HOCHSCHULE FURTWANGEN UNIVERSITY HFU

RESEARCH REPORT

HUFACIN_VR

A pilot study and methodologically approach to

Headset-related stress in VR

TEAM

Project Supervisor:
Prof. Nikolaus Hottong

Faculty of Digital Media

Master-Students of the courses: Design of Interactive Media and Media Informatics:

Denise Marcella Alka
Luis Keck
Markus Schürmann
Jeanna Nechypurenko
Leonie Kahn
Icaro Orengo

SUBMITTED ON: 29.07.2022
REVISED: 30.08.2022

ABSTRACT

While we have seen an increase in the production and consumption of Virtual Reality applications in recent years, there seems to have been no consistent and methodologically solid approach to evaluating the impact of *VR technologies* on users. Our research focuses on highlighting the possible relationships between the occurrence of stressors and the use of VR headsets. Therefore, it was decided to methodologically address the issue by developing a general taxonomy for possible stressors in VR and firstly carrying out a series of laboratory tests isolating only the baseline of headset-related stress.

Twelve young adults were tested in a pilot study by a specific five minute VR item for qualitative evaluations of the content they consumed, their experience with the equipment, and their general experience running through our study. Same test were carried out with a comparison group in a boxed TV-environment. For both groups biomedical parameters were derived that indicate the stress level of the subjects. ECG, EDA and Cortisol test were used for these measurements. Additionally the subjects were screened by questionnaires regarding changes in comfort.

The data show experimental validity and robustness. Although knowing that if the samples are too small, the power of our tests may not be sufficient to detect an precise difference between the means, we decided to perform an experimental statistical analysis of the pilot study subject data.

According to the results of this analysis and within its limitations, *young adults are not likely to experience stress from using a VR headset for a more short-term use where all other VR stressors have been eliminated.* Ten hypotheses in this regard have been rejected.

However, the methods used to record stress or relaxation in VR seem generally suitable for further investigations verifying the results in a future larger quantitative study or to identify other stress factors caused by VR-content, through equipment mismatch or by conflicts between user's sensoric reality and the virtual world.

For these reasons, we encourage further study and expansion of the research at a later stage.

The pilot test was conducted by a student team of the Master courses *Design of Interactive Media* and *Media Informatics* at Furtwangen University (HFU).

The research project "*Human Factors in Virtual Reality (HuFacIn_VR)*" was initiated in 2021 by Prof. Nikolaus Hottong (Faculty of Digital Media) and Prof. Dr. Verena Wagner-Hartl (Faculty of Industrial Technologies).

CONTENTS

CONTENTS

LIST OF FIGURES

LIST OF FIGURES

LIST OF TABLES

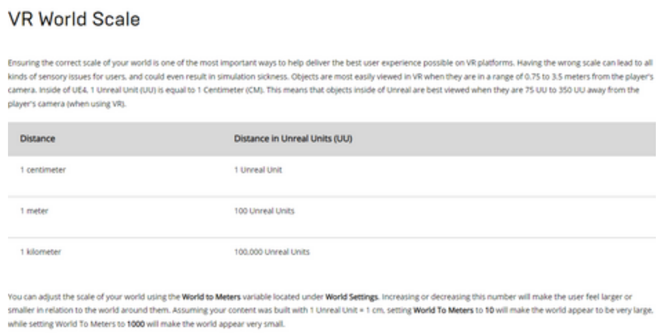
1 INTRODUCTION: STRESS IN VR

The increasing digitalization of our society has given the living and working world immense opportunities, but also created numerous new questions that will have to be scientifically processed in the coming years. For example, new lines of business and closer networking of different people and institutions worldwide emerged. In this context, the industry, including large companies such as Google, Apple and mobile companies, has increasingly embraced newer technologies such as virtual reality (VR). Immersing yourself in the computergenerated reality increases the immersion and the feeling via large screens in special rooms (Cave Automatic Virtual Environment, also CAVE) or by wearing a head-mounted display such as video or VR glasses as well as VR-Headsets to be more present in a created reality.

With the help of this technology, a multitude of new application possibilities is offered in sectors such as the media, medical technology, education and the automotive industry. The reason why their use is still very hesitant, especially in education, is related to various factors. One reason relates to the equipment or technical devices. In addition, the previous handling and work with VR applications as well as the development and use is still immature. Above all, fundamental research questions remain insufficiently answered, such as possible physiological and psychological borderline situations or research on the objective measurement of stress in VR systems.

The information about a stress-free or unproblematic stay in simulation or VR systems differs. There is mostly no scientific evidence of the informations mentioned. For example in the scene comfort documentation of the Unreal Engine, one produced by Epic Games for the development of console and computer games, Game engine, clarified on their web pages.

> “Objects are most easily viewed in VR when they are in a range of 0.75 to 3.5 meters from the player's camera. Inside of UE4, 1 Unreal Unit (UU) is equal to 1 Centimeter (CM). This means that objects inside of Unreal are best viewed when they are 75 UU to 350 UU away from the player's camera (when using VR).”
>
> (Virtual Reality Best Practices | Unreal Engine Documentation, n.d.)



VR World Scale

Ensuring the correct scale of your world is one of the most important ways to help deliver the best user experience possible on VR platforms. Having the wrong scale can lead to all kinds of sensory issues for users, and could even result in simulation sickness. Objects are most easily viewed in VR when they are in a range of 0.75 to 3.5 meters from the player's camera. Inside of UE4, 1 Unreal Unit (UU) is equal to 1 Centimeter (CM). This means that objects inside of Unreal are best viewed when they are 75 UU to 350 UU away from the player's camera (when using VR).

Distance	Distance in Unreal Units (UU)
1 centimeter	1 Unreal Unit
1 meter	100 Unreal Units
1 kilometer	100,000 Unreal Units

You can adjust the scale of your world using the **World to Meters** variable located under **World Settings**. Increasing or decreasing this number will make the user feel larger or smaller in relation to the world around them. Assuming your content was built with 1 Unreal Unit = 1 cm, setting **World To Meters** to **10** will make the world appear to be very large, while setting World To Meters to **1000** will make the world appear very small.

Figure 2: Unreal Engine VR World Scale recommendations (*Virtual Reality Best Practices | Unreal Engine Documentation*, n.d.)

1 INTRODUCTION: STRESS IN VR

It is stated there that objects outside of a distance range of 0.75 - 3.5 meters cause sensory problems, but the scientific source itself remains unnamed. Maybe the near point problem is caused by the vergence / accommodation conflict but this statement is not linked to any scientific paper. However, this was the fundamental starting point of our investigations towards stress in VR which could be further expanded to improve the existing VR apparatus, VR tools or best-practice UX design of VR content. Among other impacts as nausea, motion sickness and eye strain.

But at the start there is one big question: how can stress generally be measured in a system where you leave reality whith a specific equipment attached to your body?

Figure 3: Minh Pham (2020)

In the course of processing the scientific basis in the field of VR, a team of the Faculty of Digital Media at Furtwangen University is currently researching manageable methods to identify stress in virtual reality.

The results should be publicly accessible and contribute to improving VR-UX design, VR headset technology or other VR equipment.

1 INTRODUCTION: STRESS IN VR

This project was based on an initial investigation of objective measurement options for emotions and stress in VR, which was carried out at the HFU Campus in Tuttlingen by Prof. Dr. Wagner-Hartl and Master-students of the course Human Factors (HF) since summer semester of 2021. This experiment provided first approaches and ideas, to identify the basic stress through the use of a VR headset.

