadowsky & Stanney, 2002)

Presence Measure	*Type of Measure*	*Description of Measure*	*Strengths*	*Weaknesses*
Rating scales	Subjective	Rate level of presence experienced in a VE.	Direct perception of user.	Interrater reliability may be weak; limited utility for comparison of experiences across diverse environments.
Subjective report	Subjective	Directed, open-ended questions of reactions and impressions related to presence.	Direct perception of user.	Interpreting results may be difficult due to response variability.
Comparison-based prediction	Subjective	Distinguish between a real-world scene and a VE simulating that scene or between alternative VE systems (standardized comparison cases may be used).	No bias due to definition of terms.	Size of just noticeable difference dependent on stimulus characteristic compared.
Cross-modality matching	Subjective	Represent experience of one sensory modality through another modality.	Can be used when verbal scaling is inappropriate or difficult to quantify.	Interpreting results may be difficult due to response variability.
Behavioral measures	Objective	Reflexive physical reactions or behaviors evoked by stimuli or events occurring in a VE.	Directly influenced by the VE, no subjective bias.	May not reflect influences of stimuli or events on total presence.
Physiological indicators	Objective	Physiological responses that occur when experiencing VE stimuli or events occurring within it.	Can serve to isolate influences on presence, objectively measured.	Internal or external "noise" may affect measures, intrusiveness, and reliability problems.

The distinction between qualitative and quantitative measures in this field of research located between engineering and psychology is not as simple as assumed. Especially, there is no comprehensive and general definition for subjective and objective measures in the field of psychology (Muckler & Seven, 1992). To establish a necessary definition for this research, subjective measures were defined as self-assessments or reports, biased by the participant, while objective measures rely on a ‘neutral’ measurement by an instrument, a test, or an observer. In this view, objective measures can also have subjectivity. The following parts describe first the subjective and second the objective measures.

2.6.2.1 Subjective Measures

Main subjective measures for presence are questionnaires which are applied at different phases of the experiment. With the aim to keep the task itself uninterrupted, the questionnaires are administered mostly before or after the task. In the following, different approaches for questionnaires will be discussed.

In an effort to develop a better understanding of the underlying concepts of presence and their relationships to performance, Witmer and Singer (1994) developed two questionnaires: the Immersive Tendencies Questionnaire (ITQ) as pre-test and the Presence Questionnaire (PQ) as post-test. Asking questions about four presence affecting factors – control, sensory, distraction, and realism (Table 2) – they developed the PQ as self report.

Table 2: Factors of Presence (Witmer & Singer, 1998)

Control Factors	Sensory Factors	Distraction Factors	Realism Factors
Degree of control	Sensory modality	Isolation	Scene realism
Immediacy of control	Environmental richness	Selective attention	Information consistent with objective world
Anticipation of events	Multimodal presentation	Interface awareness	Meaningfulness of experience
Mode of control	Consistency of multimodal information		Separation anxiety/ disorientation
Physical environment modifiability	Degree of movement perception Active search		

They found promising results of correlation between questionnaires, presence and performance (Witmer & Singer, AD-A286 183, 1994). The ITQ measures the individual and not necessarily simulation- or teleoperation-related tendency to become involved in activities, to maintain focus on current activities, and to play video games (18 items). The PQ indirectly measures the degree of experienced presence in VE by inquiring about the four related factors (32 questions). Both, ITQ and PQ use 7-point scales, and the final score for presence is the sum of the answers. An analysis of correlation and validity showed that the PQ measures a single construct of presence, where as the ITQ measures immersive tendencies. Witmer and Singer stated that presence should be related to simulator sickness, task performance, modes of interaction, ITQ, and spatial ability (Witmer & Singer, 1998). The ITQ and PQ were used by Finkelstein (1999) who found a positive relationship between the PQ-defined presence and performance.

Slater, based on his more stringent separation between immersion and presence, criticized the PQ for confounding individual differences (presence) and the VE characteristics (immersion) (Slater, Measuring Presence: A Response to the Witmer and Singer Presence Questionaire, 1999). He developed an alternative questionnaire which was later defined as 'SUS', Slater-Usoh-Steed (Slater, Steed, McCarthy, & Maringelli, 1998). The SUS does not use the indirect approach of the PQ, instead it questions the user directly about his/her experience of 'being in' the VE, which necessitates some understanding of the participant about the construct of presence.

In an attempt to resolve the dispute about the two approaches, Usoh et al. used the PQ and SUS to compare the results between a real and a simulated task. But neither questionnaire established a presence difference between the real and simulated task (Usoh, Catena, Arman, & Slater, 2000). The authors, however, emphasized the low power of their experimental design. Another finding was the high influence of the user's own experience and hence understanding of 'being in' a simulation. According to this effect, a novice in simulations might feel immersed much earlier than an experienced VE-user. This concurs with prior described effects of user experience.

The relationship of presence with multiple factors and the resulting complexity of the construct did not only lead to the described dispute about validity of questionnaires but also to a significant number of them (for one possible overview see van Baren and Jsselsteijn, 2005).

One approach to implement the Witmer and Singer PQ and to respect the remarks of Slater et al. is the 'igroup presence questionnaire' (IPQ). Schubert et al. stated that presence is a subjective experience and thereby only quantifiable by the experiencing user, which supports the use of self reporting questionnaires (Schubert, Friedmann, & Regenbrecht, 2001; Schubert T. W., 2003). Based on Schubert's et al. model of presence, immersion is the technological aspect while conception is the cognitive perception of the VE. Both are necessary for the experience of presence in a VE (Figure 6).

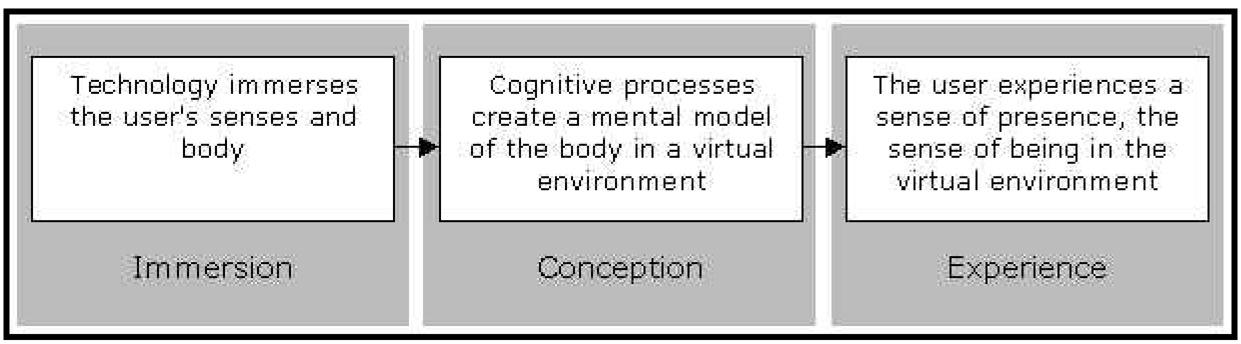


Figure 6: Development of Presence (igroup, 2004)

Schubert et al. also stated that presence emerges out of the user's mental model and his attention allocation as result of his cognitive processes. Presence is thereby related to the sense of acting in and with the VE and the concentration on the VE while ignoring the reality. This is in line with the attention allocation theory which is discussed later.

Schubert's et al. efforts aimed at the exploration of possible factors loading on presence, contrary to Witmer and Singer who concentrated on presence itself. They conducted two sequential studies to develop an according questionnaire. In a first study, a combined questionnaire of 75 items (including PQ and SUS) was presented, and the data were analyzed. A principal component analysis with oblique rotation showed eight major components (Table 3).

Table 3: Factor Analysis Study 1 (Schubert, Friedmann, & Regenbrecht, 2001)

Table 1. *Factor Analysis Study 1. Factors, Numbers of Items, and Explained Variance*

Component	Name	Label	Number of Items	Eigenvalue	% of Variance Explained
1	Spatial presence	SP	14	14.087	20.717
2	Quality of immersion	QI	8	4.574	6.726
3	Involvement	INV	10	3.824	5.624
4	Drama	DRA	7	3.083	4.533
5	Interface awareness	IA	7	2.485	3.655
6	Exploration of VE	EXPL	6	2.262	3.326
7	Predictability & interaction	PRED	6	1.967	2.893
8	Realness	REAL	5	1.901	2.795

The analysis showed three components related to presence: SP[7], INV[8], and REAL[9], with a total of 29 items loading on them. The constructs of immersion and interaction are covered by the other factors. By using confirmatory factor analysis those

[7] Spatial Presence
[8] Involvement
[9] Realness

items were selected, which only loaded on one component. This resulted in five items for SP, four for INV, and three for REAL.

A 2nd study was run with the improved questionnaire in which only presence (SP, INV, REAL) and interaction (EXPL[10], PRED[11]) related questions were used. A factor analysis of the collected data supported the findings of study 1 and resulted in five loading components (Table 4).

Table 4: Factor Analysis Study 2 (Schubert, Friedmann, & Regenbrecht, 2001)

Component	Name	Label	Number of Items	Initial Eigenvalue	% of Variance Explained
1	Spatial presence	SP	14	11.767	31.804
2	Exploration of VE	EXPL	8	2.682	7.248
3	Realness	REAL	3	2.070	5.594
4	Predictability & interaction	PRED	2	1.717	4.640
5	Involvement	INV	10	1.559	4.214

The resulting igroup presence questionnaire (IPQ) was further optimized and is able to distinguish between three factors for presence: sense of spatial presence, level of involvement (attention allocation), and judgment of VE's reality (Figure 7).

The post-test questionnaire consists of 14 questions (Appendix A) in which one covers sense of presence in general, five questions cover spatial presence, four in-

[10] Exploration of VE
[11] Predictability and interaction

volvement, and four experienced realism. The answers have a scale from 0 to 6, and the sense of presence can be calculated as the sum of all questions.

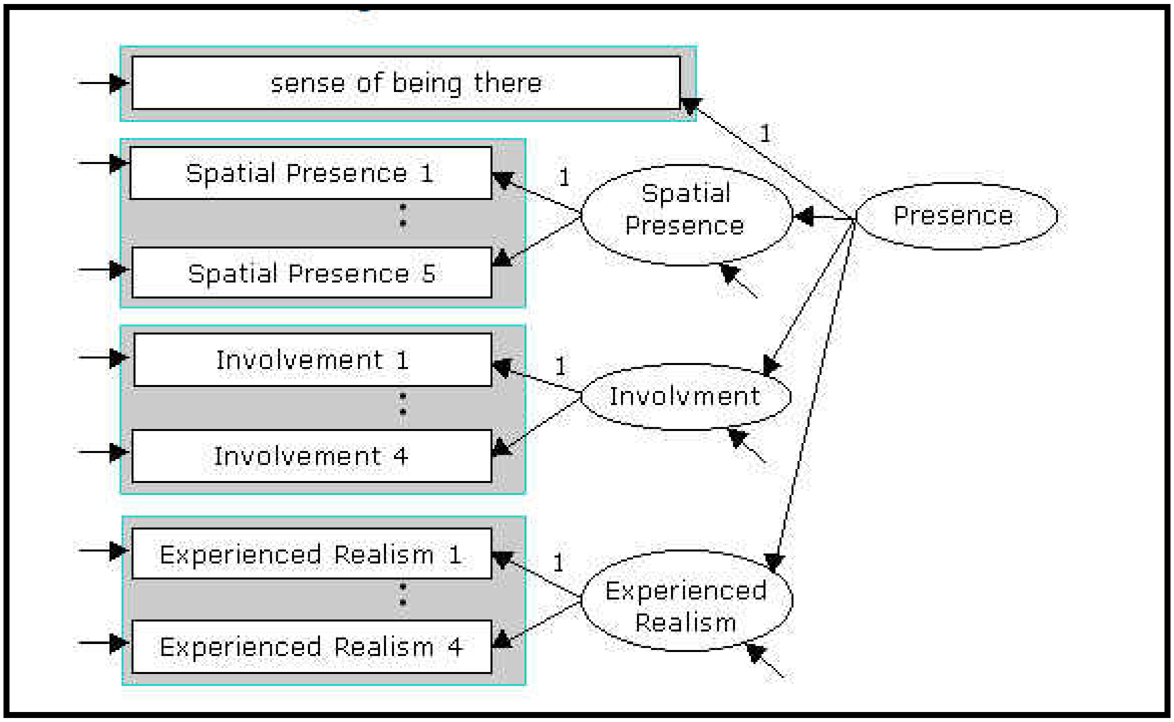


Figure 7: IPQ Factor Analysis (igroup, 2004)

In further studies related to 3D-Games, Regenbrecht and Schuster (2002) used the IPQ and the three component approach to further analyze their relationships. They found a high correlation between perceived level of interaction and spatial presence but lower relation to involvement and sensed reality.

The IPQ combines the questionnaires of Witmer and Singer and Slater et al. into a new questionnaire which showed in a factor analysis three major components which were also observed by other studies. This mediating approach and the possibility to distinguish between the three components of presence advocated the use of the IPQ in this experiment.

Despite the advantages of post-test questionnaires for presence, they have the disadvantage of being subjective and are delayed to the actual experience (Lee & Perez, 2008; Riley, 2001). This introduced the desire to establish direct, objective, and quantifiable measures of presence related factors.

2.6.2.2 Objective Measures

The desired objective measures have to overcome the user's subjective influences of questionnaires. One solution seems to be the relationship of physical measures to the construct of presence.

Wiederhold et al. (2001) used the Tellegen Absorption Scale (TAS), Dissociative Experience Scale (DES), Simulator Sickness Questionnaire (SSQ), and the Presence Questionnaire (PQ) to analyze their relationship with quantitative measures as heart rate and skin conductance. The TAS questionnaire (Tellegen & Atkinson, 1974) replaced the ITQ and as additional measure for immersive tendencies Wiederhold et al. used the DES questionnaire (Carlson, Waller, & Putnam, 1996). The SSQ (Kennedy, Lane, Berbaum, & Lilienthal, 1993) assesses the possible experience of simulator sickness[12]. Wiederhold et al. found that heart rate and skin resistance had high correlations with presence (PQ), degree of realism, and immersiveness. For the experiment, they used an emotionally challenging VE (high fidelity flight simulation). Also, they found a (expected) correlation between TAS and DES, and between the DES and PQ.

The positive correlation of heart rate to presence was also supported by Meehan (2001). Additionally, Jang et al. (2002) found that the variance of the heart rate is also a possible measure for reactions to the VE. They used an active flight and a passive driving simulation for the experiments. Zimmons (2004) used a highly emotional and stressful VE to test the relationship between physiological reactions (heart beat, skin temperature, and galvanic skin response) and aspects of the visual interface design (resolution and lightning). He used the SUS as questionnaire for presence and did not find a correlation between the SUS and presence, contrary to the objective measures which showed significant differences between the two test environments. Slater's research to establish physiological measures for presence showed in one study that even in not stressful scenarios participants generated increased anxiety based on psychological preconditions, in this case social anxiety (Slater, et al., 2006). The results concerning perceived presence indicate that the emotional content of the VE has a much higher influence on the physical measures than the perceived sense of presence. The partici-

---

[12] Witmer and Singer (1998) found that simulator sickness is negative related to presence and it also affects performance.

pant's psychological pre-condition in relation to the emotional content of the simulation seems to affect the perceived level of stress and hence the physiological measures.

The literature showed a correlation especially between heart rate and presence in virtual environments, but the studies often used emotional arousing scenarios to foster the physical responses. With the assumed relation of presence and an emotionally unchallenging application, these objective measures were assumed to be less significant. Nevertheless, an inclusion of the objective measure heart rate into the experiment was expected to bolster and complement the subjective IPQ.

In the desire to establish quantifiable and objective measures for presence - beside physical measures - the associations of presence related constructs like attention allocation and situation awareness were utilized.

Based on the attentional resource theory, Draper et al. (1996) assumed that the more attention is allocated to the VE, the higher the sense of presence should be. In their study, limited by a small sample size, they nevertheless found that the theory of a relationship between presence and allocated resources has value. Riley (2001) also used the attention allocation theory and found that presence is influenced by the attentional allocation to the simulation. She also found an influence of task difficulty (workload) and situation awareness on presence.

But, the high correlations of attentional resource allocation with task challenge (workload), emotional connection, situation awareness, and immersive tendencies make it difficult to establish attentional resource allocation as a measure for presence. Measures of other constructs connected to the attentional resource theory, like situational awareness, were developed.

Endsley defined situation awareness (SA) in three levels as (1) the perception of the elements in the environment, (2) the comprehension of their meaning, and (3) the projection of their status in the near future. Based on this definition Endsley expected a strong relationship between performance and SA. She therefore developed the Situation Awareness Global Assessment Technique (SAGAT) to optimize performance by optimizing SA (Endsley, Toward a Theory of Situation Awareness in Dynamic Systems, 1995; Endsley, Measurement of Situation Awareness in Dynamic Systems, 1995).

In the SAGAT approach, the experiment is randomly stopped ('frozen') and randomized questions covering all three levels of SA are answered. The answers are then compared to the correct data of the VE and summarized in a percentage of correct answers. The SAGAT was primary developed to assess and improve designs technologies for highly demanding tasks (e.g. displays in fighter aircrafts). Draper, Kaber, and Usher (1998) concluded that the user of a virtual environment experiences two distinct types of situation awareness: in the actual surrounding environment and in the VE. The level of situation awareness in the VE, which on the opposite side points to the disregard of the real environment, allows conclusions about the level of presence (Draper, Kaber, & Usher, Telepresence, 1998). Although the use of SA as an indirect but objective measure for presence seemed viable, the applied scenarios needed a certain depth and contend for the participant to develop testable SA-levels. The experimental scenario would have to support all three levels of SA and should have a certain length to enable a valid questionnaire. But, the proposed experimental design with the intention to minimize technical impacts on performance and a limited reality will not support such complexity of the scenario. The current experimental design asks for minimal effects of the scenario on presence, which defies the necessary scenario complexity for the SAGAT.

2.7 Differentiation of Planned Research from Existing Literature

The review of the current literature showed the struggle to capture the concept of presence and to define its relation to performance and other constructs. The proposed and assumed relationship between performance and presence drives the discussion, since it is the aim to develop simulations and teleoperation systems that enhance performance. Often, the proposed approach to enhance presence to foster performance is expensive and not proven in all aspects. Further, the relationship between presence and performance is difficult to establish since the related factors are not easy to isolate in the experimental setting. Comparisons within simulations are hampered by the use of the same technology, which - based on this study's hypothesis - affects presence negatively. Additionally they are often confounded with the concept of immersion. The comparison of simulation to reality (e.g. live exercises or use cases), on the other side, is not easy to evaluate since many additional factors are influencing the outcome of the real task. The current review established correlations between the constructs of presence, spatial awareness, cognitive workload, immersion, attention allocation, and performance, but often the results are confounded within the constructs itself or highly task related. Subsequently, the results differ significant between studies.