In order to be able to generally determine stress in VR and to recognize the potential effects, a more thorough investigation of the fundamental stress phenomenon is required, based on limitations and exclusion methods in order to obtain a more accurate knowledge of stress in virtual reality. Since winter semester 2021/22, the basic stress of wearing a VR Headset has been the focus of the research project by Prof. Hottong and master-students in the courses Design of Interactive Media and Media Informatics at Furtwangen University.

This project started in cooperation with Industrial Technologies/Tuttlingen and was completely taken over in summer semester 2022. Due to the above-mentioned prerequisites, we focussed on the improvement of basic stress by wearing a VR headset, which has to be clearly differentiated from other stress triggers as we explained in the following chapters.

Maybe the results of our pilot study could be a relevant contribution, particularly with regard to other investigations on stress or emotional arousal in VR.

2 DEFINITIONS & RELATED WORK

2.1 DEFINITIONS

First, there are several preconditions before being able to conduct such a study. One of them is definitions, which need to be clarified. It should be clear, what is meant by VR or stress in general or what emotional arousal is, as those will be very important core aspects in further investigations.

2.1.1 VR

VR is the shortcut for virtual reality in this context. It is a medium that perceives the position and actions of the user (head movement, hands, etc.). Stimuli must be natural so that the participant can accept them as a substitute for real stimuli and feels to be present in the simulation (cf. Sherman, 2019).

Virtual reality can be used in combination with VR glasses as well as VR-Headsets, via large screens in special rooms (Cave Automatic Virtual Environment, also CAVE) or by wearing a head-mounted display such as video.

2.1.2 Stress

According to Seyle (Seyle, 1981) stress is declared as a "non-specific response of the organism to any demand" [ibid., p. 170] while Schandry (Schandry, 2016) describes it as a result of one process that consists of three components.

Those are:

1. “An interaction of the individual with his stimulus environment, which is characterized by the character of demands” [ibid., p.316];

2. “One or more attempts to cope with this situation” [ibid., p.316];

3. "A deviation from the optimal balance of physical and psychological functions as a result of this effort by the individual ("stress" in the narrower sense)" [ibid., p.316]

All three components apply to the use of a VR-Headset as a user is exposed to different stimuli while interacting with the equipement. The user tries to compensate those stimuli for example proceeding from the content itself or the equipment by adapting his eyes or tightening the muscles (cf. Mon-Williams & Rushton, 1993).

Also, stress is a real or perceived threat to a person's physiological and psychological integrity that results in physiological and/or behavioral responses (cf. Collet et al., 2021).

Generally it can be measured by the size of the pupil, or other measurements such as the heart rate, cortisol etc..

Our research is addressing the basic stress caused by the apparatus (such as VR Headsets, controllers, etc..) and not by VR content which is mostly called mental stress (see chapter 2.2).

2.1.3 Delimitation to emotional arousal
Stress and arousal both activate the sympathetic nervous system, so they can feel very similar physically (high heart rates, increased muscle tension and less coordination. etc.). Generally stress is subjectively perceived as negative and arousal as positive.

Arousal is a physiological and psychological activity. Stokes and Kite (2001) suggested that arousal is considered the basic energetic state of an organism and „A hypothetical construct that represents the level of central nervous system activity along a behavioral continuum ranging from sleep to alertness.“ (Stokes & Kite, 2001).

Stress on the other side is a state of mental or emotional strain or tension resulting from adverse or demanding circumstances. It is a normal phenomenon that is characterized by a feeling of emotional or physical tension (cf. Tran et al., 2021).

According to the circumplex model of affect (Figure 4), affective states are dependent on each other (cf. Russell, 1980). The model defines the emotion in arousal and valence dimensions. In this case arousal, it refers to the person’s agitation level, and valence refers to how pleasant or unpleasant the feeling is. Psychological research suggests, that these two dimensions are correlated with each other (cf. Lewis et al., 2007). Similarly, stress can be interpreted both positively and negatively depending on its origin. It can occur from various events.

For example, getting a new job with high demand can keep a person excited while in a stressed state, while the lack of desired success in any activity can trigger negative stress. High arousal and negative valence are characteristics of emotional stress (cf. Christianson, 1992). Additionally any stress or negative motivational state will shut down our playfulness (cf. Pfaff, 2006).

Adjectives such as “stressed”, “nervous”, and “tense”, are commonly used to describe emotional stress, while the opposites are “relaxed”, “calm”, and “at ease” (cf. Barrett & Russell, 1999) . Anger, disgust, and fear are specific emotions associated with acute affective states due to their high arousal and negative valence (cf. Wichary et al., 2016).

Our research is addressing the negatively denoted stress and not emotional arousal.

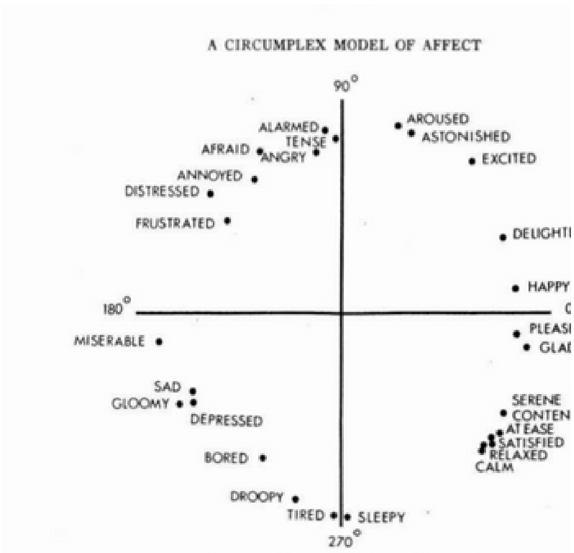


Figure 4: Direct circular scaling coordinates for 28 affect words (based on Russel’s “Circumplex Model of Affect”, 1980)

2.1.4 Relaxation

The term relaxation refers to the physio-psychological process that leads from a state of agitation to mental, physical and emotional calm. The length of stay in this state before its variation is influenced by numerous psychological, social and environmental factors.

The feelings associated with this particular state are well-being, serenity and inner peace. Since this is the result of a change in state, its effect is particularly noticeable after a previous strain. The state of relaxation is often confused with that resulting from intense physical exertion or arousal. In fact, as a result of these manifestations particular hormones are released that simulate the relaxed state, but nevertheless must be associated with fatigue and exhaustion.

The main discriminant between these two states is that in the relaxed state alertness, attention, and reaction performance are not significantly impaired (Schandry, 2016; D'Amelio, 2009).

According to our goal to eliminate content-based stressors in our tests, we inevitably have to create a VR-scene which consistently fulfills all audiovisual and motion features of relaxation (see chapter 5.6).

2.1.5 Cognitive Load

Cognitive Load is the mental effort required to perform a task or learn something new. The theory states, that the successful completion of a task depends on a complex interplay between the sensory input, the long-term memory (which serves as a repository for previously acquired knowledge and skills) and the working memory (shortterm memory).

Working memory is the intermediate stage between sensory and long-term memory, it integrates newly learned information into long-term memory. Sensory and long-term memory have flexible capacities, working memory is limited. The attention directs the working memory function to relevant sensory information and stored knowledge. It directs the learning process and increases (or decreases) the efficiency of working memory. The mix of these attentional, sensory and memory processes determines a person's cognitive load at any given moment (cf. Siegel et al., 2021).

There are three types of cognitive load: Intrinsic cognitive load arises due to the complexity of the content of the task. Extraneous cognitive load is caused by the externals of the learning material, the learning environment and the general conditions while learning. It puts unnecessary strain on working memory and can be avoided. Intrinsic and extraneous co-gnitive load are additive. This leads to an overload of the working memory when both of them are high (cf. Zheng, 2017).