The direct comparison between a confined simulation and an also confined task in a miniature world minimizes the influencing factors and leads to a direct comparison of the perceived sense of presence and the performance in relation to the implementation type. Despite the numerous observations in conducted experiments, this approach was not yet taken. The experiment will lead to a better understanding of the construct of presence and its relation to performance. The possibility to use different measures for presence in very similar setting will also contribute to a better understanding of the different measures and their relation to presence.

2.8 Summary

Originating from the hypothesis that the user's sense of presence and subsequent his performance is different based on the application type, the review of the current literature led to several conclusions about the hypothesis and its experimental implementation.

Firstly, the concept of presence is not conclusive and generally established. Especially the differentiation between presence and immersion is an ongoing discussion. The hypothesis and the experimental design acknowledged this by clear distinction between presence and immersion. This study thus followed the approach to see immersion as the technical and quantifiable contributor to the psychological construct of presence. The uncertainty connected with the construct of presence and its relation to performance had led to different and numerous measures. This study utilized a subjective (IPQ) and an objective (heart rate) measure for presence.

Secondly, the user's cognition to perform the given task is essential for the evaluation of performance. One aspect is spatial ability, which has a high influence on teleoperation tasks (esp. UGV operation) and was addressed in the experimental design. Other cognitive aspects were the perceived workload, which also influences the performance, and immersive tendencies, which affect presence.

Thirdly, the extreme high influence of the application's technical characteristics to performance had to be eliminated as much as possible. The experiment, by using a simulation and a miniature teleoperation task, followed this requirement.

The performance measures had to focus on the task and were not related to the technical application. Measures were number of runs [runs], errors in robot manipulations [errors], and average robot speed [speed].

A summary of the related construct, variables, and measures is shown in Table 5.

Table 5: Factors and Experimental Design

Category	Factor	Remarks	Experimental Design	Reference
Technical – Interface Hardware	Visual clues e.g. depth, resolution, update rate	Ability of the display to stimulated human vision	Constant	(May & Badcock, 2002), (Nash, Edwards, Thompson, & Barfield, 2000)
	Auditory cues	Ability to support the user's auditory perception o f the VE	Constant	(Shilling & Shinn-Cunningham, 2002)
	Haptic cues	Manipulation of objects with active interfaces that allow 'feeling'	Not used	(Biggs & Sriniwasan, Haptic Interfaces, 2002)
	Body acceleration	Perception of acceleration in the VE	Constant	(Lawson, Sides, & Hickinbotham, 2002), (Nash, Edwards, Thompson, & Barfield, 2000)
	Motion tracking	Ability to track user's movement	Not used	(Foxlin, 2002)
	Gesture recognition	Ability to interact with VE	Not used	(Turk, 2002)
	Input device design	Task dependent degrees of freedom, two handed operation	Constant	(Bowman, 2002), (Nash, Edwards, Thompson, & Barfield, 2000), (Sadowsky & Stanney, 2002)
	Spatial resolution	Finesse of spatial detail.	Constant	(Nelson & Bolia, 2002)
	Temporal resolution	Temporal mismatch between multiple inputs	Constant	(Nelson & Bolia, 2002)
	Breadth	Level of modality	Constant	(Nash, Edwards, Thompson, & Barfield, 2000)
	Ease of interaction, mapping	See interface design	Constant	(Sadowsky & Stanney, 2002)
	User initiated control, range of interactions		Constant	(Sadowsky & Stanney, 2002), (Witmer & Singer, AD-A286 183, 1994), (Nash, Edwards, Thompson, & Barfield, 2000)
Technical - Interface Software	Pictorial realism		Constant	(Sadowsky & Stanney, 2002)
	Length of exposure		Constant	(Sadowsky & Stanney, 2002)
	Social factors	Interaction with other avatars	Not used	(Sadowsky & Stanney, 2002)
	System factors	Realism, interface, interaction	Constant	(Sadowsky & Stanney, 2002)
	Simulator Sickness		Measured	(Kennedy, Lane, Berbaum, & Lilienthal, 1993), (Sadowsky & Stanney, 2002)
	Immersion	Objective extend of technical modality and fidelity	Constant	(Slater & Wilbur, A framework for immersive environments (FIVE), 1997)
	Consistency	Predictability of reaction in VE	Constant	(Nash, Edwards, Thompson, & Barfield, 2000)

Category	Factor	Remarks	Experimental Design	Reference
Technical - Interface Software	Self representation	Avatar or representation of user	Constant	(Nash, Edwards, Thompson, & Barfield, 2000)
Cognition Personality Traits	Affects	Motivation, feelings, attitudes, emotions, anxiety, values	Randomized	(Sherman & Craig, 2003), (Ford Morie, Tortell, & Williams, 2008), (Nash, Edwards, Thompson, & Barfield, 2000), (Tennyson & Breuer, 2002)
	Attention and concentration, involvement, situation awareness	Psychological level of attention committed to the virtual stimuli.	Measured variable	(Kaber, Draper, & Usher, 2002), (Foxlin, 2002)(Munro, Breaux, Patrey, & Sheldon, 2002), (Witmer & Singer, 1998)
	Spatial aptitude	Position and orientation in VE	Randomized, measured	(Kaber, Draper, & Usher, 2002), (Witmer, Sadowsky, & Finkelstein, Technical Report 1103, 2000)
	Reasoning aptitude	Strategy development	Randomized	(Kaber, Draper, & Usher, 2002)
	Immersive tendencies		Variable, measured	(Kaber, Draper, & Usher, 2002), (Witmer, Sadowsky, & Finkelstein, Technical Report 1103, 2000)
	Adaptability	Adjustment to new circumstances	Randomized	(Nash, Edwards, Thompson, & Barfield, 2000)
Cognitive Experience	User experience Memory	User's -Knowledge -Skills -Mental models of task and simulation -Gaming experience	Randomized, partly monitored	(Kaber, Draper, & Usher, 2002), (Sherman & Craig, 2003)
	Cultural background		Randomized	(Sherman & Craig, 2003)
	Workload	Perceived or real	Manipulated, measured	(Lee & Perez, 2008)
Cognitive Psycho-physiological	Age / Disabilities	Confounded with other individual constructs (e.g. experience, aptitudes)	Randomized	(Kaber, Draper, & Usher, 2002)
	Sex	Contrary study results, often related to other constructs (e.g. spatial abilities)	Stratified	(Chen, Durlach, Sloan, & Bowsen, 2005), (Peters, 2005)

CHAPTER 3: RESEARCH DESIGN

3.1 Research Questions

The main hypothesis stated that the user's mental state depends on his/her assessment of the simulation as 'not real' and that this limits his/her sense of presence. Limited presence, in turn, leads to a more careless behavior that is exempt from the fear of negative outcomes, which in the end affects performance.

Each single relationship between (a) type of application/danger to equipment, (b) sense of presence, and (c) performance had to be established in this research. Also, it had to be established that (d) presence affects performance and not the reverse. Subsequently, it was the aim of this research to answer the following questions:

1. In a teleoperation task, will the perceived danger to equipment presented in a miniature world lead to a higher sense of presence, compared to a non dangerous simulation? ($A \rightarrow B$)
2. Will a higher sense of presence lead to a change in performance (higher in quality, lower in speed)? ($A \rightarrow C$)
3. Will a changed workload of the teleoperation influence performance but not presence? ($B \rightarrow C$)

To establish these relationships, the experimental setting had to enable the measurement of presence, performance, and workload based on the two possible application types with two different workload levels. To establish the argument, that the perceived danger to equipment affects presence, the operator's assessment of such danger also had to be monitored. Additional affecting constructs like immersive tendencies and the user's (task related) cognitive abilities were observed.

The requirements of technical comparability were fulfilled by using a miniature-world-based 'real' application and a virtual simulation, each with two levels of difficulty. The levels of difficulty had to be related to the task itself and were not connected to the interface, since differences of the interface design would have affected presence.

The miniature world allowed a laboratory-like environment in which the performance-influencing effects of a natural environment (e.g. changing weather, ground conditions) were eliminated. Additionally, anecdotal observations of the behavior in the miniature world suggested that the prospect of 'damaging' the miniature robot already affected the operator. In the experimental setting, the simulation replaced the robot (UGV) while the real operational controls were kept.

3.2 Research Design

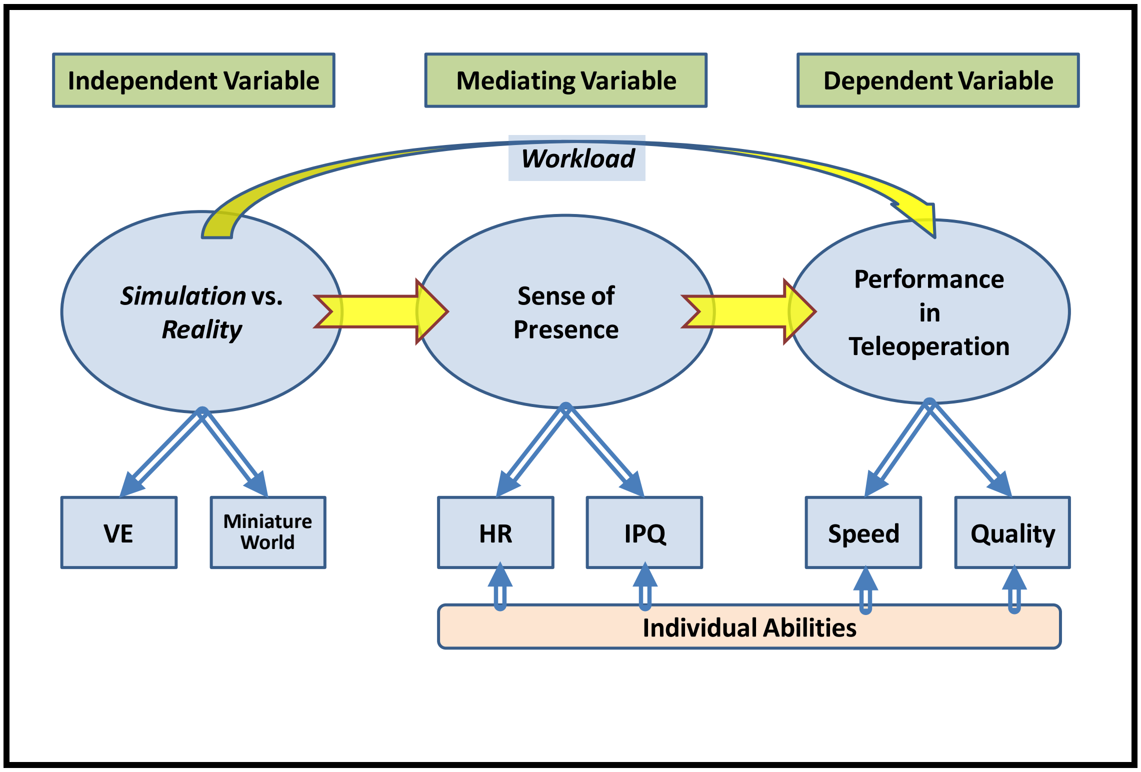


Figure 8: Research Model

The graphical research model (Figure 8) shows the assumed relationships between the constructs, factors, and variables. The independent variables were application type (virtual environment or miniature world) and level of difficulty (workload). The difference between the two types of application was limited to the way the robot existed: as a virtual model (VE) or as a real mini robot (miniature world).

Following the research question it was expected that there is - depending on the application - a difference in presence and hence in performance. Although the relationship between presence and performance is not commonly established in similar research, the literature has shown evidence for the argument that increased presence fosters better performance for tasks with VE-based movement (Sadowsky & Stanney, 2002). This also applied for the given teleoperation task. The findings of Sadowsky and Stanney were supported by Riley (2001), who also found a positive correlation between presence and performance in a teleoperation task. Nevertheless, this study enhances the knowledge about this relationship. The introduction of performance- but not presence-affecting workloads was intended to document that the performance differences were triggered by the changed sense of presence and not vice versa.

Given the expected relation of presence and performance, the theoretical model defined performance as a dependent variable and presence as a mediating variable (Garson, 2002). The sense of presence was measured by the IPQ and average heart rate (HR), while the performance measures were (a) number of surrounded cylinders (‘runs’), (b) operational errors, and (c) average speed.

The external influence of the user’s cognitive abilities and state were measured and assessed. To eliminate the possible technical effects on performance, the experimental design had to ensure a similar capabilities and behavior of the robot in both applications.

3.3 Sub-Hypotheses

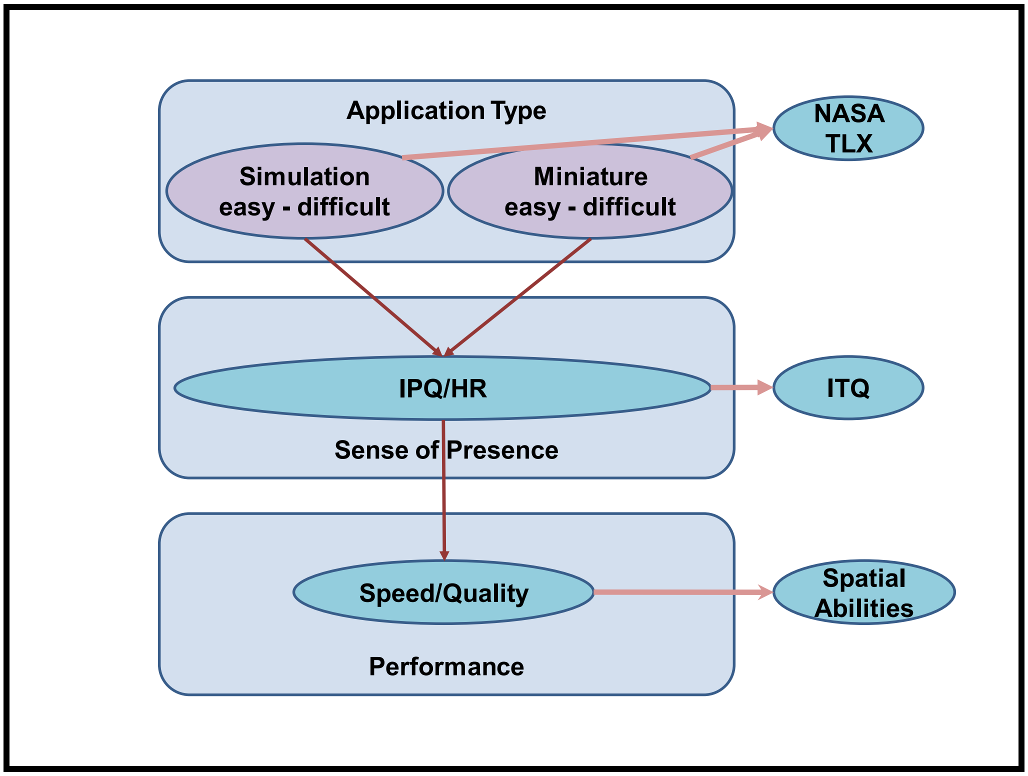


Figure 9: Path Diagram of Model and Measures

Derived from the research questions and the research design, the following relationships (Figure 9) were assumed:

1. Application type and the perceived sense of presence are related. The sense of presence in the simulation would be lower than in the real teleoperation task. The measurements of presence should be correlated between simulation and miniature world. With the given measures for presence this translated into:

H_{11a}: IPQ/HR[13] (miniature world) > IPQ/HR (simulation).

H_{11b}: IPQ/HR positively correlated between simulation and miniature world.

2. The application type is correlated to the performance. The number of runs, average speed, and number of errors would be higher in the simulation.

   H_{12a1}: Runs (simulation) > runs (miniature).

   H_{12a2}: Speed (simulation) > speed (miniature).

   H_{12a3}: Errors (simulation) > errors (miniature).

   H_{12b}: Runs/speed/errors are positively correlated with miniature world/simulation.

3. The perceived sense of presence would be negatively related to the level of performance. Higher presence leads to lower performance measures.

   H_{13a}: IPQ-scales and performance are negatively related.

   H_{13b}: HR and performance are negatively related.

4. To conclude the argument that performance was affected by presence and not vice versa, the level of difficulty was introduced as an independent variable. If performance was affected by level of difficulty but not presence, we could assume that the relationship is causal from presence to performance.

   H_{04a}: IPQ-scales between levels of difficulty are not significantly different.

   H_{04b}: HR between levels of difficulty are not significantly different.

   H_{14c}: Runs/speed/errors are higher in the easy tasks.

---

[13] HR: average heart rate

5. The discussion of the used measures revealed sub-hypotheses which needed to be observed.

   H_{15a}: There exists correlation between IPQ and HR.

   H_{15b1}: Spatial visualization and runs/speed/errors should be positively correlated.

   H_{15b1}: Spatial orientation and runs/speed/errors should be positively correlated.

   H_{15c}: Immersive tendencies and presence should be positively correlated.

CHAPTER 4: METHODOLOGY

4.1 Population, Sampling, and Data Acquisition

The population for the experiment was comprised of UCF students and members of the UCF's Institute for Training and Simulations (IST). The participants were equally distributed to the two experimental groups, defined by the sequence of experimental environments (VE-Mini or Mini-VE). In prior experiments the participants sometimes lacked the necessary commitment to perform at peak level. Since performance is one of the key factors of the experiment a financial incentive was introduced, depending on the overall performance in the tasks. All four tasks were conducted by the participants in one sequence to decrease variance and to enable a correlation analysis. The sequence of treatments was stratified to control and eliminate sequence-based influences.

To calculate the statistical power a-priory, G*Power 3 (Faul, Erdfelder, Lang, & Buchner, 2007) with effect sizes of 0.5 (medium) for differences of means and 0.4 (medium-high) for correlation analyses[14], was used. The a-priory analysis resulted in estimated sample sizes of 27 for means and 37 for correlation tests.

4.2 Experimental Design

4.2.1 Instrumentation

The experiment took place at the Team Performance Laboratory (TPL) in the Partnership II building of IST at UCF. The operator station both for the simulation and

[14] Defined by Cohen (Cohen, 1988), as cited by Faul et al. (Faul, Erdfelder, Lang, & Buchner, 2007).

the miniature robot were in one room. The miniature world was set up in a neighboring room and was not visible for the participant.

The virtual simulation used an exact replica of the miniature environment, including the robot (Figure 10). The simulation was running on two networked PCs, one as operator station with the egocentric view (Figure 11) and the second PC for observation and recording (top-down view, Figure 11). The VE was developed in Unity3D[15] and then implemented into the Sarge[16] simulation. The robot was operated with a joystick.

Figure 10: Robots (VE and Mini)

[15] Unity3D: multiplatform game development tool (http://unity3d.com/)
[16] SARGE: Search and Rescue Game Environment (http://sarge.sourceforge.net)

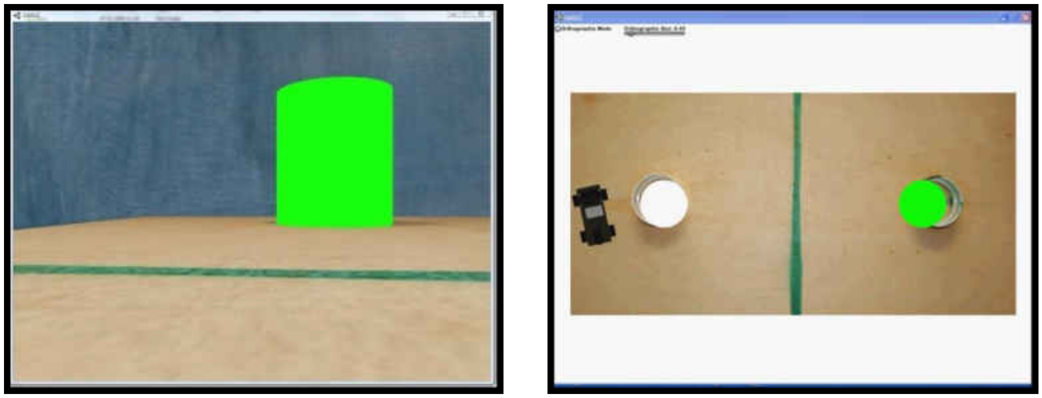


Figure 11: Views VE (Egocentric and Top-down)

The miniature experiment also used a PC with joystick as operator station. In this case, the robot was radio-controlled. The robot itself was a customized miniature vehicle with a fixed forward view camera; the additional 360 degree camera was not used (Figure 10). The egocentric camera view was then transmitted to the operator station. To support a post analysis of the experiment a top-down camera recorded the robot's movements.

Although the robot had wheels, it moved like a tank as so far as for forward/backwards movements all wheels turned forwards/backwards together. For turns while standing, the wheels on one side moved forward while the ones on the other side moved backwards. While driving forwards/backwards *and* steering, the respective side stopped while the other moved on (Figure 12). As result the robot turned on the spot when standing and turned more car-like when driving.

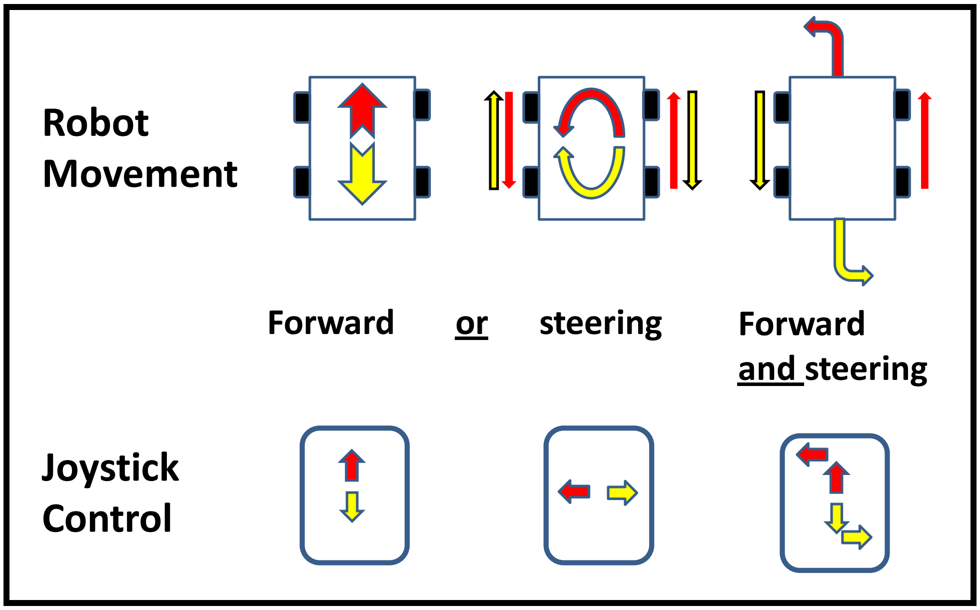


Figure 12: Robot Control

Heart rate was monitored with a POLAR Transmitter (Figure 13) and an RS 400 receiver which measured HR at 1/sec intervals. The data was transferred to a PC via the POLAR Trainer 5 software, which calculated the average HR during the tasks.

Figure 13: POLAR Transmitter

4.2.2 Staging of Scenario and Manipulation Check

An important feature of the experiment was the perceived danger to the miniature robot, affecting the sense of presence. During the experiment, two aspects had to be acknowledged with respect to the operator's psyche: monitoring the perceived danger ('manipulation check') and insuring this perceived danger to the equipment ('staging').

The assumption that the application type changes the user's attitude based on the perceived danger to equipment had to be confirmed in the experiment. This measurement of the independent variable is called the manipulation check (Sigall & Mills, 1998). The application of this check must ensure that it does not affect the participant's perception of this factor: 'Why, if what I was told is true, do they need my opinion?' (Sigall & Mills, 1998, p. 219). Hence, the manipulation check to confirm a changed perceived danger to the robot was administered as a questionnaire after the treatments in the respective environments.

Ford Morie et al. (2008) found that the setting and introduction of a simulation changed the perception of the task itself. They 'staged' identical scenarios in a gaming environment and as a military task. Soldiers reacted to this staging by higher involvement in the military environment. Applied to the given experimental setting it was necessary to enhance the motivation of the participant both to ensure the full commitment to the given tasks and to foster the perceived danger to the robot in the miniature world. Means to ensure the motivation and enhance the perceived danger were:

- A point-system based on the performance which increased or decreased the level of incentives the participant received after finishing the whole experiment. The higher performance (number of runs), the higher the financial incentive.
- The 'value' of the robot was exaggerated during the introductory phase and in the wording of the written introductions.

4.3 Tasks and Measurements

The task was to drive of the robot around two cylinders on a table (Figure 14). The imminent danger to the robot was to fall off the table. The level of difficulty was manipulated by different distances between the cylinders and the table edge: the distance was 190 mm (2x vehicle width) for the easy task and 135 mm for the difficult one (1.2x vehicle width). One 'run' was defined as completion of one lap around a cylinder and the crossing of the middle line (Figure 14).

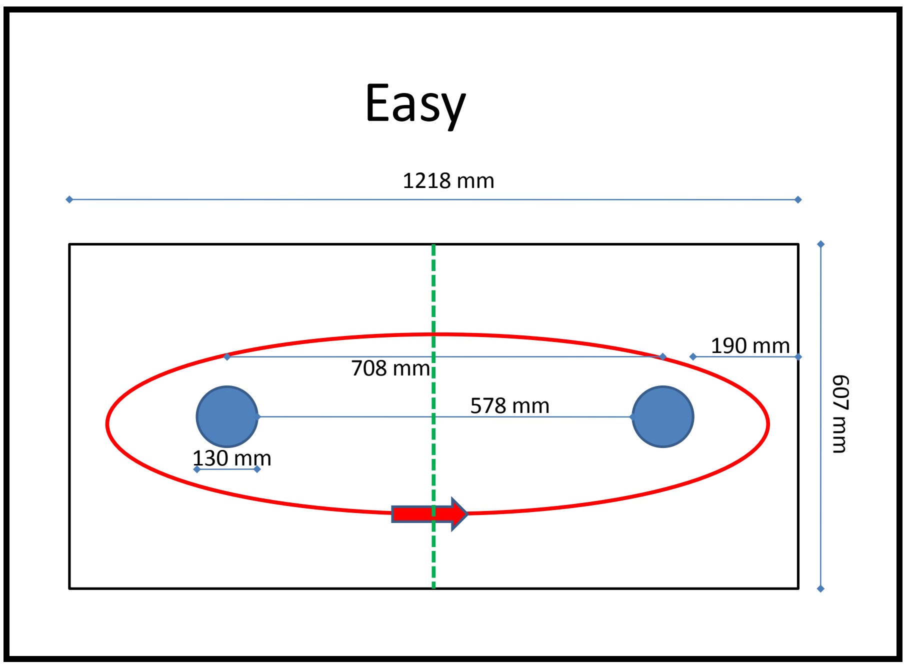


Figure 14: Task Outline

Quality/accuracy in the given tasks was defined as the avoidance of mistakes like hitting the cylinders or falling off the edge. If the robot fell off or was not operable at the edge, the task had to be restarted. Measure for quality was the number of restarts during one time period [errors]. The resetting, especially of the miniature world, required some time which was covered by an administrative penalty time of 1 minute. Hitting or sliding along the cylinders resulted in a loss of speed and thereby needed no immediate reaction/penalty.

Beside quality/accuracy of the robots operation, speed was an additional measure. It was technically difficult to establish a reliable measure for speed in the miniature world, but higher speed showed itself in a high number of runs in the given time [runs].

Based on the penalty time for errors, the number of runs as one measure for speed was directly correlated to the number of errors: the more errors the less operational time. An isolated measure for speed was derived by calculating the number of runs within the operational time [runs/min.]. The operational time was calculated from experimental time minus penalty time[17]. This measure allowed for the distinction between a participant who had a low number of runs due to a high error rate and another who had few runs due to a very careful operation (few errors).

As objective measure for presence, average heart rate (HR) was taken continuously during the tasks including the training phase to accustom the participant to the used system. The IPQ, as subjective measure, was administered after each treatment/task, including the practice phases, as was the NASA-TLX.

[17] The operational time cannot always be derived from the number of restarts. If an operational error happened within the last minute of the task the operational time was higher than 5 – N minutes and speed was calculated accordingly

Cognitive abilities were assessed in the initial phase of the experiment by the Guilford-Zimmerman tests for Spatial Visualization and Spatial Orientation, and the ITQ. The assessment of simulator sickness (SSQ) was conducted during the initial phase and after the final virtual experiment. Additional information about the participants (age, gender, gaming experience, etc.) was collected in the initial phase in a general questionnaire.

4.4 Summary of Measures

A summary of measures is shown in Table 6.

Table 6: Summary of Measures

Variable	Measure	Remarks
Sense of Presence	IPQ HR (average)	Relationship between IPQ and HR had to be observed.
Cognitive Abilities • spatial • immersive tendencies	Guilford-Zimmerman tests ITQ	
Performance	• number of runs • number of operational errors • average robot speed	
Level of difficulty/workload	NASA TLX	
Perceived danger to equipment	Questionnaire	
Simulator Sickness	SSQ	Affecting performance and presence
User Experience	Questionnaire	Gaming/VE experience

The design of the experiment ensured a minimal technical effect on performance between the two applications. Identical tasks on each difficulty level and within each application were used to prevent undesired performance differences between applications. The elimination of technical performance influences and the control of cognition-based effects assured that the observed performance changes were based on a difference in presence.

4.5 Procedure

Based on the task to circle the two cylinders in two different tasks levels and a familiarization phase, the sequence for each environment was practice phase (3 min.), easy task (5 min.), and difficult task (5 min.). Figure 15 shows the flowchart of the experiment.

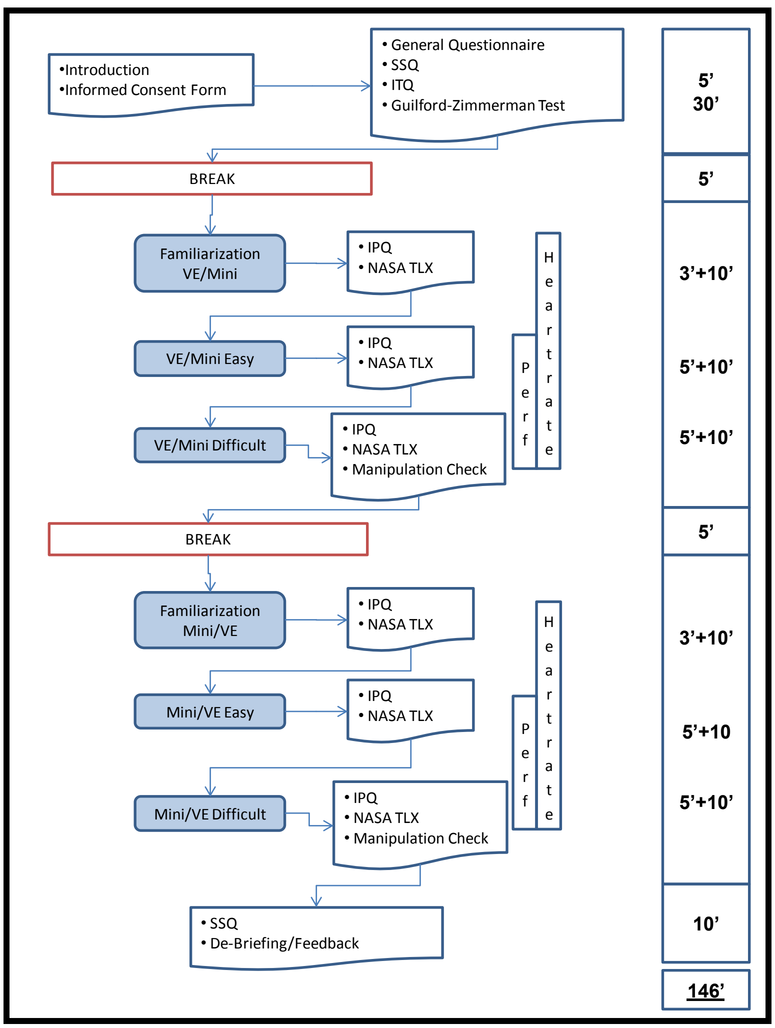


Figure 15: Flow Chart of Experiment

4.6 Summary

The aim of the experimental design was to measure presence, performance, and workload in four different treatments/settings. Based on the measurements the relationship between application type, presence, and performance could then be analyzed.

To minimize individual effects on the measurements, a within-subjects design for all four treatments was selected. This approach allowed the analysis of differences in the results between the treatments. The influence of training and eventually boredom based on the sequence and/or simplicity of the task had to be observed. The hypotheses and the respective relationships of measures are shown in table 7.

Table 7: Summary of Hypotheses and Measures

Hypothesis	Measures
H_1: Presence increases from simulation to miniature world.	H_{11a}: IPQ/HR (mini) > IPQ/HR (VE) H_{11b}: IPQ/HR correlated with VE/Mini
H_2: Performance decreases from simulation to miniature world.	H_{12a1}: runs (VE) > runs (Mini) H_{12a2}: speed (VE) > speed (Mini) H_{12a3}: errors (VE) > errors (Mini) H_{12b}: runs/speed/errors correlated with VE/Mini
H_3: Performance decreases when presence increases.	H_{13a}: IPQ negatively related to performance H_{13b}: HR negatively related to performance
H_4: Workload affects performance but not presence.	H_{04a}: IPQ (easy) ≈ IPQ (difficult) H_{04b}: HR (easy) ≈ HR (difficult) H_{14c}: performance (easy) > performance (difficult)
H_5: Assumptions from literature review	H_{12a1}: IPQ correlated with HR H_{12a1}: Spatial visualization correlated with performance H_{12a1}: Spatial orientation correlated with performance H_{12a1}: ITQ correlated with IPQ

Note: **Parametric statistics: t-test, Pearson's Correlation**
Non-parametric statistics: Wilcoxon Signed Rank test, Spearman's Rho

CHAPTER 5: RESULTS

5.1 Descriptive Statistics of Measurements

The statistical analysis was conducted with SPSS 12.5 and, if not otherwise stated, using α=0.05. In all cases data were tested for normality and parametric statistical analyses used were t-tests for means and Pearson correlation for relationships. For the non-parametric measures, the Wilcoxon-Signed Ranks test for medians and Spearman's Rho for correlations were used.

Table 8 to 12 show the means, standard deviations, minimum/maximum, and correlations for the demographic data and tests administered. The correlation tests are 2-tailed since they were not hypotheses-related. Additional analyses and results were shown in context with their respective hypothesis.

Overall 40 participants were recruited for the experiment; two did not finish due to acute nausea/dizziness. Additionally, two heart rate readings were not usable due to wrong data transfer of the transmitter. Thus, 38 datasets were available for the analysis using the presence questionnaire and 36 for heart rate based analyses.