The learning process itself and the effort to understand the learning material, leads to germane or effective cognitive load. It is influenced by the investigator. Unlike extraneous cognitive load, which impairs learning, germane cognitive load has a positive effect on learning. Through higher motivation and an increase of effort the resources used for a task can also be increased (Renkl et al., 2016).

We roughly observed the cognitive load of all subjects running through our VR tests using a built-in method by the VR manufacturer (see chapter 7.1.4)

2.2 RELATED WORK

While even in common sense it is known and everyone has had empirical experience that each individual perceives situations differently and reacts differently to what can be called stress, it is also important to note scientifically that some individuals are more vulnerable to the effects of stress than others. While our bodies can handle acute exposure to stressors, it is equally true, that chronic exposure can lead to psychological disorders, such as anxiety and depression, and even physical changes, such as cardiovascular changes associated with hypertension or stroke (cf. Cohen et al., 2016).

Despite considerable efforts to find an unambiguous method for identifying people vulnerable to developing stress-related disorders, the methodological problem on quantitative data analysis turns out to be central. In fact, most research relies on self-assessment and subjective clinical rankings or exposes subjects to unnatural environments. Through the use of wearable devices and other noninvasive sensing technologies, it has been possible to greatly expand both the range of experimental subjects and the range of survey methodologies. This is how behavioral researchers led by Carmen Sandi of EPFL's School of Life Sciences developed a virtual reality (VR) method that measures a person's vulnerability to psychogenic stressors. They were able to create an approach that captures information about an experimental subject's movement while exploring two virtual environments to predict heart rate variability when exposed to threatening or highly stressful situations. Heart rate variability has been shown in the field to be a strong predictor of vulnerability to physiological stress as well as the development of psychopathology and cardiovascular disorders (cf. Rodrigues et al., 2020).

Other foundational research for methodological and deontological construction came to our support, providing us with a remarkable and well-detailed series of research aimed at analyzing the rise or decline of stressors in experimental subjects as a result of the use of VR technology. (cf. Annerstedt et al., 2013) However, this research focused more on the content used to create the virtual environments rather than on the equipment that enables its use.

2.2.1 Conflicts virtuality vs. reality

In order to enhance the VR experience and reduce psychophysical stressors, an interesting approach was presented at the "IEEE Symposium on 3D User Interfaces" (cf. Wang et al., 2019) in Orlando that offers the user a meaningful interpretive metaphor for entering a new world. This is based on the interdisciplinary concept of imitation to recreate a true physical translation from one place to another. Otherwise VR users would be catapulted from one world to another in a brutal manner. A 5-step conceptual model of transition is presented that describes an Implementation and possible evaluation procedures.

This phenomenon was of great inspiration to be able conceptually and practically to precipitate the experimental subjects of our research from a physical world to the real world and then back again trying to limit as much as possible the transition process between the two worlds described above and sometimes known as Translocation-Effect.

2.2.2 Content-related stress

A great deal of interest from the perspective of research and scientific publications has been reserved for the topic of the effects of content enjoyed in VR in experimental subjects.

As highlighted by Annerstedt (cf. Annerstedt et al., 2013) and his team, a relationship can be found between naturethemed virtual environments, nature sounds and stress management. The pilot study's conclusion that stress recovery occurs more efficiently in natural settings provided the foundation of thought for the creation of our three-dimensional test scene. In fact, even though Wang and his team's study (cf. Wang et al., 2019) only subjected the experimental subjects to 3D and therefore non-stereo-scopic videos, the results highlighted con-firmed and supported our thesis that em-ploying and immersing the experimental subjects in comfortable and non-bored na-tural environments might have been the best choice to eliminate all those disturbing fac-tors that somehow could have affected the goal of neutrality of the VR scene that we set out to achieve.

Following other interpretive schemes and objectives, Hofmann's research and his teams (cf. Hofmann et al., 2020) moved to investigate the link between emotional arousal and parieto-occipital alpha power under naturalistic stimulation. Thus 37 healthy young adults completed an im-mersive VR experience, which included roller coaster rides, while their EEG was recorded. Of particular interest is the investigation methodology that allowed the research team, through the use of linear regression techniques peculiar to Machine Learning, and the use of EEG combined with VR tech-nology, to expand the areas of in-vestigation and methods of noninvasive analysis.

Despite the excellent results presented in this Paper our research team opted not to use EEG technology as it presented too high risks of generating a source of stress due to the use of the headset compared to the possible investigative benefits. In fact, the comfort of the Headset was a very important point of attention in reducing uncontrollable external variables as much as possible.

In defining equipment suitable for the analysis and retrieval of quantitative data of biomedical parameters in order to identify stressors, we focused on the analysis and possible integration of sensors built into the HP headset. These by calculating and comparing different parameters, such as heart rate and variability, pupil dilation, speed, amount of puti fixed and facial cameras, deliver an algorithmic analysis that should represent what HP defines as Cognitive Load (cf. Siegel et al., 2021). As specified the cognitive load recorded in the VR measurement is based on multitasking focused on performing tasks in VR environments. The results were analyzed through the use of a Nasa Task Load Index (TLX) and from measurement devices mentioned above. Precisely because Cognitive Load describes the cognitive/ mental load required to perform a task, learn a new activity or something new, our research team decided to use this technology in order to make comparisons and thereby be able to identify matches of particular statistical value.

When setting out to investigate phenomena within virtual ecosystems, it is standard practice to be aware of and try to contain as much as possible all those phenomena known as cybersickness. These are configured in the research design as Confounding Variables (Pourhoseingholi et al., 2012) that if not properly controlled could undermine the methodology of the experiment.

Cybersickness is a type of motion sickness that can occur when using simulators or spending time in virtual reality. Effects on those affected include nausea, dizziness or headaches. The cause is believed to be the same as motion sickness, which can occur when traveling by car or ship, among other things. When using VR headsets, for example, there is a sensory conflict between what the user sees and what the other senses, such as the sense of balance, perceive. This means for example, that users are moving in a virtual world, while the real body is sitting in a chair.

Several factors are believed to be associated with cybersickness. (cf. Kemeny et al., 2020) Although the creation of the scenes was done with great care and based on parameters, that have already been studied (cf. Davis et al., 2014), and in order to reduce the occurrence of such effects, the design of the experiment itself underwent two different internal reviews for evaluating possible confounding effects on biophysical parameters. As pointed out in previous research (cf. Dennison et al., 2016), the undesirable effects could increase as the discrepancy between the movements in the digital environment and the immobility of the subjects or with the presence of unnatural movements in the 2D video increases. For this reason, a number of precautions have been taken to reduce unwanted effects to a minimum.

Another key parameter we took into consideration is the need to limit the "Uncanny Valley" effect (cf. Seyama & Nagayama, 2007; Choe et al. 2020) by excluding all scenes with overly realistic humanoids. Additionally, authoritative literature produced to date, shows that natural environments have a greater relaxing effect on subjects than non natural environments (cf. Moore, 2012). We decided to opt for scenes in which there are no individuals and to focus on scenes with a naturalistic context.

2.2.3 Equipment-related stress

The third and final group listed as a source of stress comes from the use of equipment designed to enjoy multimedia content. In fact, each multimedia production needs a particular Setup that the user must employ to enjoy the content properly. Clearly, the complexity and interference of equipment increases for the enjoyment of VR content when compared to two-dimensional content on traditional screens. The studies reviewed were very useful in order to guide us in choosing a latest generation headset, which, according to the manufacturer, offers an excellent compromise between the possible sources of discomfort and performance.