Table 8: Means, Standard Deviations, and Correlations of Personal Data

		Age	Gender	Familiarity with Games	ITQ (Immersive Tendencies Questionnaire)	Spatial Orientation Test	Spatial Visualization Test
Age	Spearman Correlation	*M=19.87*	.205	.201	-.071	-.169	-.028
	Sig. (2-tailed)	*18/27*[b]	.218	.226	.672	.309	.866
	N	*38*	38	38	38	38	38
Gender [a]	Spearman Correlation		*M=0.6*	-.336	-.262	-.245	-.352
	Sig. (2-tailed)			.039	.107	.127	.026
	N		*40*	38	39	40	40
Familiarity with Games	Pearson Correlation			*M=3.34*	.539	.223	.357
	Sig. (2-tailed)			*SD=1.665* *1/6*[b]	.000	.177	.028
	N			*38*	38	38	38
ITQ (Immersive Tendencies Questionnaire)	Pearson Correlation				*M=4.226*	.117	.211
	Sig. (2-tailed)				*SD=.929* *2.44/6.39*[b]	.477	.198
	N				*39*	39	39
Spatial Orientation Test	Pearson Correlation					*M=15.056*	.639
	Sig. (2-tailed)					*SD=9.633* *1.5/39.0*[b]	.000
	N					*40*	40
Spatial Visualization Test	Pearson Correlation						*M=16.344*
	Sig. (2-tailed)						*SD=9.402* *-2.5/37.0*[b]
	N						*40*

[a] **Male = 0, Female = 1, *M* shows percentage of female in sample population.**
[b] **Min./Max.**

Table 9: Means, Standard Deviations, and Correlations for Measurements of the Simulation's Easy Task (VEE)

		IPQVEE	TLX VEE	HR VEE	MajVEE	SpVEE	RunsVEE
IPQVEE (Presence Questionnaire)	Pearson Correlation	*M=37.16*	-.070	.285			
	Sig. (2-tailed)	*SD=15.09*	.675	.087			
		3/68[d]					
	N	38	38	37			
TLX VEE (Workload Index)	Pearson Correlation		*M=709.24*	-.090			
	Sig. (2-tailed)		*SD=305.559*	.595			
			135/1300[d]				
	N		38	37			
HR VEE (Average Heart Rate)	Pearson Correlation			*M=79.34*			
	Sig. (2-tailed)			*SD=11.278*			
				60/105[d]			
	N			38			
MajVEE (Errors [a])	Spearman Correlation	.109	.375	.041	*M=0.44*		
	Sig. (2-tailed)	.513	.020	.806	*0/3[d]*		
	N	38	38	38	39		
SpVEE (Speed [b])	Spearman Correlation	.069	-.305	-.202	-.077	*M=6.378*	
	Sig. (2-tailed)	.679	.062	.224	.639	*SD=1.677*	
						1.8/8.2[d]	
	N	38	38	38	39	39	
RunsVEE (Number of Runs [c])	Spearman Correlation	-.023	-.398	-.061	-.487	.787	*M=28.99*
	Sig. (2-tailed)	.890	.013	.714	.002	.000	*SD=9.947*
							3/41[d]
	N	38	38	38	39	39	39

[a] Number of falls from the table.
[b] Number of rounded cylinders per minute.
[c] Number of rounded cylinders in 5 minutes.
[d] Min/Max

Table 10: Means, Standard Deviations, and Correlations for Measurements of the Simulation's Difficult Task (VED)

		IPQVED	TLX VED	HR VED	MajVED	SpVED	RunsVED
IPQVED (Presence Questionnaire)	Pearson Correlation	*M=35.58*	.056	.183			
	Sig. (2-tailed)	*SD=17.415*	.744	.285			
		1/71[d]					
	N	38	36	36			
TLX VED (Workload Index)	Pearson Correlation		*M=789.86*	-.052			
	Sig. (2-tailed)		*SD=315.181*	.769			
			25/1360[d]				
	N		36	34			
HR VED (Average Heart Rate)	Pearson Correlation			*M=76.61*			
	Sig. (2-tailed)			*SD=10.608*			
				57/102[d]			
	N			36			
MajVED (Errors [a])	Spearman Correlation	-.170	.206	.039	*M=1.03*		
	Sig. (2-tailed)	.308	.229	.823	*0/3*[d]		
	N	38	36	36	38		
SpVED (Speed [b])	Spearman Correlation	.196	-.448	.032	-.391	*M=4.955*	
	Sig. (2-tailed)	.238	.006	.855	.015	*SD=2.130*	
						1.22/8.8[d]	
	N	38	38	36	38	38	
RunsVED (Number of Runs [c])	Spearman Correlation	.204	.438	.023	-.669	.926	*M=20.97*
	Sig. (2-tailed)	.218	.008	.896	.000	.000	*SD=10.620*
							3/38[d]
	N	38	36	36	38	38	38

[a] Number of falls from the table.
[b] Number of rounded cylinders per minute.
[c] Number of rounded cylinders in 5 minutes.
[d] Min/Max

Table 11: Means, Standard Deviation, and Correlations for Measurements of the Miniature World's Easy Task (MiE)

		IPQ MiE	TLX MiE	HR MiE	MajMiE	SpMiE	RunsMiE
IPQMiE (Presence Questionnaire)	Pearson Correlation	*M=48.71*	.053	.190			
	Sig. (2-tailed)	*SD=16.064*	.754	.266			
		17/84[d]					
	N	38	37	36			
TLX MiE (Workload Index)	Pearson Correlation		*M=808.51*	-.013			
	Sig. (2-tailed)		*SD=321.830*	.942			
			75/1500[d]				
	N		37	36			
HR MiE (Average Heart Rate)	Pearson Correlation			*M=80.64*			
	Sig. (2-tailed)			*SD=12.278*			
				59/115[d]			
	N			36			
MajMiE (Errors [a])	Spearman Correlation	.034	.159	-.310	*M=0.18*		
	Sig. (2-tailed)	.839	.347	.066	*0/2*[d]		
	N	38	37	36	38		
SpMiE (Speed [b])	Spearman Correlation	-.122	-.311	.058	-.090	*M=5.791*	
	Sig. (2-tailed)	.467	.061	.738	.591	*SD=1.752*	
						2.0/8.6[d]	
	N	38	37	36	38	38	
RunsMiE (Number of Runs [c])	Spearman Correlation	-.123	-.312	.109	-.282	.975	*M=28.05*
	Sig. (2-tailed)	.461	.060	.528	0.86	.000	*SD=9.174*
							10/43[d]
	N	38	37	36	38	38	38

[a] Number of falls from the table.
[b] Number of rounded cylinders per minute.
[c] Number of rounded cylinders in 5 minutes.
[d] Min/Max

Table 12: Means, Standard Deviation, and Correlations for Measurements of the Miniature World's Difficult Task (MiD)

		IPQMiD	TLX MiD	HR MiD	MajMiD	SpMiD	RunsMiD
IPQMiD (Presence Questionnaire)	Pearson Correlation	*M=47.11*	-.043	.278			
	Sig. (2-tailed)	*SD=16.776*	.796	.091			
		11/84[d]					
	N	38	38	38			
TLX MiD (Workload Index)	Pearson Correlation		*M=962.63*	-.126			
	Sig. (2-tailed)		*SD=357.245*	.449			
			75/1510[d]				
	N		38	38			
HR MiD (Average Heart Rate)	Pearson Correlation			*M=79.45*			
	Sig. (2-tailed)			*SD=11.123*			
				60/106[d]			
	N			38			
MajMiD (Errors [a])	Spearman Correlation	-.274	.472	-.256	*M=1.42*		
	Sig. (2-tailed)	.096	.003	.121	*0/4*[d]		
	N	38	38	38	38		
SpMiD (Speed [b])	Pearson Correlation	-.081	-.246	.220	-.183	*M=3.9842*	
	Sig. (2-tailed)	.629	.137	.184	.271	*SD=1.779*	
						1.4/8.0[d]	
	N	38	38	38	38	38	
RunsMiD (Number of Runs [c])	Pearson Correlation	-.084	-.308	.319	-.401	.888	*M=14.95*
	Sig. (2-tailed)	.615	.060	.051	.013	.000	*SD=9.317*
			38				*1/36*[d]
	N	38		38	38	38	38

[a] Number of falls from the table.
[b] Number of rounded cylinders per minute.
[c] Number of rounded cylinders in 5 minutes.
[d] Min/Max

Table 13: Means of Main Measures

	IPQ (Presence)	HR (Heart Rate)	Errors	Speed	Runs	TLX (Workload Index)
Simulation Easy Task	37.16	79.34	.44	6.37	28.99	709.24
Simulation Difficult Task	35.58	76.61	1.03	4.95	24.47	784.86
Miniature World Easy Task	48.71	86.64	.18	5.79	28.05	808.51
Miniature World Difficult Task	47.11	79.45	1.42	3.98	14.95	962.63

5.2. Test of Manipulations

5.2.1. Fear of Damage

The hypotheses and hence the experiment were based on the theory that the operator experienced concern for the outcome of his actions. This manipulation, although supposed to be implemented in the experimental design, had to be validated. A questionnaire consisting of three items was used to evaluate the perceived threat to the robot in the two environments. It was hypothesized that the operator's perceived danger to the robot increased in the miniature world.

The manipulation check for fear of damage showed that the perceived threat to the robot increased from 2.6 in the simulation to 3.5 in the miniature world (on a scale from 1-7). The Wilcoxon Signed Ranks test confirmed a statistical significance of this difference between simulation and miniature world ($z = -2.962$, p (1-tailed) = .0015), with a medium effect size of r=0.34 (Cohen 1988). This confirmed that the different application types affected the perceived danger to the robot of the participants.

5.2.2 Task Difficulty

The high difficulty condition of the experimental design induced a higher risk by reducing the gap between cylinders and table edge, but a validation of this alteration's impact on the perceived workload between the two levels of difficulty was necessary. It was hypothesized that the workload index is higher for the difficult task (TLX Easy < TLX Diff).

Table 14: Paired Sample t-test for Workload between Difficulty Levels within Environment

	Paired Differences							
	Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference				
				Lower	Upper	t	df	Sig. (1-tailed)
TLX VE Easy - TLX VE Diff [a]	-87.618	207.407	35.570	-159.986	-15.250	-2.463	33	.0009
TLX Mi Easy - TLX Mi Diff [b]	-130.588	214.715	36.823	-205.506	-55.671	-3.546	33	.0005

[a] **Difference of workload between the simulation's easy and difficult tasks**
[b] **Difference of workload between the miniature world's easy and difficult tasks.**

The perceived workload, measured by the NATO-TLX, increased from 709.24 (easy task) to 789.86 (difficult task) in the simulation and from 808.51 to 962.63 in the miniature world (on a scale from 0-1500). The t-test for paired samples showed a significant increase in workload from easy to difficult tasks for both environments (Table 14). It was concluded that the change in the cylinder position introduced a significant change in perceived workload. To allow for an assessment of the effects size Eta Squared was calculated and with 0.17 for easy and 0.33 for difficult tasks, the effects were assessed as large (Cohen, 1988).

5.3 Tests of Hypotheses

5.3.1 Hypothesis 1

The internal consistency test for IPQ and HR (Table 15) between difficulty levels but within the application type (VE/Mi) resulted in Cronbach's alpha coefficients of 0.912 for presence in the simulation, 0.898 for presence in the miniatures world, 0.978 for heart rate in the simulation, and 0.922 for heart rate in the miniature world. This supported the use of collapsed values for IPQ/HR for each environment as an average of the respective difficulty levels. The calculated average measurements are shown in Table 16. The means are 36.33 for presence in the simulation, 47.91 for presence in the miniature world, 79.14 BPM[18] for simulation, and 79.66 BPM for the miniature world.

Table 15: Reliability Analysis for Presence and Heart Rate Within the Same Environment

Items	Cronbach's Alpha	N
IPQVEE – IPQVED [a]	0.912	38
IPQMiE – IPQMiD [b]	0.898	38
HRVEE – HRVED [c]	0.978	36
HRMiE – HRMiD [d]	0.922	36

[a] Presence in simulation easy to difficult task.
[b] Presence in miniature world easy to difficult task.
[c] Heart rate in simulation easy to difficult task.
[d] Heart rate in miniature world easy to difficult task.

[18] Beats per minute

Table 16: Descriptive Statistics for Means IPQ and HR

	N	Minimum	Maximum	Mean	Std. Deviation
IPQVE	38	3.50	67.00	36.3289	15.63359
IPQMi	38	14.00	84.00	47.9079	15.64896
HRVE	38	58.50	103.50	79.1447	10.93155
HRMi	38	59.50	109.50	80.526	11.39985

Note: IPQ = Presence, HR = Heart Rate, VE= Simulation, Mi = Miniature World

H_{11a}: IPQ/HR (miniature world) > IPQ/HR (simulation):

It was hypothesized that perceived presence was higher in the miniature world [Mi] than in the simulation [VE]. The paired sample t-test for the IPQ resulted in a statistically significant increase of presence from simulation (*M*=36.33, *SD*=15.63) to miniature world (*M*=47.91, *SD*=15.65), $t(37)=-5.632$, $p<.0005$ (1-tailed). The eta squared statistic (.74) indicated a large effect size.

Table 17: Paired Sample t-test IPQ and HR between Environments

	Paired Differences							
				95% Confidence Interval of the Difference				
	Mean of Difference	Std. Deviation	Std. Error Mean	Lower	Upper	t	df	Sig. (1-tailed)
IPQVE – IPQMi [a]	-11.57895	12.67394	2.05598	-15.74477	-7.41313	-5.632	37	.000
HRVE – HRMi [b]	-.789	4.00482	.64967	-2.22425	.40846	-1.397	37	.0855

[a] Difference in presence between simulation and miniature world.
[b] Difference in heart rate between simulation and miniature world.

The t-test for HR showed a statistically significant increase from simulation (*M*=79.15, *SD*=10.93) to miniature world (*M*=80.53, *SD*=11.4) at α=0.1, $t(37)=-1.397$, $p=.0855$ (1-tailed). The eta squared statistic (.05) indicated a small effect size.

The hypothesis was supported for a significant difference in perceived presence (IPQ) between the simulation and the miniature world (Table 17). It was also supported for heart rate, but with less statistical significance.

H_{11b}: IPQ/HR positively correlated between simulation and miniature world:

The perceived presence measured either with IPQ or HR was hypothesized to be positively correlated with the application type: the higher IPQ/HR in the simulation [VE], the higher in the miniature world [Mi].

The relationships between IPQ/HR within the simulation and the miniature world were investigated using Pearson product-moment correlation coefficients. There was a strong, positive correlation between perceived presence (IPQ) in the simulation and in the miniature world (r=0.672, N=38, $p<.0005$). For HR the correlation also was strong and positive (r=0.912, N=38, $p<.0005$). The results supported the hypothesis of a linear and positive correlation for perceived presence between the environments.

5.3.2 Hypothesis 2

The hypotheses were based on the assumptions that performance would be higher in the simulation.

H_{12a1}: runs (simulation) > runs (miniature).

H_{12a2}: speed (simulation) > speed (miniature).

H_{12a3}: errors (simulation) > errors (miniature).

An internal consistency test for performance (Table 18) within the application type (VE/Mi) using Spearman's rho showed strong correlations for speed (r_s=.625 [VE], r_s=.741 [Mi]) and runs (r_s=.656 [VE], r_s=.592 [Mi]) but not for errors (r_s=.287 [VE], r_s=.193 [Mi]) (Table 18). This supported the use of collapsed values for speed and runs, but not errors, the calculated values are shown in Table 19. In the following, the hypotheses will be tested with those collapsed means for speed (SpeedVE, SpeedMi) and runs (RunsVE, RunsMi), while errors related hypotheses will be tested using all four treatment results.

Table 18: Reliability Analysis Performance Measures using Spearman's Rho

Items	Spearman's Rho	Sig. (2-tailed)	N
Speed Simulation Easy – Speed Simulation Diff	.625	.000	38
Speed Miniature Easy – Speed Miniature Diff.	.741	.000	38
Errors Simulation Easy – Errors Simulation Diff.	.287	.080	38
Errors Miniature Easy – Errors Miniature Diff.	.193	.246	38
Runs Simulation Easy – Runs Simulation Diff.	.656	.000	38
Runs Miniature Easy – Runs Miniature Diff.	.592	.000	38

Table 19: Descriptive Statistics Speed and Runs

	N	Minimum	Maximum	Mean
Speed Simulation	38	1.78	8.10	5.6897
Speed Miniature	38	1.90	7.80	4.8864
Runs Simulation	38	3.83	39.00	25.0613
Runs Miniature	38	7.50	37.00	21.5000

Hypothesis 2 assumed that speed, runs, and errors would increase in the simulation. The Wilcoxon signed rank test (Table 20) showed a significant decrease in speed from simulation (*M*=5.69) to miniature world (*M*=4.89), z=-3.61, p<.001 (1-tailed). The test results for runs also showed a significant decrease from simulation (*M*=25.06,) to miniature world (*M*=21.5), z=-2.2822, p<.005 (1-tailed). The results for errors in the easy tasks showed a significant *decrease* in errors from simulation (*M*=.44) to miniature world (*M*=.18), z=-2.236, p<.05 (1-tailed), while the results for difficult tasks showed a significant *increase* in errors from simulation (*M*=1.03) to miniature world (*M*=1.42), z=-1.842, p<.05.

Table 20: Wilcoxon Signed Rank Test for Performance Measures

	Speed Mi - VE	Runs Mi - VE	Errors Easy Task Mi - VE	Errors Diff. Task Mi - VE
Z	-3.606	-2.822	-2.236	-1.842
Asymp. Sig. (1-tailed)	.000	.0025	.0125	.0325

Note: miniature world = Mi, simulation = VE.

In summary the hypothesis of decreased performance in the miniature worlds compared to the simulation was supported for speed, runs, and errors in easy tasks, but not for errors in difficult tasks.

H_{12b}: Runs/speed/errors are positively correlated with miniature world/simulation.

This hypothesis was based on the assumption that a higher performance in the simulation would be associated with a higher performance in the miniature world. The

Spearman's rho correlation test showed a significantly strong correlation for speed (r_s=.714, N=38, p<.001) and runs (r_s=.734, N=38, p<.001), but no statistically significant results for errors (Table 21).

Table 21: Spearman's Rho for Performance Measures between Applications (VE, Mi)

Relationship	Spearman's rho	Sig. (1-tailed)	N
SpeedVE - SpeedMi	0.714	.000	38
RunsVE - RunsMi	0.734	.000	38
Errors VE easy – Errors Mi easy	0.228	.085	38
Errors VE diff. – Errors Mi diff.	0.043	.399	38

Note: simulation = VE, miniature world = Mi.

The hypotheses were supported for speed and runs but not for errors. A further analysis of the errors data showed that the number of errors was low, which introduced a large number of ties into the Spearman's correlation test. A different, proportion-based, analysis of the errors showed that 33 (87%) of the participants improved their error rate or stayed at 0-errors in the easy task, and 11 (29%) improved in the difficult task. A z-test of the proportions for easy tasks showed significant support for a trend in error-improvement in easy tasks, z=4.5617, p<.001, with supports the hypothesis of a positive relationship. The low trend in difficult task (<50%) did not support the hypothesis.

5.3.3 Hypothesis 3

H_{13a}: IPQ-scales and performance are negatively related.

H_{13b}: Heart rate and performance are negatively related.

This hypothesis was based on the assumption that a higher presence leads to better performance, in this case: with higher presence a more ‘real’ performance was expected. In the given setting this meant lower speed, runs, and fewer errors due to an increased fear of damage to the robot.

An analysis of the data showed that there was no numerical correlation between increase of IPQ/HR and performance. A higher increase of IPQ/HR did not translate into a comparable equal decrease of performance. Both, presence and performance were individual and not comparable in their numerical values, so that a proportional approach was used: the percentage of participants who followed the hypothesis was calculated and tested for its significance.

Table 22 and Table 24 show the calculated changes in presence (or heart rate), speed, runs, and errors from simulation to miniature world. The hypothesis assumed positive changes in presence or heart rate and subsequently negative changes in performance measures. For the rare cases of a decreased presence (or heart rate) the performance measures were supposed to increase. All cases that followed this hypothesis were marked with ‘+’.

H_{13a}: IPQ negatively related to performance.

Table 22: Changes in IPQ and Performance from Simulation to Miniature World

	Change of Presence	Change of Speed		Change of Runs		Change of Errors Easy Task [a]		Change of Errors Diff, Task [a]	
1	31.5	-1.6	+	-4.00	+	-1	+	-2	+
2	40.5	-3.3	+	-12.00	+	0	+	0	+
3	39.5	-0.9	+	-4.50	+	0	+	0	+
4	1.5	1.1	-	5.00	-	0	+	1	-
5	-9.0	-2.1	-	-9.00	-	0	-	1	-
6	3.0	-1.1	+	-1.50	+	0	+	0	+
7	4.0	-1.7	+	-13.00	+	-1	+	2	-
8	14.5	-2.4	+	-8.50	+	0	+	-1	+
9	25.0	0.4	-	-0.50	+	0	+	2	-
10	3.0	-1.3	+	-3.00	+	0	+	-1	+
11	20.0	0.0	-	-1.00	+	0	+	1	+
12	37.5	-1.0	+	8.67	-	0	-	-2	-
13	10.5	-1.1	+	-13.00	+	0	+	1	-
14	14.0	-0.2	+	-4.50	+	0	-	2	-
15	14.0	-1.9	+	-8.50	+	-1	+	2	-
16	1.0	-3.3	+	-14.00	+	-1	+	0	+
17	15.0	0.4	-	4.50	-	-1	+	0	+
18	4.5	1.0	-	5.00	-	0	+	0	+
19	31.0	-1.7	+	-10.50	+	0	+	3	-
20	17.5	-0.2	+	2.00	-	-1	+	-1	+
21	10.5	-1.8	+	-2.00	+	-2	+	0	+
22	-4.0	-0.2	-	-8.00	-	0	-	2	-
23	10.5	-0.7	+	-3.50	+	0	+	2	-
24	5.5	0.2	-	3.00	-	-1	+	0	+
25	4.0	-0.8	+	-4.00	+	-2	+	3	-
26	8.5	0.7	-	7.50	-	-1	+	0	+
27	2.5	-2.3	+	-9.00	+	0	+	0	+
28	1.5	-0.7	+	-8.00	+	0	+	2	-
29	24.0	-1.2	+	-13.50	+	1	+	2	-
30	24.5	-0.4	+	-3.50	+	0	+	-1	+
31	8.5	-1.1	+	-3.50	+	0	+	-2	+
32	-1.5	0.0	-	6.50	-	-1	-	-1	-
33	-0.5	0.1	+	1.50	+	0	-	-1	-
34	11.0	2.0	-	9.00	-	0	+	0	+
35	9.0	-2.1	+	-10.50	+	0	+	0	+
36	15.0	-0.4	+	-5.00	+	1	-	0	+
37	-4.5	-1.5	-	-11.50	-	1	+	1	-
38	-3.0	0.3	+	1.50	+	0	-	0	-

<u>Note:</u> + = measure followed hypothesis, -= measure did not follow,
[a] No errors in both environments was assessed as +.

Table 22 showed that 26 participants (68%) changed their speed in accordance with the change in IPQ, 26 (68%) for number of runs, 30 (79%) for errors (easy task), and 20 (53%) for errors (difficult). To support the hypothesized relationship between presence and performance it had to be shown that the probabilities are significant higher than 50% (randomness). More than 50% probability would show the findings of negative relationships between presence and performance measures were not random. A z-test for large samples was conducted; with H_0: p=0.5, H_1: $p>p_0$, $n*p_0>4$, and $n*(1-p_0)>4$. For all four performance measures, the z-test for proportions showed the assumed significant direction for speed ($p<.05$), runs ($p<.05$), and errors in easy tasks ($p<.001$). The hypothesis was supported for runs, speed, and errors in easy tasks, but not for errors in difficult tasks (Table 23).

Table 23: Z-test for Performance Changes

Negative Relationship	Proportion following the hypothesis	p-value
Presence - Speed	68%	0.0136
Presence - Runs	68%	0.0136
Presence -Errors (easy task)	79%	0.0005
Presence -Errors (diff. task)	20%	0.34

Note: H_0: p=0.5, H_1: $p>p_0$. $n*p_0>4$ and $n*(1-p_0)>4$.

H_{13b}: HR positively correlated to performance.

Table 24: Changes in Heart Rate and Performance from Simulation to Miniature World

	Change of Heart Rate	Change of Speed		Change of Runs		Change of Errors Easy Task [a]		Change of Errors Diff- Task [a]	
1	-2.50	-1.56	-	-4.00	-	-1	-	-2	-
2	1.00	-3.25	+	-12.00	+	0	+	0	-
3	-0.50	-0.90	-	-4.50	-	0	-	0	-
4	3.00	1.12	-	5.00	-	0	+	1	-
5	-2.00	-2.06	-	-9.00	-	0	-	1	-
6	1.50	-1.10	+	-1.50	+	0	+	0	+
7	-5.50	-1.72	-	-13.00	-	-1	-	2	-
8	3.00	-2.40	+	-8.50	+	0	+	-1	+
9	1.50	0.43	-	-0.50	+	0	+	2	-
10	6.00	-1.25	+	-3.00	+	0	+	-1	+
11	0.00	0.00	-	-1.00	-	0	-	1	-
12	0.50	-1.02	+	8.67	-	0	-	-2	+
13	-0.50	-1.08	-	-13.00	-	0	-	1	-
14	8.50	-0.18	+	-4.50	+	0	-	2	-
15	2.00	-1.87	+	-8.50	+	-1	+	2	-
16	-0.50	-3.29	-	-14.00	-	-1	-	0	-
17	-4.50	0.40	-	4.50	-	-1	-	0	-
18	2.00	1.00	-	5.00	-	0	+	0	+
19	-1.50	-1.65	-	-10.50	-	0	-	3	-
20	-6.00	-0.20	-	2.00	-	-1	-	-1	-
21	1.00	-1.75	+	-2.00	+	-2	-	0	-
22	1.50	-0.20	+	-8.00	+	0	+	2	-
23	3.50	-0.70	+	-3.50	+	0	+	2	-
24	-3.50	0.19	+	3.00	+	-1	+	0	-
25	-2.00	-0.80	-	-4.00	-	-2	-	3	-
26	-3.00	0.69	+	7.50	+	-1	-	0	-
27	0.00	-2.28	-	-9.00	-	0	+	0	+
28	-2.00	-0.74	-	-8.00	-	0	-	2	-
29	3.50	-1.16	+	-13.50	+	1	-	2	-
30	3.50	-0.43	+	-3.50	+	0	+	-1	+
31	16.00	-1.09	+	-3.50	+	0	+	-2	+
32	4.50	-0.04	+	6.50	-	-1	+	-1	+
33	2.00	0.13	-	1.50	-	0	+	-1	+
34	3.50	2.00	-	9.00	-	0	+	0	-
35	2.00	-2.10	+	-10.50	+	0	+	0	+
36	-2.50	-0.44	-	-5.00	-	1	-	0	-
37	3.00	-1.53	+	-11.50	+	1	+	1	+
38	2.00	0.30	-	1.50	-	0	+	0	+

Note: + = measure followed hypothesis, -= measure did not follow
[a] **No errors in both environments was assessed as +.**

Similar to presence, this hypothesis was based on the assumption that a higher heart rate was related to a higher concern for the outcome and hence a decrease of performance. Table 24 showed that 18 (47%) participants changed their speed in accordance with the change in HR, 17 (45%) in number of runs, 19 (50%) in errors (easy task), and 13 (34%) for errors (difficult). Since all proportions are ≤ 50%, a relationship between HR and performance measures cannot be assumed.

5.3.4 Hypothesis 4

Hypothesis 4 was based on the possibility that perceived performance can affect perceived presence in post-test questionnaires. To counteract this possibility the construct of workload was introduced to test for this potential performance-presence relationship. It was hypothesized that in the experimental setting, different levels of workload, hence performance, resulted in similar presence in equal environments. A change of performance based on workload but not presence would prove that the observed presence measures were not performance related. The hypothesis consisted of two parts: (a) part 1 assumed no change of presence between levels of difficulty, while (b) part 2 assumed a significant change in performance between difficulty levels. For part 1 the null hypothesis should not be rejected, while the null hypothesis should be rejected for part 2.

H_{04a}: IPQ-scales between levels of difficulty are not significantly different.

H_{04b}: HR between levels of difficulty are not significantly different.

Table 25: T-Test for IPQ and HR within Environments

	Paired Differences							
	Mean Difference	Std. Deviation	Std. Error Mean Difference	95% Confidence Interval of the Difference				
				Lower	Upper	t	df	Sig. (2-tailed)
IPQVEE – IPQVED [a]	1.485	9.824	1.710	-1.999	4.968	.868	32	.392
IPQ MiE – IPQMiD [b]	1.788	10.568	1.840	-1.959	5.535	.972	32	.338
HR VE easy – HR VE diff [c]	.529	3.203	.549	-.588	1.647	.964	33	.342
HR Mini easy - HR Mini diff [d]	1.265	3.840	.659	-.075	2.605	1.920	33	.063

[a] Difference in presence between easy and difficult task in simulation.
[b] Difference in presence between easy and difficult task in miniature world.
[c] Difference in hear rate between easy and difficult task in simulation.
[d] Difference in heart rate between easy and difficult in miniature world.

With an absence of significant proof for a difference in means for all presence and heart rate measures, the null hypothesis was not rejected. Although it is not possible to support the null hypothesis, it can be assumed that there was no workload/performance induced change of presence.

H_{14c}: *Runs/speed/errors are higher in the easy tasks.*

Table 26: Wilcoxon Signed Ranks Test for Performance Measures between Difficulty Levels

	Speed Easy - Difficult		Runs Easy - Difficult		Errors Easy - Difficult	
	Simulation	Miniature	Simulation	Miniature	Simulation	Miniature
Z	-4.882	-5.130	-4.558	-5.305	-3.065	-4.634
Asymp. Sig. (1-tailed)	.000	.000	.000	.000	.001	.000

The Wilcoxon signed ranks test for changes in performance based on workload showed a statistically significant difference from easy to difficult (Table 26). In the simulation the mean decrease was 2.44 [runs/min.] for speed and 8.18 for runs, and an increase of 0.58 for errors. In the miniature world the decrease was 1.8 for speed and 12.1 for runs, while the errors increased by 1.24. This showed that the introduced levels of difficulty significantly affected performance.

Overall the hypothesis that the introduced level of difficulty affected performance but not presence (as measured by HR and IPQ) was supported. Hence it was concluded that the observed changes in presence were not induced by changes in performance.

5.3.5 Hypothesis 5

H_{15a}: *There exists positive correlation between IPQ and HR.*

Since HR and IPQ were assumed to measure the same construct of presence, a positive correlation was expected. The Person correlation test showed a small positive correlation between the IPQ and HR in both environments (simulation: r=.262, p=.028

(1-tailed), *N*=38; miniature world: *r*=.209, *p*=.052(1-tailed), *N*=38). The hypothesis was supported with an increased $\alpha = 0.1$.

H_{15b1}: Spatial visualization and runs/speed/errors should be positively correlated.

H_{15b2}: Spatial orientation and runs/speed/errors should be positively correlated.

The review of literature had shown that the spatial ability tests and performance should be positively correlated.

Table 27: Spearman's Rho for Spatial Abilities and Performance Measures

		Speed		Runs		Errors			
		Simulation	Miniature	Simulation	Miniature	Simulation Easy	Miniature Easy	Simulation Diff.	Miniature Diff.
Spatial Orientation Test	Correlation Coefficient	.380	.099	.305	.199	-.039	-.197	.193	-.345
	Sig. (1-tailed)	.009	.277	.031	.115	.408	.117	.123	.017
Spatial Visualization Test	Correlation Coefficient	.327	.175	.300	.216	-.056	-.215	-.050	-.208
	Sig. (1-tailed)	.022	.146	.033	.097	.369	.098	.383	.105

Note: *N*=38

The analyses showed a positive correlation between the spatial orientation tests and speed/runs in the simulations (speed: r_s=.380, *N*=38, *p*<.05; runs: r_s=.305, *N*=38, *p*<.05). In the miniature world only a negative correlation between spatial orientation and errors could be found for difficult tasks (r_s=-.345, *N*=38, *p*<.05). For spatial visualization only a correlation for speed and runs in the simulation was found with statistical significance (speed: r_s=.327, N=38, p<.05; runs: r_s=.300, N=38, p<.05) All data can be found in Table 27. Overall the hypotheses only were significantly supported for speed and runs in the simulations.

H_{15c}: Immersive tendencies and presence should be positively correlated.

The immersive tendencies questionnaire (ITQ) was supposed to test the individual tendency to experience presence and hence should be positively correlated to the presence questionnaire (IPQ). The analysis showed a small correlation for the relationship between ITQ and IPQ in the simulation with an increased used α=0.1 (r=.175, N=38, p=.0735). The correlation was medium between ITQ and IPQ in the miniature world with a used α=0.05 (r=.425, N=38, p=.0025). The hypothesis was supported for both environments.

CHAPTER 6: DISCUSSION AND CONCLUSION

6. 1 Overview of Results

Table 28: Overview of Hypothesis Tests

			Comparison of Means	Tests for Relationships
H_1	IPQ		+ [a]	+ [a]
	HR		+ [b]	+ [a]
H_2	Runs		+ [a]	+ [a]
	Speed		+ [a]	+ [a]
	Errors	Easy	+ [a]	+/- [c]
		Difficult	-	-
H_3	IPQ			+/- [d]
	HR			-
H_4	IPQ		+ [e]	
	HR		+ [e]	
	Runs		+ [a]	
	Speed		+ [a]	
	Errors	Easy	+ [a]	
		Difficult	+ [a]	

Note: hypothesis supported: +, not supported: - .
[a] α=0.05
[b] α=0.1
[c] No significance in correlation tests, but positive results for test of proportions.
[d] Trend supported for speed, runs, errors (easy), but not errors (diff.).
[e] Null hypothesis not rejected.

Table 28 shows the results of the hypothesis tests. In the following, the results and their implications with respect to the used constructs and assumed relationships are discussed.

6.2 Relationships of Constructs

6.2.1 Relationships of Presence Measures and Application Type

Presence or telepresence as the feeling of 'being in' the VE or remote environment was measured by the subjective igroup presence questionnaire (IPQ) and the objective average heat rate (HR). It was assumed that perceived presence increases in the miniature world. While IPQ-scores increased by a statistically significant 32%, HR only increased by 0.6% using an α of 0.1.

The correlation analysis as the second part of the hypothesis showed significant positive correlation between IPQ-scores in the simulation and the miniature world. Although the results for HR were also significant, the small difference in means does not advocate the use of HR as a measure for presence.

The IPQ showed that the participants clearly assessed the miniature world as more real than the VE. At the same time, HR was not equally distinct as an objective measure for increased presence. As mentioned before, the previous studies using HR as an objective measure used arousing or frightening scenarios, which was not the case for the scenario used here to drive around two cylinders. The findings answer the question by Slater et al. about HR measured in beats per minute as presence-measure in non-stressful environments (Slater, et al., 2006): the average heart rate was not a clear measure in the given scenario. It seems that the given miniature world did not induce sufficient stress to trigger a significant change of HR.

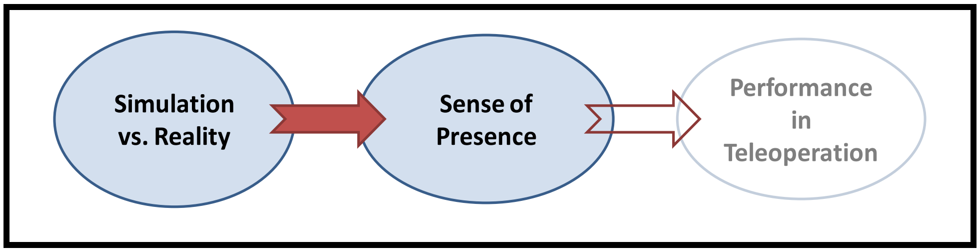


Figure 16: Constructs Hypothesis 1

Nevertheless it was concluded that the first part of the proposed construct interactions, the direct relationship between presence and application type is supported by the experimental data (Figure 16). The more 'real' the environment was perceived, the more increased presence. The basic idea of the construct of presence as a measure for the operator's 'disbelief' towards the remote environment was supported since the real environment scored statistically significant higher in the IPQ.

6.2.2 The Relationship of Performance Measures and Application Type

It had been hypothesized that the performance measures speed, runs, and errors would decrease in the miniature world. Indeed, speed and runs *decreased* with significance by 14% and errors in easy tasks by 60%. The number of errors in the difficult task showed a statistically significant *increase*, contrary to the hypothesis, by 38%.

Observations during the experiment helped to explain the unexpected findings for errors in difficult tasks. The easy task did not enforce errors by its setting since the distance between cylinders and table edge was large (2x vehicle). Concurrently, major errors (falling off the table) were purely based on unnecessary operational mistakes, which seemed to be based on lower concern for the outcome. The difficult task with its much smaller distance between cylinder and edge (1.2x vehicle) required much more operational skill to prevent falls. Hence the number of errors was also a measure of such skills, not only carelessness. Additionally, the controls of the robot had an unexpected impact on performance: the more car-like behavior when the operator drove forward while steering reduced the margin of error especially in the area between cylinders and table edge; a more ‘courageous’ behavior more often seen in the simulation. If the participant got too careful/scared and only used skid steering (steering while standing) behind the cylinders, the margin of error increased significantly. Skid steering behind the cylinders almost always resulted in a fall in the difficult task (Figure 17), since the robot pushed itself off the table. This error was more often observed in the miniature world. A quote from one participant:”I was surprised how easy the difficult task got, when I was no longer scared”. This participant performed significantly better in the second difficult task, when she drove faster. In summary this operational behavior of the robot turned around the hypothesis for difficult tasks: the more afraid the participants were, the more errors occurred. This case showed how easily an increase of operational requirements can lead to confounded relationships between presence, skills, and performance. Future experiments must account for skill related changed in performance.