This results in widespread stress due to a whole range of physiological and psychological responses due to the general comfort characteristics of the equipment.

Indeed, as highlighted in previous studies, the weight of VR equipment (cf. Yan et al., 2018) or the creation of microclimates (cf. Wang et al., 2020) while using VR headsets leads experimental subjects to report a feeling of discomfort, which then results in stress at home psychophysically.

Another element responsible for raising stress levels is due to suboptimal calibration of VR equipment. In this regard, manufacturing companies often provide ad hoc calibration software (in our specific case HP Eye Tracking Calibration).

In addition, it is common practice to find a manual interpupillary distance (IPD) calibration mechanism on the glasses of the headset.

Effective calibration massively affects the experience with VR equipment. In fact, psychophysical sources of stress can result from the dysfunctional setting of the headset or visual apparatus of the headset. Stress in this area is caused e.g. by insufficient VR headset calibration, limited field of view (FoV), technically induced motion blur, i.e., lack of focus, lag or sudden effects, or accommodation/vergence conflict in the near field of the user, which is on the other hand the important personal space for interaction.

In the context of measurement, only this third group “equipment-related stress” is considered as a permant baseline for basic stress in the VR area.

To enable optimal measurement of this third group, the first two groups, i.e., stress resulting from conflicts between virtuality and reality and content-related stress, must be largely eliminated in the item of our test series.

2 DEFINITIONS & RELATED WORK

Three groups (Figure 5) of potential stress that can be triggered when using a VR headset have been identified:

The first group is stress resulting from conflicts between virtuality and reality. This includes, for example, sensory discrepancies such as staying in a real, warm room while at the same time being in a VR room that displays a cool environment. In addition, cybersickness, simulation sickness and widespread phobias such as the fear of heights, arachnophobia or claustrophobia. Furthermore, translocation effects also belong to the conflicts between virtuality and reality, which arise because a person has to reorient himself or herself when entering virtual space or reality, and this can possibly cause stress. Motor restrictions of the body and virtuality, including a possible loss of control and the resulting physical and visual discrepancies as well as emotional bias or cross talk also fall into this area. Last but not least, the experience of a user through the use of a VR headset also plays a role.

The second group includes content-related stress, where stress is caused by the content presented in the virtual space. This includes audiovisual effects such as flickering lights or shrill noises. Design principles, the placement of content or the choice of colors could also have a certain impact on stress. Other inputs that can influence the individual stress level are jump scares, for example from objects or figures suddenly appearing, a field of view restricted by the content, narrow, angled rooms, crowds of people or loneliness through wide, dreary plains. Further stress could arise from motor, cognitive or timing challenges that put a person under pressure and thus trigger stress. Fast movements, especially if they were not triggered by the user, can affect well-being and thus stress. A content-related loss of control could also arise here.

The third and thus last of the defined groups consists of equipment-related stress. This generally includes the wearing comfort and the VR equipment. Stress in this area is caused, for example, by insufficient calibration of the VR headset, a limited field of view (FoV), technically induced motion blur, i.e. a lack of focus, lag or sudden effects or the accommodation-vergence conflict in the close-up range.

Now that causes of stress in VR are identified, a vital step for the planned experiment was to find ways of measuring stress and determine whether they are suited for our research and realizable with the available equipment (see chapter 5.6).

3 ISOLATING THE POTENTIALLY STRESS CAUSED BY THE VR-APPARATUS

As a result, the following Figure 5 shows an unweighted taxonomy of possible sources of stress when using virtual reality. The issue can be separated into three categories: stress caused by conflicts of reality vs. virtuality, content-related stress and equipment-related stress. Unweighted means that the occurrence of one factor may potentially mask other factors with minor impact at the same time.

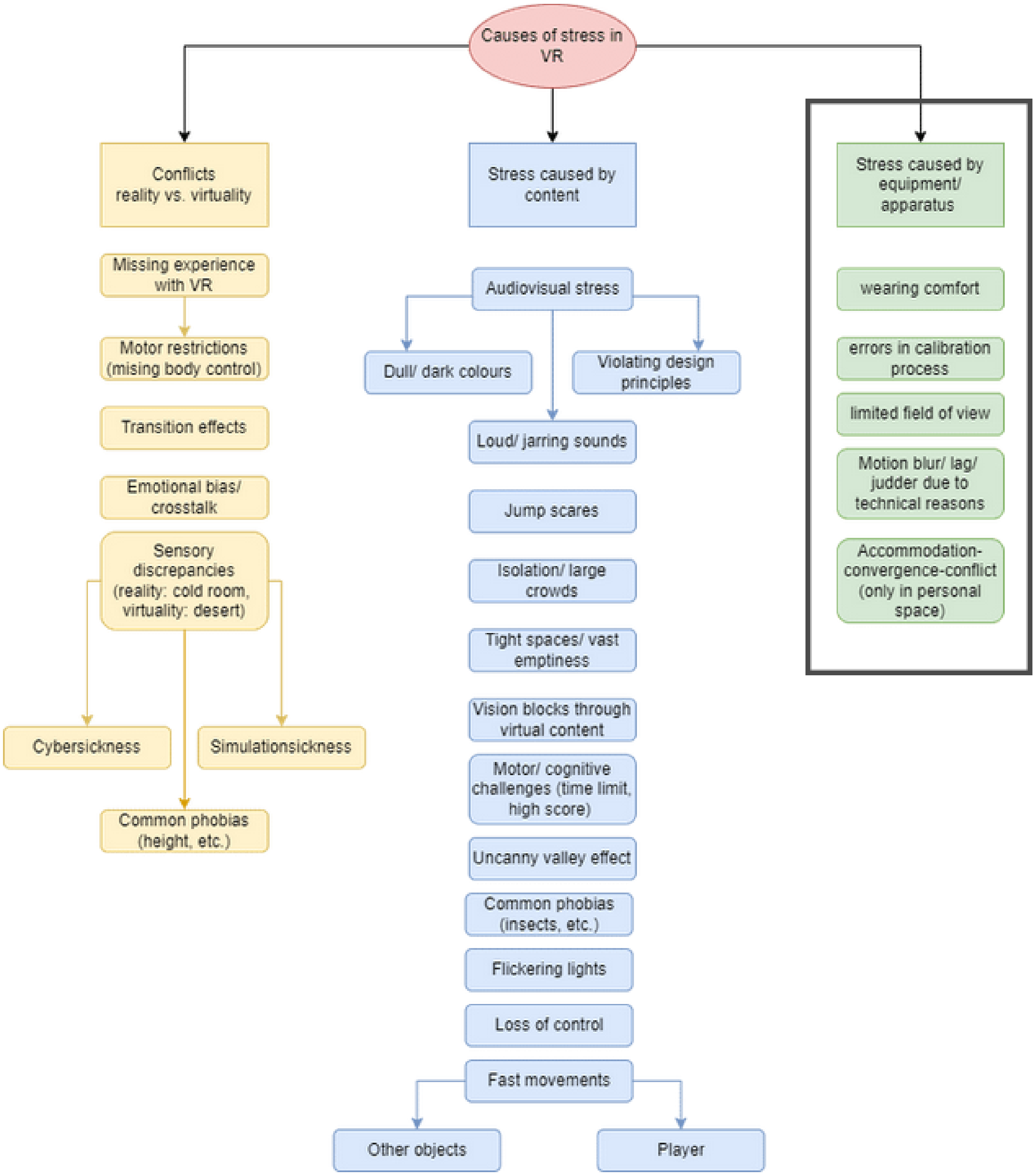


Figure 5: Taxonomy of stress in VR, unweighted (Alka et al., 2022)

4 MEASURING STRESS

It was important that the intended biomedical sensors would not become additional sources of stress due to how uncomfortable they are. Furthermore all methods have to be noninvasive, since none of the team have medical training and we are therefore not allowed to take blood samples e.g..