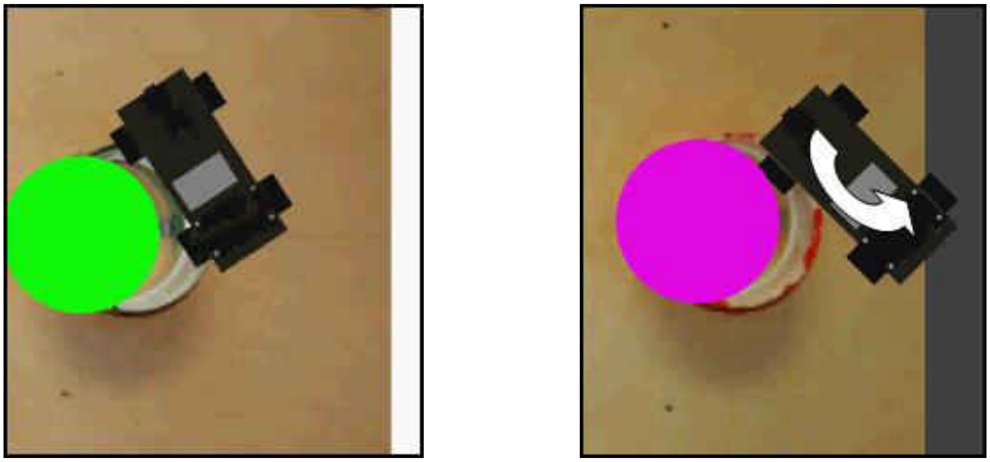


Figure 17: Skid Steering Behind Cylinder (easy/difficult)

The second part of the hypothesis, a positive correlation between the performance measures and the application type, was supported with significance and large correlation factors for speed and runs, but not for errors. A secondary analysis of the probabilities for errors showed significant hypothesis support for the easy task but not the difficult one, based on proportions. The missing positive relationships between application type and error rate in the difficult task was based in the task's operational demands, as described before. The prior described aspect of confounded constructs, here operational skills and presence, occurred in this setting and the available data did not allow for a distinction between skill-induced errors and presence-/fear-induced ones.

But overall, the assumed relationship between performance and application type was supported for speed, runs, and errors in easy tasks: the participants performed more carefully (slower, less errors) in the miniature world, compared to the simulation.

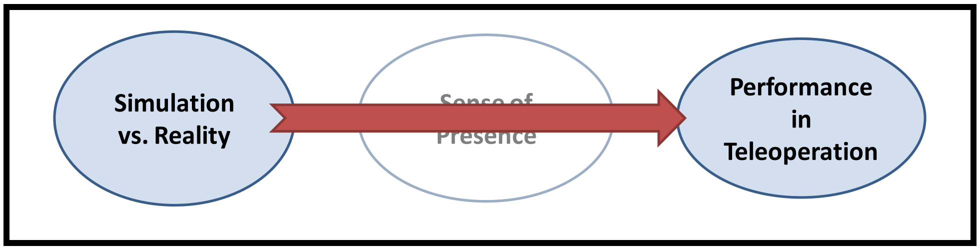


Figure 18: Constructs Hypothesis 2

6.2.3 Relationship of Presence and Performance Measures

To complete the model, the third hypothesis was based on the assumption that increased presence leads to more careful behavior, hence decreased speed, number of runs, and errors. The previous analysis of the difficult task had shown the confounded relationship of operational skills and presence, which made the separation and analysis of presence-related performance changes impossible. Consequently, the following analysis excludes the measures for errors in difficult tasks.

Performance and sense of presence are highly individual constructs for which the measurements are also highly individual. Hence, a primary correlation analysis of the change in presence and performance from simulation to miniature world showed that the increase of presence did not qualitatively correspond to a similar decrease in performance: if somebody perceived a large increase in presence (measured by IPQ), it did not result in an equally large decrease in performance. On the other hand, the z-test for proportions showed that there was a significant probability that an increase in presence (IPQ-scales) results in decreased performance.

The primary correlation analysis of HR also showed no significant results, but contrary to the IPQ to performance relationship, the proportions for HR to performance were all under 50%, which did not support the hypothesis. This result was a confirmation of the previous findings for HR as an inappropriate presence measure in the given scenario.

The tests showed a negative relationship between presence (measured by IPQ) and performance. Nevertheless this relationship was qualitative and could not be quantified as statistically significant correlation.

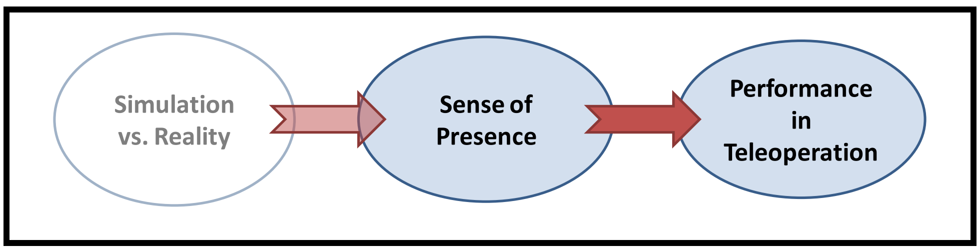


Figure 19: Constructs Hypothesis 3

6.2.4 Relationships of Workload, Performance, and Presence

Slater et al. (1999), like others, defined four factors influencing presence: (a) control factors, (b) sensory factors, (c) distraction factors, and (d) realism factors. In partial concordance, the IPQ items were divided by Schubert et al. (2001) into (a) general presence, (b) spatial presence, (c) involvement, and (d) experienced realism. With subjective influences like control factors or experienced realism, the construct of presence and the respective questionnaires are all prone to changes based on the perception of control by the participant. It could be theorized that better performance leads to a higher perception of control and hence higher perceived presence. To support the proposed directional relationship from presence to performance, two analyses with respect to workload (changed performance) were necessary: (a) a significant change in workload and hence performance did not affect presence, but (b) the different workloads affected performance.

a) The t-test for the IPQ-scores between levels of difficulty (IPQVEE – IPQVED and IPQMiE-IPQMiD[19]) showed no significant difference in means. In conjunction with the strong correlation between the respective IPQ measurements, this result supported the hypothesis that the perceived presence was not primarily affected by the change of difficulty. Based on the conclusion that HR was not a viable measure for presence, the HR results could not be used to support the hypothesis.
b) For the second part of the analysis, the Wilcoxon signed rank test showed a significant difference in all performance measures between easy and difficult tasks (Performance VEE-VED and MiE-MiD), which supported the hypothesis that changes in workload affected performance.

Overall the experimental results showed that perceived presence was not primarily affected by the level of difficulty, which on the other side implied that performance was affected by presence and not vice versa. It could be concluded that presence was the mediating variable between application type and performance (Figure 20).

[19] VE = simulation, Mi = miniature world, E = easy, D = difficult.

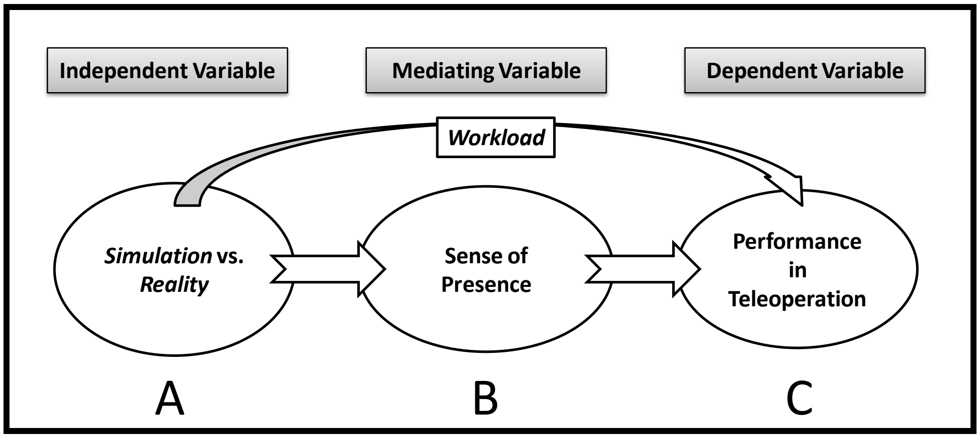


Figure 20: Relationships of the Constructs

6.2.5 Conclusions for the Proposed Relationships

In summary, the four main hypotheses (Figure 20) were supported and showed that

- the application type (simulation or miniature world) affected perceived presence (A→B),
- performance deceased in the more dangerous miniature world (A→C),
- perceived presence was negatively related to performance (B→C), and
- presence affected performance and not vice versa (not C→B).

These findings supported the proposed model of relationship that the application type affected presence and that presence affected performance. Although the change of performance and presence between simulation and miniature world were significant, the relationship between the constructs itself could only be established qualitatively. The prior, sometimes contradictory findings of other studies that the presence-performance relationship was difficult to establish and to quantify were confirmed. Main obstacle in this study was the high individuality of presence and performance. This individuality resulted in the absence of a mathematical correlation between change in presence and changes in performance. Nevertheless, the overall theory that a change in presence, based on application type and perceived sense of danger, induced a change in performance was supported.

6.3 Additional Relationships of Constructs and Measures

6.3.1 Heart Rate and 'igroup Presence Questionnaire'

The previous findings already supported skepticism in the use of HR as a measure for presence in the given scenario, which became obvious in the not established relationship between HR and performance. Based on this, although the analysis resulted in a small positive correlation at a decreased significance level ($\alpha=0.1$), it could not be concluded that the correlation is based on a different assessment of the applications and hence higher agitation in the miniature world. The changes in HR between the application types, contrary to the IPQ-scales that changed significantly, were just too small to draw further conclusions.

6.3.2 Spatial Abilities Test and Performance Measures

Surprisingly, the correlations between the Guilford-Zimmerman tests for spatial orientation and spatial visualization were not significant for all measures: speed and runs were significantly correlated to the spatial ability tests in the simulation, but not the miniature world. The numbers of errors were not correlated to the spatial ability test in all four tasks. Overall the lack of significant findings in the analyses in 6 of the 8 cases would make any discussion highly speculative. Further research would be necessary.

6.3.3 Immersive Tendencies and Presence

A positive correlation between the pre-test ITQ and the two IPQs (VE/Mi) at a decreased significance level (α=0.1) confirmed the assumed relationship and were consistent with findings of Witmer and Singer (1998). The correlation for the more intense miniature world was higher than for the simulation. This could be explained by the relationship that the higher the immersiveness of the environment and the higher immersive tendencies, the more sense of presence increases. A Pearson correlation analysis between the ITQ and change of IPQ-scales (IPQVE-IPQMi[20]) supported this theory (r=0.305, N=37, p=0.033 (1-tailed)).

6.4 Consideration of Additional Relationships

Beside the already discussed factors of the user's cognition (spatial aptitudes and immersive tendencies) other cognitive factors could have affected performance and influenced the experimental results. In the following paragraphs those possible factors will be described in their observed influence on the performance measures during the experiment.

6.4.1 Affects

Affects, according to Tennyson and Breuer (2002), include motivation, feeling, attitudes, emotions, anxiety, and values. Affects were not evaluated by questionnaires or directed observations during the experiment, but assumed to be randomized. Nevertheless anecdotal observations might be useful to derive conclusions:

[20] VE = simulation, Mi = miniature world.

- While most of the participants were highly committed to the experiment, some showed lacking motivation. The latter had problems following instructions and their performance was observably bad. Surprisingly those participant's results did not show up as outliers in the subsequent analyses, their performance-differences between simulation and miniature world followed the hypotheses in the same way as with the motivated participants. It can be concluded that motivation did not appear to affect the hypothesized performance- and presence-changes, only the baseline of performance.
- Some participants showed significant aversion to the task itself, some found the task boring after a short time and others reported dizziness due to the circling movement. This aversion to the task also negatively affected the overall performance, but as with motivation, these affects did not influence the relative change in performance or presence.

The anecdotal observations and the analysis of the individual's results showed that the presence related performance differences are a general observation, despite individual personal affects. It can be concluded that it appears that affects did not influence the performance changes between the two applications.

6.4.2 Influence of Experience and Task Sequencing

Familiarity with the task (simulation or teleoperation) could also have affected performance. Beside the familiarity and experience prior to the tasks, the learning curve during the experiment might also have been a significant factor.

A correlation analysis between reported familiarity (ego-centric games, joystick controls, teleoperation tasks) and performance measures showed no significant relationship. The reported prior experience did not affect the performance in these specific tasks.

The sequence of the application (VE-Mi or Mi-VE) had a ratio of 50:50. Figure 21 - 24 show that the performance-differences were not significantly affected by the sequence. It can be concluded that the sequencing of application did not significantly affected the results of the experiment.

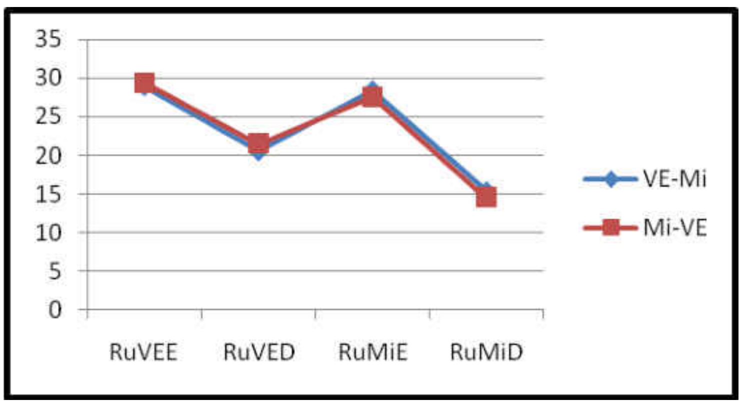


Figure 21: Runs in Relation to Tasks and Sequence

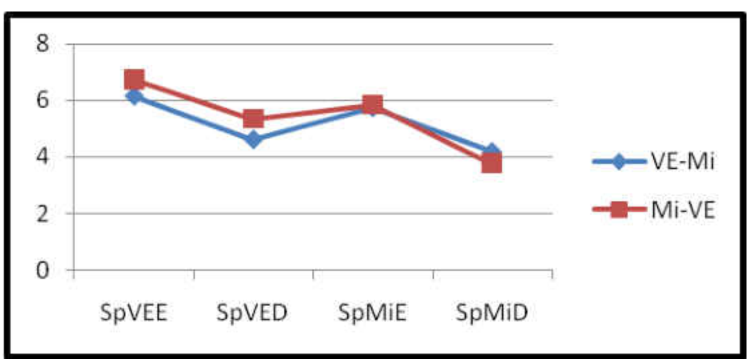


Figure 22: Speed in Relation to Tasks and Sequence

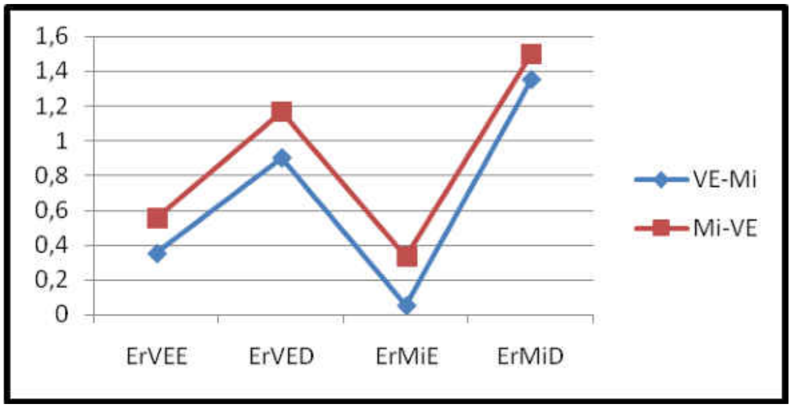


Figure 23: Errors in Relation to Tasks and Sequence

6.5 Limitations and Directions for Future Research

While the basic model of this study, the directed relationship between application type and performance with presence as mediating construct, was supported by the experimental findings, the study had a number of limitations:

First and as suspected after the review of the literature, performance and presence measures were highly task related. Although the main model was confirmed, quantitative results from the experiment (like performance change in percent) cannot be transferred to other tasks.

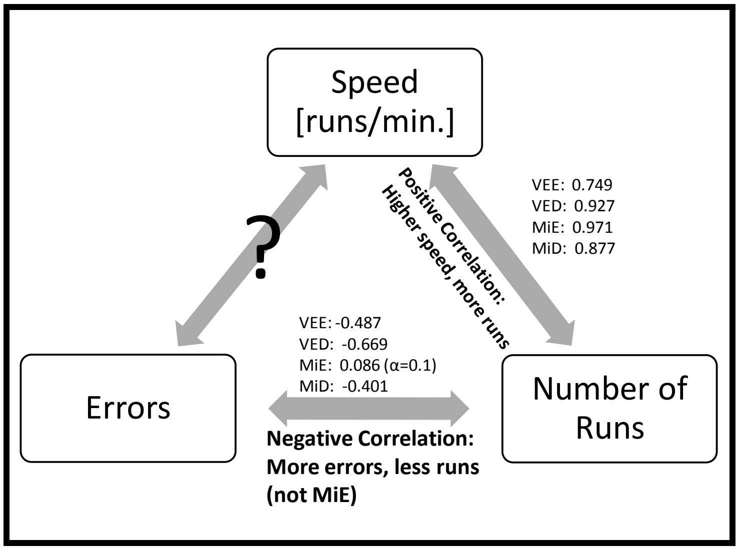


Figure 24: Relationship of Performance Measures

Second, performance measures were related to each other: speed as possible measure for the participants' riskiness was calculated from the number of runs within the operational time. Runs on the other hand depended on the number of errors, since an error was connected with an one minute penalty time. Despite the sometimes inconclusive findings for correlations (Figure 24) especially for speed and errors, speed was still related to the number of errors via runs and hence not a truly independent measure for performance.

Third, the robot's operation had an unexpected impact on the performance measures in difficult tasks: errors occurred due to increased fear, not vice versa as hypothesized. Although the observations suggested that a higher fear resulted in higher errors in the given setting, this theory was not tested and could not be scientifically analyzed.

Based on the study's limitations and observations during the experiment, further research on presence and performance could be:

- Research with truly independent performance measures; e. g. speed measured during all tasks in meter/sec by an optic analysis of the robot's movement in the miniature world.
- Questions of the participants showed that the IPQ items could be misunderstood due to their focus on virtual environments (see Appendix A, IPQ). Tests of different presence questionnaires could result in a more significant relationship between presence and performance and a better understanding of the applicability of those questionnaires to teleoperation.
- Observation during the experiment showed basically two strategies for the tasks: some participants always searched for the cylinders and used them as an orientation point; others oriented along the table edge without bothering about the cylinders at all. If successfully applied, the second strategy produced better results since it did not distinguish between easy and difficult levels. One participant, the top performer with this strategy, did not even perceive the difference between the difficulty levels. It could not be determined what triggered the decision concerning the strategy.
- Average heart rate was not a viable measure for presence in the given scenario. Future research with more valuable robots might induce more 'fear' to the operators and could establish a higher correlation between presence and heart rate.

The overall setting of a simulation and a basically identical miniature world proved to be supportive for research on the application type→presence→performance relationship; it allowed for detailed observation of performance differences while at the same time it introduced a significantly different assessment of the application by the participant.

6.6 Scope of Generalization

The aim of the study was the establishment of the presence-performance relationship based on a changed perception of danger between two applications. Although the study was not able to find quantifiable relationships to 'forecast' performance based on the level or presence, a positive relationship was established: the higher the presence, the more 'real' the behavior in the task. The study found a significant change in performance based on the operator's threat-assessment, which can be generalized to any simulation-based result. Although the quantifiable findings are highly task related, the general model was supported and can be used in the development and evaluation of simulation-based experiments.

6.7 Summary of Findings

This study was not only about the expected performance differences between simulations and real applications. The main focus was on the underlying reason for that performance change: perceived presence. It was hypothesized that, depending on the user's perception of danger to the equipment, performance between a simulation and a 'real' miniature world will change. In addition the sense of presence as a construct to describe the user's disbelief or feeling of being in the simulation/teleoperation was hypothesized to be the mediating variable: a higher presence leads to a more real behavior.

In summary, the hypothesis was supported in the experiment. Performance was significantly lower in the miniature world while presence increased. In conjunction with additional tests it could be concluded that presence affected performance and not vice versa.

It was shown that the sense of presence was related to performance, although these findings were more qualitative and did not result in a numerical relationship between the two due to the individual's differences in presence and performance. Nevertheless the relationship between presence and performance itself was established, which could not be stipulated in the beginning.

6.8 Implications for Theory and Practice

The significant change in performance between the simulation and the miniature world has considerable consequences for simulation-based studies. The transfer of data to the real application has to consider these differences, it cannot be assumed that performance and performance-related factors like workload are identical. The perceived danger to 'only' a miniature robot led to a performance change of 14%; thus it can be expected that this factor would increase in real world applications where a much higher fear of negative outcomes exists.

The aforementioned conclusions for simulation are also valid for simulation-based teleoperation studies. In addition it has been observed that simulation oriented presence questionnaires can raise misunderstandings in their wording. In the described experiment with a real and a virtual environment, this danger was even higher since the participants could confuse the wording of 'virtual environment' with the prior simulation environment, when actually the item concerned the remote environment in the miniature world.

The confirmed relationship between presence and performance does not support the conclusion: so more presence equals better performance. Since this study covered a specific part of presence, fear of negative outcome, it comes in some aspects to the contrary conclusion: an increase in spatial presence for example might have increased performance in both the simulation and the miniature world, but it would not affect the performance change between the two applications. Research of presence is not about optimizing performance in the simulation, instead an increase in presence should focus on the minimization of differences between simulations and real applications. The bottom line of this study is that in the effort to transfer simulation based experimental results fear of negative outcomes as a specific part of presence has to be addressed.

APPENDIX A: IPQ (IGROUP PROJECT CONSORTIUM, 2004)

Table 29: Items IPQ

Number	IPQ item name	shortcut	loading on …	English question	English anchors	Copyright (item source)
1	G1	sense of being there	PRES	In the computer generated world I had a sense of 'being there'	not at all--very much	Slater & Usoh (1994)
2	SP1	sense of VE behind	SP	Somehow I felt that the virtual world surrounded me.	fully disagree--fully agree	IPQ
3	SP2	only pictures	SP	I felt like I was just perceiving pictures.	fully disagree--fully agree	IPQ
4	SP3	no sense of being in virtual space	SP	I did not feel present in the virtual space.	did not feel--felt present	
5	SP4	sense of acting in VE	SP	I had a sense of acting in the virtual space, rather than operating something from outside.	fully disagree--fully agree	IPQ
6	SP5	sense of being present in VE	SP	I felt present in the virtual space.	fully disagree--fully agree	IPQ
7	INV1	awareness of real environment	INV	How aware were you of the real world surrounding while navigating in the virtual world? (I.e. sounds, room temperature, other people, etc.)?	extremely aware-moderately aware-not aware at all	Witmer & Singer (1994)
8	INV2	not aware of real environment	INV	I was not aware of my real environment.	fully disagree--fully agree	IPQ
9	INV3	no attention to real environment	INV	I still paid attention to the real environment.	fully disagree--fully agree	IPQ
10	INV4	attention captivated by VE	INV	I was completely captivated by the virtual world.	fully disagree--fully agree	IPQ

Number	IPQ item name	shortcut	loading on …	English question	English anchors	Copyright (item source)
11	REAL1	VE real (real/not real)	REAL	How real did the virtual world seem to you?	completely real--not real at all	Hendrix (1994)
12	REAL2	experience similar to real environment	REAL	How much did your experience in the virtual environment seem consistent with your real world experience?	not consistent-moderately consistent-very consistent	Witmer & Singer (1994)
13	REAL3	VE real (imagined/real)	REAL	did the virtual world seem to you?	about as real as an imagined world--indistinguishable from the real world	Carlin, Hoffman, & Weghorst (1997)
14	REAL4	VE wirklich	REAL	The virtual world seemed more realistic than the real world.	fully disagree--fully agree	IPQ

PRES = General Presence, SP = Spatial Presence, INV = Involvement, REAL = Experienced Realism.

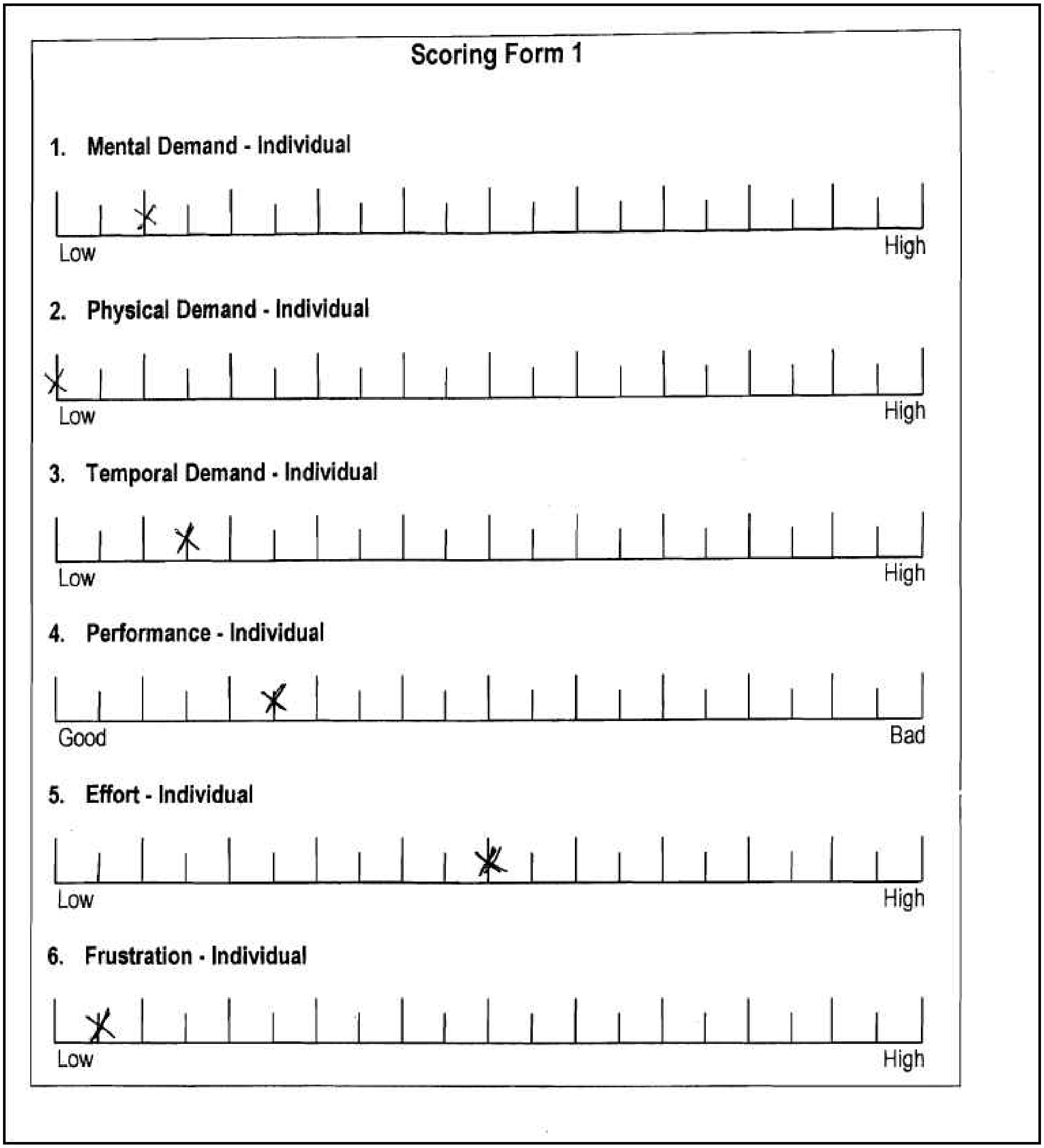
Slater, M., & Usoh, M. (1994). Representations Systems, Perceptual Position, and Presence in Immersive Virtual Environments, Presence, Vol. 2(3), 221-233.

Witmer, B.G., & Singer, M.J. (1994). Measuring Presence in Virtual Environments, ARI Technical Report, Alexandria, VA: U. S. Army Research Institute for the Behavioral and Social Sciences.

Hendrix. C.M. (1994). Exploratory Studies on the Sense of Presence in Virtual Environments as a Function of Visual and Auditory Display Parameters, Master's Thesis, Human Interface Technology Laboratory of the Washington Technology Center at the University of Washington.

Carlin, A.S., Hoffman, H.G., & Weghorst, S. (1997). Virtual reality and tactile augmentation in the treatment of spider phobia: a case report, Behavior Research and Therapy, 35(2), 153-158.

APPENDIX B: NASA TLX


Scoring Form 1

1. Mental Demand - Individual

Low High

2. Physical Demand - Individual

Low High

3. Temporal Demand - Individual

Low High

4. Performance - Individual

Good Bad

5. Effort - Individual

Low High

6. Frustration - Individual

Low High

Figure 25: NASA TLX Scoring Form 1

Scoring Form 2

For each of the pairs (e.g., mental demand vs. effort) choose which one of the two items was more important to *your experience of workload* (Circle).

Circle one of each pair:

Sources of Workload Comparisons

Effort or Performance	Temporal Demand or Frustration
Temporal Demand or Effort	Physical Demand or Frustration
Performance or Frustration	Physical Demand or Temporal Demand
Physical Demand or Performance	Temporal Demand or Mental Demand
Frustration or Effort	Performance or Mental Demand
Performance or Temporal Demand	Mental Demand or Effort
Mental Demand or Physical Demand	Effort or Physical Demand
Frustration or Mental Demand	

Figure 26: NASA TLX Scoring Form 2

APPENDIX C: IMMERSIVE TENDENCIES QUESTIONNAIRE

Please complete the following questions. Any information you provide is voluntary and will be kept strictly confidential. A participant number will be assigned to your responses and in no way will your name be associated with the data. The information you provide will be used only for the purposes of this study. If you have any questions, please ask.

Indicate your preferred answer by marking an "X" in the appropriate box of the seven point scale. Please consider the entire scale when making your responses, as the intermediate levels may apply. For example, if your response is once or twice, the second box from the left should be marked. If your response is many times but not extremely often, then the sixth (or second box from the right) should be marked.

1. Do you easily become deeply involved in movies or tv dramas?

|________|________|________|________|________|________|________|

NEVER OCCASIONALLY OFTEN

2. Do you ever become so involved in a television program or book that people have problems getting your attention?

|________|________|________|________|________|________|________|

NEVER OCCASIONALLY OFTEN

3. How mentally alert do you feel at the present time?

|________|________|________|________|________|________|________|

NOT ALERT MODERATELY FULLY ALERT

4. Do you ever become so involved in a movie that you are not aware of things happening around you?

|________|________|________|________|________|________|________|

NEVER OCCASIONALLY OFTEN

5. How frequently do you find yourself closely identifying with the characters in a story line?

|________|________|________|________|________|________|________|

NEVER OCCASIONALLY OFTEN

6. Do you ever become so involved in a video game that it is as if you are inside the game rather than moving a joystick and watching the screen?

|________|________|________|________|________|________|________|

NEVER OCCASIONALLY OFTEN

7. How physically fit do you feel today?

|________|________|________|________|________|________|________|

NOT FIT MODERATELY FIT EXTREMELY FIT

8. How good are you at blocking out external distractions when you are involved in something?

|________|________|________|________|________|________|________|

NOT VERY GOOD SOMEWHAT GOOD VERY GOOD

9. When watching sports, do you ever become so involved in the game that you react as if you were one of the players?

|________|________|________|________|________|________|________|

NEVER OCCASIONALLY OFTEN

10. Do you ever become so involved in a daydream that you are not aware of things happening around you?

|________|________|________|________|________|________|________|

NEVER OCCASIONALLY OFTEN

11. Do you ever have dreams that are so real that you feel disoriented when you awake?

|________|________|________|________|________|________|________|

NEVER OCCASIONALLY OFTEN

12. When playing sports, do you become so involved in the game that you lose track of time?

|________|________|________|________|________|________|________|

NEVER OCCASIONALLY OFTEN

13. How well do you concentrate on enjoyable activities?

|________|________|________|________|________|________|________|

NOT AT ALL MODERATELY WELL VERY WELL

14. How often do you play arcade or video games? (OFTEN should be taken to mean every day or every two days, on average.)

|________|________|________|________|________|________|________|

NEVER OCCASIONALLY OFTEN

15. Have you ever gotten excited during a chase or fight scene on TV or in the movies?
|_________|_________|_________|_________|_________|_________|_________|
NEVER OCCASIONALLY OFTEN

16. Have you ever gotten scared by something happening on a TV show or in a movie?
|_________|_________|_________|_________|_________|_________|_________|
NEVER OCCASIONALLY OFTEN

17. Have you ever remained apprehensive or fearful long after watching a scary movie?
|_________|_________|_________|_________|_________|_________|_________|
NEVER OCCASIONALLY OFTEN

18. Do you ever become so involved in doing something that you lose all track of time?
|_________|_________|_________|_________|_________|_________|_________|
NEVER OCCASIONALLY OFTEN

APPENDIX D: IRB APPROVAL

University of Central Florida

University of Central Florida Institutional Review Board
Office of Research & Commercialization
12201 Research Parkway, Suite 501
Orlando, Florida 32826-3246
Telephone: 407-823-2901, 407-882-2012 or 407-882-2276
www.research.ucf.edu/compliance/irb.html

Notice of Expedited Review and Approval of Requested Addendum/Modification Changes

From: **UCF Institutional Review Board**
FWA00000351, Exp. 10/8/11, IRB00001138

To: **Andre Huthmann and Florian Jentsch**

Date: **December 03, 2008**

IRB Number: **SBE-08-05811**

Study Title: **Presence-Dependent Performance Differences between Virtual Simulations and Miniature Worlds**

Dear Researcher:

Your requested addendum/modification changes to your study noted above which were submitted to the IRB on 11/29/2008 were approved by **expedited** review on 12/01/2008.

Per federal regulations, 45 CFR 46.110, the expeditable modifications were determined to be minor changes in previously approved research during the period for which approval was authorized.

Use of the approved, stamped consent document(s) is required. The new form supersedes all previous versions, which are now invalid for further use. Only approved investigators (or other approved key study personnel) may solicit consent for research participation. Subjects or their representatives must receive a copy of the consent form(s).

This addendum approval does NOT extend the IRB approval period or replace the Continuing Review form for renewal of the study.

On behalf of Tracy Dietz, Ph.D., IRB Chair, this letter is signed by:

Signature applied by Janice Turchin on 12/03/2008 01:11:56 PM EST

IRB Coordinator

Internal IRB Submission Reference Number: 004440

Figure 27: IRB Approval

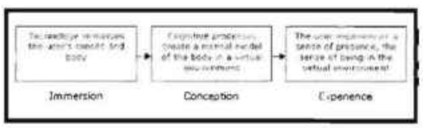
APPENDIX E: COPYRIGHT PERMISSION

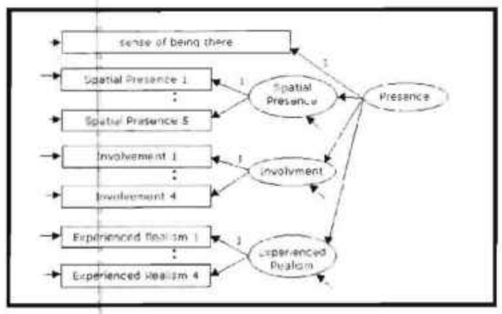
Andre Huthmann
16255 Bristol Lake Circle
Orlando, FL 32828

January 10, 2009

I am completing a doctoral dissertation at the University of Central Florida entitled "PRESENCE-DEPENDENT PERFORMANCE DIFFERENCES BETWEEN VIRTUAL SIMULATIONS AND MINIATURE WORLDS." I would like your permission to reprint in my dissertation excerpts/graphics from the following home-page: http://www.igroup.org/pq/ipq/construction.php.

The excerpts to be reproduced are:





The requested permission extends to any future revisions and editions of my dissertation, including non-exclusive world rights in all languages, and to the publication of my dissertation on demand by UMI. These rights will in no way restrict republication of the material in any other form by you or by others authorized by you. Your signing of this letter will also confirm that you own or your company owns the copyright to the above-described material.

If these arrangements meet with your approval, please sign this letter where indicated below and return it to me. Thank you for your attention in this matter.

Sincerely,

Andre Huthmann

PERMISSION GRANTED FOR THE USE REQUESTED ABOVE:

By: ____________________

Name and signature

Dr. Thomas Schubert

Date: 11.02.2009

REFERENCES

Baker, E. L., & Delacruz, G. C. (2008). A Framework for the Assessment of Learning Games. In H. F. O'Neil, & R. S. Perez (Eds.), *Computer Games and Team and Individual Learning* (pp. 21-37). Amsterdam: Elsevier.