Measuring stress in individuals is a common part of research, so there are several reliable methods already established that can be used in this experiment. Both objective and subjective data will be collected and compared during analysis. One objective method of measuring stress is monitoring a subject's heart rate, as a person's heart rate will increase compared to their resting heart rate when confronted with a source of mental stress (cf. Taelman et al., 2009). What needs to be considered when evaluating heart rates is, that the values can vary between individuals so that the subjects' data can not be compared directly. Instead it will be necessary to take baseline measurements to identify a subject's resting heart rate and then observe how it changes when exposed to our test item.

For measuring heart rates, our team was equipped with a "movisens EcgMove 4" sensor that can be attached around a subject's chest via a strap. Since this method does not interfere with the VR headset, it can be used in our experiment but there is the need to address this sensor in an interview held with the subject after the experiment, to determine if they were irritated by it, to exclude it as a possible cause of stress.

The ECG sensor also allows measuring the heart rate variability (HRV), "the variation over time of the period between consecutive heartbeats" (Acharya et al., 2007, p. 1031). It is a preferred way to measure stress. Its only drawback is that it can only be measured every 30 seconds. Even if this resolution is sufficient, it could lead to a case in which artifacts could be masked by its resolution. In a case of stress the HRV would decrease.

Also available was a "movisens EdaMove 4" sensor that records a subject's skin conductance level. A common stress response in humans is an increased sweat production. An increase in skin conductance would indicate that a subject's stress level increased (cf. Bach et al., 2010). The used sensor will be attached to a subject's wrist with a band and two electrodes will be stuck to the inside of their palm, measuring the conductance level. Comparing different subject's values will also not work here. Instead each subject's data will be compared to their baseline measurement.

Another method of measurement that was considered for this experiment was to measure a subject's brain activity, more specifically their alpha power, with an EEG or BCI, because high alpha power correlates with relaxation (Hofmann et al., 2020). This method could not be used for our experiment for several reasons. For one, the needed equipment was not available at the time and the team was lacking expertise to execute the complex AI-based data analysis in the given time.

4 MEASURING STRESS

Another methodological problem was, that these devices would not fit on subjects' heads if they wore a VR headset. A similar issue occurred when looking into the possibility of using electromyography to measure face muscle tension. This would require applying electrodes around the subject's eyes where the headset would put pressure on them rendering the data useless.

Stress can also cause spikes in blood pressure (cf. Mayo Clinic Staff, 2021), but the equipment to observe this value was not available at the time or the devices were not suitable for our test setup. Measuring blood oxygen levels with pulse oximetry was also considered (cf. Yoo & Lee, 2011), and while obtaining the necessary equipment it was decided to exclude this method in an effort to minimize the amount of devices a subject would need to wear. A type of measurement that does not require additional sensors is observing changes in a subject's saliva's cortisol level (cf. Perogamvros et al., 2009). Cortisol is an hormon commonly associated with stress. While the absolute level is dependent on time of day and lighting conditions, an increase correlates with rising stress levels (cf. Takai et al., 2004). This makes it possible to take several saliva samples during the experiment and observe if a subject's level changes compared to the baseline measurement. Analyzing the cortisol levels of our samples requires sending them to an external laboratory.

One reason why the HP Reverb G2 Omnicept Edition was chosen as the VR headset for this experiment were its built-in sensors such as a Tobii eye tracker, a heart rate sensor and a face camera. The data of these sensors can be accessed through the game engine "Unity" that was used to create the item that subjects would watch during the experiment. Additionally the combined data of these sensors is used to calculate the wearer's cognitive load, a value between 0 and 1 that indicates a subject's mental stress (cf. Siegel et al., 2021). This value is meant to evaluate the difficulty of a task, with low values correlating to easier tasks.

The built-in eye trackers also allow to capture the wearer's pupil diameter. When stressed, our pupil's diameter expands (Yamanaka & Kawakami, 2009), so this data can be used in our analysis, although some circumstances have to be considered. Due to a lack of equipment, it was not possible to measure a subject's pupil diameter in the correct lighting conditions as a baseline. The comparison group that watched the content on a TV also had no eye data measured. Additionally, the content the subjects viewed was not consistent in brightness so that changes in pupil diameter could correlate to eye or head movements over the scene . That being said, it was still possible to observe if there is an increase in pupil diameter over time. Since the headset was purchased during this research, there was no prior experience when it comes to the reliability of the eye trackers. Therefore a prior test was performed to observe its behavior under changing light conditions (see chapter 5.3)

4 MEASURING STRESS

The eye tracker’s data would theoretically allow us to analyze gaze tracking as an indicator for stress (cf. Jyotsna & Amudha, 2018). This method was not used due to the lack of eyetrackers in the TV comparison group.

In addition to these objective methods of measuring stress, there is also the possibility to observe a subject's subjective perception of their own stress level with the help of questionnaires. Four different types of questionnaires were chosen for this experiment: Raw-TLX, VRSQ, Semantic Differentials and an interview with qualitative questions.

Raw-TLX is an unweighted version of the NASA Task Load Index that is meant to measure subjective workload (cf. So & Gore, 2020). The Raw-TLX is answered after the experiment because the questions relate to it. VRSQ (Virtual Reality Sickness Questionnaire) asks for common symptoms of cybersickness that can occur when using VR headsets (cf. Kim et al., 2018). It is important to exclude cybersickness as a cause for stress in our setup. Since the subjects will fill out this questionnaire before and after the experiment, it is also possible to see if they arrived at the test location in poor health.

The questionnaire with semantic differentials lets subjects give information about their mental state and mood. This questionnaire also needs to be used before and after the experiment to analyze if a mood change occured.

After the experiment is finished the subjects are interviewed with a number of qualitative questions that are meant to reveal any flaws in our test setup that could have caused unwanted stress. This interview includes questions about the equipment, the virtual scene and the whole experimental setup respectively the test execution.

5 RESEARCH QUESTION & METHODOLOGY

5.1 RESEARCH QUESTION

As mentioned, the focus of the research is equipment-related stress of VR headsets. The experimental setup will aim to eliminate all other causes of stress in VR that were found in our research. The dominating research question of this project will be: Does equipment-related stress in VR exist? This means the experimental setup must isolate this from all other known causes of stress in VR, especially conflicts between reality and virtuality, and content-related stress. Additionally the setup has to provide a proper calibration of the headset to avoid mismatches in use. Design principles of VR scenes had to be considered to avoid accommodation-vergence etc. in use. With all these causes of stress taken out of the equation any still measured indicators of stress can be linked to equipment-related stress caused by discomfort or technical limitations of the headset. If an unavoidable stress in VR caused by the equipment does exist, future research of stress in VR will have to keep this in mind and measure a baseline in a similar setup to ours as a way to provide a baseline that can be compared to the measurements of the actual item. Otherwise all measured stress can not be precisely linked to the actual field of interest.

It is important to note that our results will only apply to the headset we used in our research (HP Reverb G2 Omnicept Edition) and the existence and intensity of equipment-related stress in other headsets may vary.