Biggs, S. J., & Sriniwasan, M. A. (2002). Haptic Interfaces. In K. M. Stanney (Ed.), *Handbook of Virtual Environments.* Mahwah, NJ, USA: Lawrence Earlbaum Associates.

Blade, R. A., & Padgett, M. L. (2002). Virtual Environments Standards and Terminology. In K. M. Stanney (Ed.), *Handbook of Virtual Environments* (pp. 15-27). Mahwah, NJ, USA: Lawrence Earlbaum Associates.

Bowman, D. A. (2002). Principles for the Design of Performance-oriented Interaction Techniques. In K. M. Stanney (Ed.), *Handbook of Virtual Environments.* Mahwah, NJ, USA: Lawrence Earlbaum Associates.

Bystrom, K.-E., Barfield, W., & Hendrix, C. (1999). A Conceptual Model of the Sense of Presence on Virtual Environments. *Presence , 8* (2), 241-244.

Carlson, E. B., Waller, N. G., & Putnam, F. W. (1996). Types of Dissociation and Dissociative Types: A Taxometric Analysis for Dissociative Experiences. *Psychological Methods , 1* (3), 300-321.

Chen, J. Y., & Joyner, C. (2006). *Concurrent Performance of Gunner's and Robotic Operator's Tasks on a Simulated Mounted Combat System Environment.* Army Research Laboratory, Aberdeen.

Chen, J. Y., Durlach, P. J., Sloan, J. A., & Bowsen, L. D. (2005). *Robotic Operator Performance in Simulated Reconnaisance Missions.* Aberdeen: Army Research Laboratory.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (Vol. 2nd ed.). Hilsdale: Erlbaum.

Colman, A. M. (2001). *A Dictionary of Psychology.* Oxford: Oxford University Press.

Dictionary of Contemporary English. (1981). Gütersloh, Germany: Langenscheidt-Longman.

Draper, J. V., & Blair, L. M. (1996). Workload, Flow, and Telepresence During Teleoperation. *Proceedings of the 1996 IEEE International Conference on Robotics ans Automation* (pp. 1030-1035). Minneapolis: IEEE.

Draper, J. V., Kaber, D. B., & Usher, J. M. (1998). Telepresence. *Human Factors , 40.3*, 354.

Endsley, M. R. (1995). Measurement of Situation Awareness in Dynamic Systems. *Human Factors , 37* (1), 65-84.

Endsley, M. R. (1995). Toward a Theory of Situation Awareness in Dynamic Systems. *Human Factors , 37* (1), 32-64.

Endsley, M. R., Selcon, S. J., Hardiman, T. D., & Croft, D. G. (1998, October). A Comparative Analysis of SAGAT and SART for Evaluations of Situation Awareness. *Paper presented at the 42nd Annual Meeting of the Human Factors & Ergonomics Society* . Chicago, IL, USA.

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods , 39* (2), 175-191.

Finkelstein, Neal Mark (1999). Charting the retention of tasks learned in synthetic virtual environments. Ph.D. dissertation, University of Central Florida, United States -- Florida. Retrieved April 9, 2008, from Dissertations & Theses @ University of Central Florida-FCLA database. (Publication No. AAT 9950336).

Ford Morie, J., Tortell, R., & Williams, J. (2008). Would Yoy Like to Play a Game? Experience and Expectation in Game-Based Learning Environments. In H. F. O'Neil, & R. S. Perez (Eds.), *Computer Games and Team and Individual Learning* (pp. 269-286). Amsterdam: Elsevier.

Foxlin, E. (2002). Motion Tracking Requirements and Technologies. In K. M. Stanney (Ed.), *Handbook of Virtual Environments.* Mahwah, NJ, USA: Lawrence Earlbaum Associates.

Freeman, J., Avons, S. E., Pearson, D. E., & Ijsselstein, W. A. (1999). Effect of Sensory Information and Prior Experience on Direct Subjective Ratings of Presence. *Presence , 8* (1), 1-13.

Garson, G. D. (2002). *Guide to Writing Empirical Papers, Theses, and Dissertations.* New York: Marcel Dekker, Inc.

Guilford, J. P., & Zimmerman, W. S. (1948). The Guilford-Zimmerman Aptitude Survey. *Journal of Applied Psychology , 32* (1), 24-34.

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. In P. A. Hancock, & N. Meshkati (Eds.), Human Mental Workload (pp. 139-178). New York: Elsevier Science Publishing Company, Inc.

Hettinger, L. J. (2002). Illusory Self-motion in Virual Environments. In K. M. Stanney (Ed.), *Handbook of Virtual Environments* (pp. 471-491). Mahwah, NJ, USA: Lawrence Earlbaum Associates.

igroup. (2004, Sep 08). *IPQ Scale Construction*. Retrieved Apr 27, 2008, from http://www.igroup.org/pq/ipq/construction.php

igroup project consortium. (2004, Oct 19). *igroup presence questionnaire (IPQ) Item Download*. Retrieved March 21, 2008, from www.igroup.org: http://www.igroup.org/pq/ipq/download.php#English

Jang, D. P., Kim, I. Y., Nam, S. W., Wiederhold, B. K., Wiederhold, M. D., & Kim, S. I. (2002). Analysis of Physiological Response to Two Virtual Environments: Driving and Flying Simulation. *CyberPsychology & Behavior , 5* (1), 11-18.

Jentsch, F., & Bowers, C. A. (1998). Evidence for the Validity of PC-Based Simulations in Studying Aircrew Coordination. *The International Journal of Aviation Psychology , 8* (3), 243-260.

Kaber, D. B., Draper, J. V., & Usher, J. M. (2002). Influences of Individual Differences on Application Design for Individual and Collaborative Immersive Virtual Environments. In K. M. Stanney (Ed.), *Handbook of Virtual Environments* (pp. 379-402). Mahwah, NJ, USA: Lawrence Earlbaum Associates.

Kamsickas, G. M. (2003). FCS Unmanned Combat Demonstration - Soldier Task Loading Results. Orlando, FL, USA. Retrieved Apr 27, 2008, from proceedings.ndia.org/3570/session7/Kamsickas.ppt

Kennedy, R. S., Lane, N. E., Berbaum, K. S., & Lilienthal, M. G. (1993). Simulator sickness questionnaire: an enhanced method for quantifying simulatior sickness. *International Journal of Aviation Psychology , 3* (3), 203-220.

Kihlstrom, J. F. (2006, April 24). *Tellegen Absorption Scale*. (University of Berkeley) Retrieved April 1, 2008, from http://socrates.berkeley.edu/~kihlstrm/TAS.htm

Hale, Kelly Sue Kingdon (2006). Enhancing situational awareness through haptics interaction in virtual environment training systems. Ph.D. dissertation, University of Central Florida, United States -- Florida. Retrieved April 9, 2008, from Dissertations & Theses @ University of Central Florida-FCLA database. (Publication No. AAT 3242439).

Lathan, C. E., & Tracey, M. (2002). The Effects of Operator Spatial Perception and Sensory Feedback on Human-Robot Teleoperation Performance. *Presence , 11* (4), 368-377.

Lawson, B. D., Sides, S. A., & Hickinbotham, K. A. (2002). User Requirements for Perceiving Body Acceleration. In K. M. Stanney (Ed.), *Handbook of Virtual Environments.* Mahwah, NJ, USA: Lawrence Earlbaum Associates.

Lee, A. Y., & Perez, A. M. (2008). The Effects of Changing Resources on Game Performance, Subjective Workload, and Strategies. In H. F. O'Neil, & R. S. Perez (Eds.), *Computer Games and Team and Individual Learning* (pp. 169-184). Amsterdam: Elsevier.

Lee, K. M. (2004). Presence, Explicated. *Communication Theory* , 27-50.

May, J. G., & Badcock, D. R. (2002). Vision and Virtual Environments. In K. M. Stanney (Ed.), *Handbook of Virtual Environments.* Mahwah, NJ, USA: Lawrence Earlbaum Associates.

Meehan, Michael John (2001). Physiological reaction as an objective measure of presence in virtual environments. Ph.D. dissertation, The University of North Carolina at Chapel Hill, United States -- North Carolina. Retrieved April 9, 2008, from Dissertations & Theses: Full Text database. (Publication No. AAT 3007847).

Muckler, F. A., & Seven, S. A. (1992). Selecting Performance Measures: "Objective" versus "Subjective" Measurement. *Human Factors , 34* (4), 441-445.

Munro, A., Breaux, R., Patrey, J., & Sheldon, B. (2002). Congitive Aspects of Virtual Environments. In K. M. Stanney (Ed.), *Handbook of Virtual Environments* (pp. 415-434). Mahwah, NJ, USA: Lawrence Earlbaum Associates.

Nash, E. B., Edwards, G. W., Thompson, J. A., & Barfield, W. (2000). A Review of Presence and Performanca in Virtual Environments. *International Journal of Human-Computer Interaction , 12* (1), 1-41.

Nelson, W. T., & Bolia, R. S. (2002). Technological Considerations in the Design of Multisensory Virtual Environments: The Virtual Field of Dreams Will Have to Wait. In K. M. Stanney (Ed.), *Handbook of Virtual Environments.* Mahwah, NJ, USA: Lawrence Earlbaum Associates.

Park, Sung Ha (1998). The effects of display format and visual enhancement cues on performance of three-dimensional teleoperational tasks. Ph.D. dissertation, Texas Tech University, United States -- Texas. Retrieved April 9, 2008, from Dissertations & Theses: Full Text database. (Publication No. AAT 9826482).

Peters, M. (2005). Sex differences and the factor of time in solving Vandenberd and Kuse mental rotation problems. *Brain and Cognition , 57* (2), 176-784.

Regenbrecht, H. T., Schubert, T. W., & Friedmann, F. (1998). Measuring the Sense of Presence and its Relations to Fear of Heights in Virtual Environments. *International Journal of Human-Computer Interaction , 10* (3), 233-249.

Regenbrecht, H., & Schubert, T. (2002). Real and illusory interaction enhance presence in virtual environments. *Presence , 11* (4), 425-434.

Rehfeld, Sherri Ann (2006). The impact of mental transformation training across levels of automation on spatial awareness in human-robot interaction. Ph.D. dissertation, University of Central Florida, United States -- Florida. Retrieved April 9, 2008, from Dissertations & Theses @ University of Central Florida-FCLA database. (Publication No. AAT 3242463).

Reid, G. B., & Nygren, T. E. (1988). The Subjective Workload Assessment Technique: A Scaling Procedure for Measuring Mental Workload. In P. A. Hancock, & N. Meshkati (Eds.), *Human Mental Workload* (pp. 185-218). New York: Elsevier Science Publishers.

Riley, Jennifer Magee (2001). The utility of measures of attention and situation awareness for quantifying telepresence. Ph.D. dissertation, Mississippi State University, United States -- Mississippi. Retrieved April 9, 2009, from Dissertations & Theses: Full Text database. (Publication No. AAT 3027468).

Rose, F. D., Attree, E. A., Brooks, B. M., Parslow, D. M., Penn, P. R., & Ambihaipahan, N. (2000). Training in Virtual Environments: Transfer to Real World Tasks in Equivalence to Real Task Training. *Ergonomics , 43* (4), 494-511.

Rubio, S., Diaz, E., Martin, J., & Puente, J. M. (2004). Evaluation of Subjective Mental Workload: A Comparison of SWAT, NASA-TLX, and Workload Profile Methods. *Applied Psychology , 53* (1), 61-86.

Sadowsky, W., & Stanney, K. (2002). Presence in Virtual Environments. In K. M. Stanney (Ed.), *Handbook of Virtual Environments* (pp. 791-806). Mahwah, NJ, USA: Lawrence Earlbaum Associates.

Schubert, T. W. (2003). A three-component scale measuring spatial presence, involvement, and realness. *Zeitschrift für Medienpsychologie , 15* (2), 69-71.

Schubert, T., Friedmann, F., & Regenbrecht, H. (2001). The epcerience of presence: Factor analytic insights. *Presence , 10*, 266-281.

Shaw, R. L., Brickman, B. J., & Hettinger, L. J. (2004). The Global Implicit Measure (GIM): Concept and Experience. In D. A. von Vincenzi, M. Mouloua, & P. A. Hancock (Ed.), *Human Performance, Situation Awareness and Automation Conference (HPSAA II)* (pp. 97-102). Mahwah, NJ: Lawrence Earlbaum Ass.

Sherman, W. R., & Craig, A. B. (2003). *Understanding Virtual Reality.* San Francisco, CA: Morgan Kaufmann.

Shilling, R. D., & Shinn-Cunningham, B. (2002). Virtual Auditory Displays. In K. M. Stanney (Ed.), *Handbook of Virtual Environments* (pp. 65-92). Mahwah, NJ, USA: Lawrence Earlbaum Associates.

Sigall, H., & Mills, J. (1998). Measures of Independent Variables and Mediators are Useful in Social Psychology Experiments: But Are They Necessary? *Personality and Social Psychology Review , 2* (3), 218-226.

Slater, M. (1999). Measuring Presence: A Response to the Witmer and Singer Presence Questionaire. *Presence , 8* (5), 560-565.

Slater, M., & Wilbur, S. (1997). A framework for immersive environments (FIVE). *Presence:Teleoperators & Virtual Environments , 6* (6), 603-617.

Slater, M., Guger, C., Edlinger, G., Leeb, R., Pfurtscheller, G., Antley, A., et al. (2006). Analysis of Physological Responses to a Social Situation in an Immersive Virtual Environment. *Presence , 15* (5), 553-569.

Slater, M., Steed, A., McCarthy, J., & Maringelli, F. (1998). The Influence of Body Movement on Subjective Presence in Virtual Envornments. *Human Factors , 40* (3).

Sloan, J. A. (2005). The Effects of Video Frame Delay and Spatial Ability on the Operation of Multiple Semiautonomous ans Tele-Operated Robots. Unpublished Thesis, Orlando, FL, USA. Retrieved April 26, 2008, from http://purl.fcla.edu.ezproxy.lib.ucf.edu/fcla/etd/CFE0000430

Snow, Michael Potter (1996). Charting presence in virtual environments and its effects on performance. Ph.D. dissertation, Virginia Polytechnic Institute and State University, United States -- Virginia. Retrieved April 9, 2008, from Dissertations & Theses: Full Text database. (Publication No. AAT 9724047).

Stanney, K. (1998). Aftereffects and Sence of Presence in Vitual Environments: Formulation of a Research and Development Agenda. *International Journal for Human-Computer Interaction* (10(2)), 135-187.

Stanney, K. M., & Zyda, M. (2002). Virtual Environments in the 21st Century. In K. M. Stanney (Ed.), *Handbook of Virtual Environments* (pp. 1-14). Mahwah, NJ, USA: Lawrence Earlbaum Associates.

Storms, R. L. (2002). Auditory-Visual Cross-Modality Interaction and Illusions. In K. M. Stanney (Ed.), *Handbook of Virtual Environments* (pp. 455-469). Mahwah, NJ, USA: Lawrence Earlbaum Associates.

synesthesia. (n.d.). Retrieved April 26, 2008, from Miriam Websters Online Dictionary: http://www.merriam-webster.com/dictionary/

Tellegen, A., & Atkinson, G. (1974). Openess to Absorbing and Self-Altering Experiences ("Absorption"), a Trait Related to Hypnotic Susceptibility. *Journal of Abnormal Psychology , 83* (3), 268-277.

Tennyson, R. D., & Breuer, K. (2002). Improve Problem Solving and Creativity Through Use of Complex-Dymanic Simulations. *Computers in Human Behavior , 18*, 650-668.

Tsang, P. S., & Velazquez, V. L. (1996). Diagnosticity and multidimensional subjective workload ratings. *Ergonomics , 39* (3), 358-381.

Turk, M. (2002). Gesture Recognition. In K. M. Stanney (Ed.), *Handbook of Virtual Environments.* Mahwah, NJ, USA: Lawrence Earlbaum Associates.

Usoh, M., Catena, E., Arman, S., & Slater, M. (2000). Using Presence Questionnaires. *Presence , 9* (5), 497-503.

van Baren, J., & Jsselsteijn, W. (2005). *Compendium of Presence Measures*. Retrieved 03 03, 2008, from http://www.presence-research.org/Overview.html

Welch, R. B. (1999). How Can We Determine if the Sense of Presence Affects Task Performance? *Presence , 8* (5), 574-577.

Wickens, C. D. (2002). Multiple resources and performance prediction. *Theoretical Issues in Ergonimics Science , 3* (2), 159-177.

Wiederhold, B. K., Jang, D. P., Kaneda, M., Cabral, I., Lurie, Y., May, T., et al. (2001). An Investigation into Physiological Responeses in Virtual Environments: An Objective Measurement of Presence. In G. Riva, & C. Galimberti (Eds.), *Towards Cyberpsychology* (pp. 175-183). Washington D.C.: IOS Press.

Wikipedia. (2008, May 05). *Confirmatory Factor Analysis*. Retrieved March 11, 2008, from Wikipedia: http://en.wikipedia.org/wiki/Confirmatory_factor_analysis

Wikipedia. (2008, March 11). *convergent validity*. Retrieved April 26, 2008, from Wikipedia: http://en.wikipedia.org/wiki/Convergent_validity

Williams, Lorraine E. P. (2001). Bisensory force feedback in telerobotics. Ph.D. dissertation, University of Houston, United States -- Texas. Retrieved April 9, 2008, from Dissertations & Theses: Full Text database. (Publication No. AAT 3032041).

Witmer, B. G., & Singer, M. J. (1994). *Measuring Presence in Virtual Environments.* Alexandria: US Army Research Institute.

Witmer, B. G., Sadowsky, W. J., & Finkelstein, N. M. (2000). Training Dismounted Soldiers in Virtual Environments: Enhancing Configuration Learning. Orlando: US Army Research Institute.

Witmer, B., & Singer, M. J. (1998). Measuring Presence in Virtual Environments: A Presence Questionnaire. *Presence , 7* (3), 225-240.

Woodman, Michael D. (2006). Cognitive training transfer using a personal computer-based game: A close quarters battle case study. Ph.D. dissertation, University of Central Florida, United States -- Florida. Retrieved April 9, 2008, from Dissertations & Theses @ University of Central Florida-FCLA database. (Publication No. AAT 3210389).

Zimmons, Paul Michael (2004). The influence of lighting quality on presence and task performance in virtual environments. Ph.D. dissertation, The University of North Carolina at Chapel Hill, United States -- North Carolina. Retrieved April 9, 2008, from Dissertations & Theses: Full Text database. (Publication No. AAT 3140421).