5.2 RESULTS OF PRIOR EXPERT TESTS

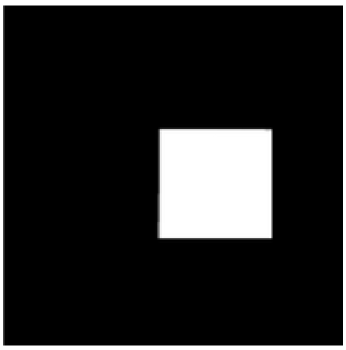
Prior to this experiment an expert test was performed in december 2021 by a collaborating team from the masters course 'Human Factors' at Furtwangen University in Tuttlingen. The experimental setup was designed by this team and was meant to verify the planned measurement procedures for the actual experiment. This was done to verify that the methods will provide usable results, reveal errors in planning and allow adjustments based on the gained experience. The largest difference that was made based on results of this expert test was that subjects will no longer be shown the scene both in VR and on a TV, due to the long duration of a single test. Instead the subjects will be separated in two groups that will either be shown the content in VR or TV. The VR group will additionally be separated on people with and without VR experience to examine if that could have influence on the subjects stress response.

The expert test allowed us to improve the experimental setup, such as choosing a suited room size, the number of team members needed to take care of one subject test and the installation of a wooden TV-box that is meant to ensure a similar field of view for TV and VR subjects. This test also helped decide which kinds of sensors will be used to gather objective data from future subjects (cf. Alka et al., 2022).

5.3 DIGRESSION: PUPIL SIZE VS. BRIGHTNESS

The built-in Eyetrackers on the HP Reverb G2 Omnicept Edition are able to measure the wearer’s pupil diameter. These measurements can be used to identify stress in subjects because there is a correlation between mental stress and an increased pupil diameter (cf. Yamanaka & Kawakami, 2009).

Since the headset was purchased we had no experience as to how accurate the measured pupil diameter data by the headset is. To test the sensor's behavior a setup was carried out in which subjects would wear the headset and view a scene that consisted of a black background and a square (Figure 6) that would over time change its color from black to white in 11 stages (Unity HSV, V-value/brightness increased from 0 to 100 in increments of 10). Only part of the display would increase in brightness to ensure that no automatic powersavings during a full white picture would occur. This test was meant to show if increasing the display’s brightness would result in a reduction of the pupil diameter as expected.



test image
(left eye)

Figure 6: Pupil size test image (Alka et al., 2022)

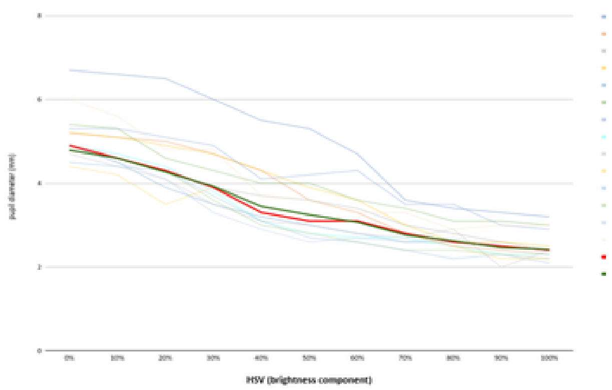


Figure 7: Pupil size results (Alka et al., 2022)

While there are some typical errors in measurement, a correlation between brightness and pupil diameter can be observed (Figure 7). This means if the scene used for our experiment has a constant brightness, stressed subjects should show an increased pupil diameter.

In realistic VR-scenes, however, this can hardly be maintained (see chapter 7.1.4)

5.4 METHODOLOGY

In order to improve the research design of the expert test of the winter semester 2021/2022, the research methods were completely overhault. It combines a series of objective and subjective research methods which are well established and designed for virtual reality research and stress research. As already stated in the chapter objective and subjective research methods, the selected objective research parameters to measure stress are heart rate, heart rate variability, cognitive load, skin conductance level and the cortisol level. The chosen subjective methods encompass of a series of questionnaires, which document the subjects demographic, their emotional state (semantic differ-entials), their experience along with some critics about the test conditions and the quality of the shown VR scene, an VRSQ before and after the test which gives an overview about the subjects VR tolerances and Weaknesses in the Application and finally an NASA Raw-TLX test which reviews the subjects temporal, physical and ministaral demands. To record the objective data several devices were used. Heart rate and heart rate variability are recorded with the EcgMove 4 ECG sensor by Movisens. The skin conductance level of a subject is recorded via the EdaMove 4 EDA sensor, also by Movisens. In order to measure the cognitive load as well as the pupil size, the integrated sensors of the HP headset are used. Cortisol levels are being measured via salivettes which are analyzed by a third party laboratory.

The subjective questionnaires on the other hand are presented and answered via a digital survey platform from the subject. Some of the questionnaires such as the VRSQ and the semantic differential questionnaire had to be answered twice to measure the differences, once before and after the test. The demographic questions are being asked right at the start of the test, this is critical as the subject has to fulfill and fit the target demographic the test aims to test, otherwise the subject has to be put into a specific test group or rejected. The NASA Raw-TLX test on the other hand had to be answered specifically after the test to measure the test difficulty to the subjects: fitness, speed and attention.

5.5 HYPOTHESES

Now that there is an overview of the data that will be collected during the experiment, we are able to formulate hypotheses that may be used to evaluate the results.
All following sub-hypotheses will have the following base hypothesis:

Equipment-related stress in VR-headsets does exist.
As mentioned, this will only apply to the headset used in this study and not include avoidable causes of equipment-related or content-related stress, because these were excluded by the experimental design. The hypotheses will be separated into two categories. Category A will include hypotheses that will be used to verify that the experimental design was executed

correctly and that the subject groups are comparable when it comes to demographic data. Category B will be used to compare hypotheses against the actually collected data and either verify or refute the base hypothesis.

Category A: Experimental conditions

1. The subject groups were comparable in terms of age, gender distribution, eyesight and VR experience.
2. The interpupillary distance (IPD) of the VR test subjects was typical for age and gender and was within the technical IPD possibilities of the VR headset.
3. The duration of the tests was comparable for each group.
4. The subjects started the test in a state of relaxation or low emotional arousal.
5. The experimental setup had no negative influence on the subjects.
6. It was possible to identify every phase of the test in the collected data.
7. Conflicts between reality and virtuality could be avoided/ minimized in the experimental setup.
8. Content-related stress could be excluded from the experimental setup.

Category B: Headset-related stress

1. The semantic differentials were more negative after VR tests and unchanged after TV tests.
2. The VRSQ results were more negative after VR tests and unchanged after TV tests.
3. The Raw-TLX results indicate that VR subjects felt more stressed after the test than TV subjects.
4. The qualitative questions show negative reactions towards the VR scene but not towards the TV scene.
5. The subjects' heart rates rose compared to the baseline measurement for VR subjects and stayed unchanged for TV subjects.
6. The subject skin conductance level rose compared to the baseline measurement for VR subjects and stayed unchanged for TV subjects.
7. The subjects' cortisol values rose compared to the baseline measurement for VR subjects and stayed unchanged for TV subjects.
8. The subjects heart rate variability declined compared to the baseline measurement for VR subjects and stayed the same for TV subjects.

Only VR group (headset sensor data):

1. The subjects' pupil diameters increased in size during the VR scene.
2. The subjects' cognitive load values rose during the VR scene.

5.6 ITEM & SETUP DESIGN

In order to support the hypothesis, a specific VR environment has been created which filters out unwanted stress. Unlike a typical VR application it should be free of all content-related and virtuality vs. reality stress factors as they are not the aim of our inquiry. The content-related stress factors include, but are not limited to, flickering lights, movement, jump scares, loud sounds, natural fears and phobias, unnatural and stress related colors like black, white, red or bright yellow. (Die Macht der Farben – Farb- und Lichttherapie als Heilmittel. n.d.)). Virtuality vs. reality related stress can occur by cybersickness and translocalization effects.
The hypothesis should therefore, conclude that the scene becomes an content-related and an virtuality vs. reality stress-free environment, which is capable to show pure symptoms of equipment-related stress.

The test scene was build in Unity's Universal Render Pipeline. Originally it was proposed and planned to build it inside the High Definition Render Pipeline, as the improved details would have brought better immersion and the scene would probably not be as game-like compared to the URP version.

The improved lighting and post processing effects would have lessened the gap to an real environment. The chosen theme for the scene was a forest (see ATTACHMENT 8, RELAXATION SCENE) because of several reasons:

For one, forests include the most relaxing colors known to the field of color psychology blue and green (Die Macht der Farben – Farb- und Lichttherapie als Heilmittel. n.d.))., also the geographic home of most subjects of the study is the black forest of germany. Which not only fits the most popular retreat spot of said subjects but also reduces translocation effects as most subjects are very familiar with such a setting. The forest scene itself is structured as follow: to the left the subject can see a small hill with some trees and a mountain range in the distance which the complete background, directly in front of the subject is a waterfall which flows into a river which leads the subject's eye movement to its directly surroundings, an second significantly larger hill on which an medieval statue can be seen. In front of said hill is another patch of large trees. In the sky, some larger and smaller birds can be spotted and in the river some fishes can be seen. These animals are supposed to occupy the subject's attention for a while so the item is not getting too bored in the five minute long timeframe of the test. Besides the visual Impressions some auditive features were also added, mainly to relax the subject further, as sound design is a big factor when it comes to relaxing scenes (see "Music designed for relaxation shows a slower tempo and can quiet your mind as well as relax your muscles" (c.f. University of Nevada, n.d.).

5 RESEARCH QUESTION & METHODOLOGY

Excluding translocation effects

The translocation effect is a phenomenon that occurs when a person is suddenly transported into a completely different environment. This phenomenon is especially common in video games and especially VR applications because of the intended purpose of sending a player into a fantasy world where they can completely immerse themselves and forget the stressful real life for a moment. For our setup however this effect could be an artifact which could distort the results as the transition from reality to virtuality can be quite stressful. Because of this an optimisation of the test setup was needed in order to minimize this effect. Therefore the item was modified to firstly match the real test environment as much as possible an then slowly tranfer into the forest scene. Consequently the black wooden box which limited the TV testers' field of view, along with the TV itself and a projection of the relaxation scene on the virtual TV, was integrated into the VR environment. This asset is shown for a few seconds when the subject enters the VR environment. After that, it slowly fades away. Finally, the forest scene becomes visible which the subject should have already seen on the screen of the virtual TV.

Excluding stereoscopic-related perception irritations

While equipment-related stress was the objective of this experiment, the goal was to execute the test in an technical ideal VR setup without any design faults or mismatches. That means all efforts were made to exclude or reduce all known causes of incorrect operation, leaving only factors that are unavoidable with the current technology.

All subjects had to perform a test evaluating their stereoscopic ability. In addition, their interpupillary distance (IPD) was measured to make sure it was within the technical limitation of the used headset. Before a subject was exposed to the relaxation scene, a calibration of the headset was performed with help of an calibration program by HP that calibrates the eye trackers and shows the IPD the wearer needs to set. To avoid the accommodation-vergence conflict, all content placed in the virtual scene had an appropriate distance to the subject's virtual location. The subjects were seated in a comfortable garden chair and are not able to walk around in the scene, head movements aside.

Comparison group setup

For reasons of methodological stringency, we build a the black wooden seating-box which limited the TV subject's field of view to the same angle as in VR (110 degrees), using a 65 inch 4k-OLED-TV in appropriate distance (see Figure 9). Item was presented monoscopic. We fell below the 60 cpd limit of the designed viewing distance of the TV, but the image impression corresponded quite good to the image resolution inside the VR headset.

The subjects were also seated in the same comfortable garden chair. VR-Headphones were exchanged to nearfield audio speakers with comparable sound characteristics.

6 EXPERIMENTAL SETUP

6.1 GENERAL CONDITIONS

As we need to guarantee the same and optimal prerequisites that allow us to get comparable data from our equipment and by survey without any problems, we searched for our subjects with general conditions which we mentioned clearly in the test subject tender.

Thus the subject should be between 18 - 34 years old, has no visual disorders (such as color blindness, squinting, other than vision correction) and no photosensitive epilepsy. Also the subject should have no illnesses during the test period (e.g. colds, fever, corona,...). There must be no consumption of coffee, alcohol or other drugs on the evening before/on the day of the experiments and he or she should have no other stressful activities/events (exams) during the test period (on the day before the test, until 3 days after). Lastly, subjects are not allowed to wear mascara or lipstick during the test to prevent false data from eye tracking or cortisol level. All this creates a solid fundation for a good set of data.

To confirm and document those defined conditions, every subject had to fill a demographic questionnaire. Additionally the investigator asked about visual disorders and led the subject through the Randot Stereotest, which is used for adult stereo testing. The other conditions concerning stressful events and the consumption of alcohol and drugs were on trust basis.

Test room conditions

To make sure every subject has comparable conditions during the test, the procedure followed a script (see Appendix 1). The room was aired before and after the test, the room temperature was documented. The experimental setup was always the same and the curtains were closed during the relaxation scene and widely open during greetings and interviews.

6.2 TEST GROUPS

The subjects are separated into two test groups, furthermore abbreviated "TV" and "VR". The VR group consists of VR experienced and VR inexperienced participants. The goal was to have 36 subjects (12 per section) to have a buffer against possible exclusions. But due to a lack of registrations this could not be achieved. A total of 23 subjects participated in the experiment from May to July 2022.

6.2.1 TV (comparision group)

The TV Group includes the subjects that see the same item/scene in the TV box, it has a size of N=11.

6.2.2 VR

This group includes the subjects that see the scene in VR, it is divided into two subgroups. One of them contains the VR experienced (VRE) and the other one the VR inexperienced (VRU) attendees. Overall five VREs and seven VRUs make up the VR group (N=12).

6.3 TEST EXECUTION

The test was designed to last over all phases between 40 and a maximum of 50 minutes per subject. One hour was scheduled for each subject due to preparations and follow-up, which means that every time slot in the appointment calendar, in which the test persons could enter themselves, could be booked on full hours. The measured time spans (each noted in the protocol) were within the contemplated time range (hypothesis A3 can be accepted).

The rough process was defined within a script (see Attachment 1), with a few points differing in VR/TV, such as calibrating the headset. This important textual requirement has been broken down into nine steps/phases:

- WELCOME - STEP 1

- RECONNAISSANCE - STEP 2

- CORTISOL TEST & EYE TEST STEP 3

- FIRST QUESTIONNAIRE - STEP 4

- PREPARATION & CALIBRATION - STEP 5

- CALIBRATION - STEP 6

- IMPLEMENTATION - STEP 7

- SECOND QUESTIONNAIRE MARATHON - STEP 8

- SANITIZE EQUIPMENT - STEP 9

First the subject is welcomed, then the person goes through the welcome phase to confirm general conditions. At step three the first objective data is collected of the subject by the first cortisol test. After this the RANDOT stereo test is done. If the subject does not meet the conditions, the person cannot take part in further investigation and is thus politely seen off. Step 4 includes the first set of questionnaires to collect subjective data before the actual test scene.

In step 5 to 7 the subject does the second cortisol test and finally puts on the VR headset/ or is sat before the TV thus dives in the relaxation scene. Here objective data (eda, ekg) are collected. In step 8 objective (Cortisol test 3) data and subjective answers are gathered to compare subjective subject sensation before and after the VR/TV test. Lastly, step 9 is to sanitize the equipment to meet hygienic standards.

6.4 TEST SETUPS

At the beginning and at the end of the test the subject are seated at a desk with a computer (Figure 8) where they fill in the declaration of consent by hand and the questionnaires digitally. On the other side of the room are the technology station from which the scene and the measuring devices are started and the equipment table where the protocol is written and items like the stereo test, the cortisol samples and a neck pillow are placed. In front of them there is the black, comfortable seating box with a garden

6 EXPERIMENTAL SETUP

chair and speakers on the left and right where the subject takes place after the first phase of questionnaires. Then the test starts (Figure 9, Figure 10).

The described setup is the same for both test groups except for the VR headset users which are additionally placed on the equipment table for the VR experimen to explain the VR-equipment. The external speakers were pushed to the side because of the VR headset's integrated headphones.

Figure 8: Setup for survey (Alka et al. 2022)

Figure 9: Setup TV box / comparison group, nearfield loudspeakers and chair (Alka et al. 2022)

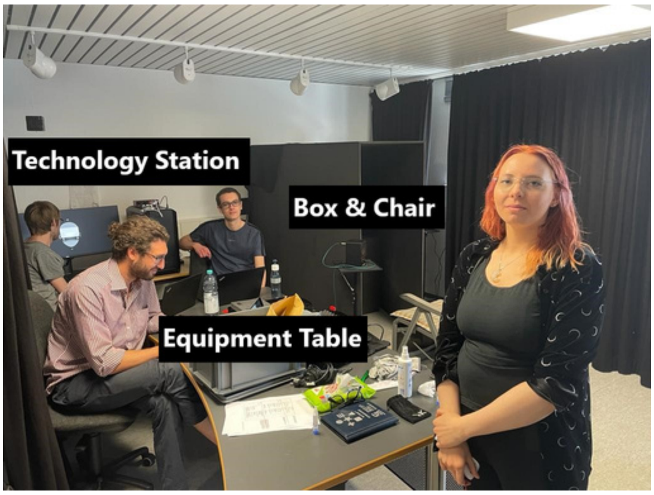


Figure 10: Setup VR, overview (Alka et al. 2022)

7 RESULTS & DISCUSSION

7.1 HYPOTHESES VS. DATA

7.1.1 Questionnaires

7.1.1.1 Demographic data

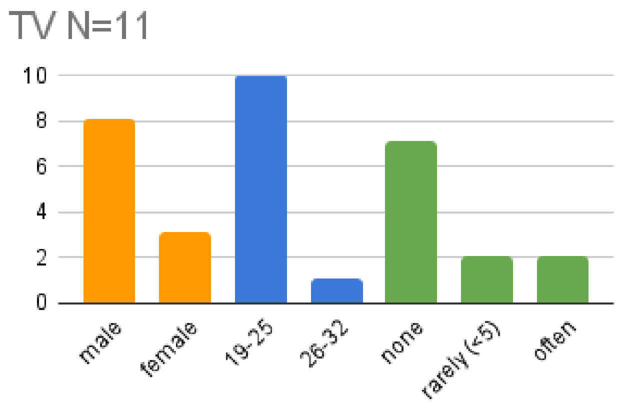


Figure 11: Demographic data TV (Alka et al. 2022)

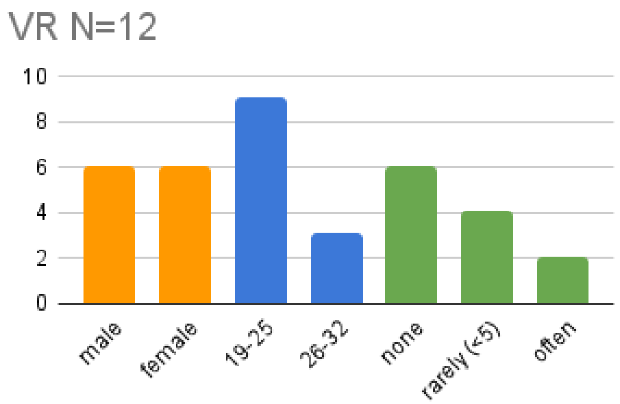


Figure 12: Demographic data VR (Alka et al. 2022)

The charts (Figure 11, Figure 12) show the distribution of gender, age and VR experience separated in TV and VR. Bboth groups contain male and female (no divers) participants. The VR group has an equal number of men and women, the TV group contains 72.73% male and 27.27% female subjects. Still this seems to have no impact on the results of the pilot study as we will see in the further evaluation.

In terms of age and VR experience, the two groups are comparable. The age ranges from 20 to 26 years with a median $\tilde{x}$ = 22 in the TV group and 19 to 32 years with a median $\tilde{x}$ = 24 in the VR group. So the VR subjects are slightly older and have a greater spread. But all are within the target age range.
The VR experience is also very similar between the two groups. The majority has never worn a VR Headset before, only two have used one between 5 and 15 times and nobody uses it on a regular basis. If we combine the above evaluations, hypothesis A1 can be verified.

Interpupillary Distance

Gender	Male	Female
Number	1771	2205
Minimum	52	52
Mean	64.67	62.31
Maximum	78	76
St'd Dev.	3.708	3.599

Table 1: Interpupillary distance gender (cf. Dogson, 2004)

According to the paper 'Variation and extrema of human interpupillary distance' (Figure 13), the average interpupillary distance is 64.67mm in males and 62.31mm in females (Table no.). In the TV group male participants have an average interpupillary distance of 64.38 mm, while the average for female participants is 63.33 mm. This is in line with the findings of the paper. For the VR group, the average interpupillary distance is 66.67mm for males and 63.50mm for females. These values are slightly higher than for the TV group (cf. Dogson, 2004).

7 RESULTS & DISCUSSION

Age group (years)	Gender	*n*	Min	Max	5th	95th	Mean ± 1st SD*	Median*
16–25	Male	150	56	72	58.06	69.58	63.65 ± 3.38	64
	Female	150	54	67	57.20	65.58	60.90 ± 2.53	61
26–40	Male	150	58	73	60.14	70.32	65.23 ± 3.14	65
	Female	150	55	68	57.33	66.58	61.67 ± 2.72	61
41–65	Male	150	58	73	60.37	70.00	65.50 ± 2.77	66
	Female	150	55	70	58.00	66.41	62.53 ± 2.76	63

Table 2: Average interpupillary distance age and gender (Pointer, 1999)

Age group (years)	Test Group	Gender	N	Min	Max	Mean	Median
19-25	TV	male	7	62	70	64,29	63
	TV	female	3	61	65	63,33	64
	VR	male	5	60	71	66,2	67
	VR	female	4	60	65	62,75	63
25-32	TV	male	1	65	65	65	65
	TV	female	0	-	-	-	-
	VR	male	1	69	69	69	69
	VR	female	2	63	67	65	65

Table 3: Average interpupillary distance age and gender TV and VR (Alka et al. 2022)

The paper 'The far interpupillary distance. A gender-specific variation with advancing age' has broken down the interpupillary distance by both gender and age. It states that males aged 16-25 years have an average interpupillary distance of 63.65mm, while males aged 26-40 have an average interpupillary distance of 65.23mm (Table 14). If we also separate by age in our experiment (Table 15), the 19-25 year old males in the TV group have an average interpupillary distance of 64.29mm, while for the over 25 year old males it is 65mm. Both are close to the values of the paper. Compared to this, male participants of the VR group have a value of 66.20mm for age 19-25 and 69mm for over 25. All values are within the standard deviation.

For women aged 16-25, this paper reports an average interpupillary distance of 60.90mm and 61.67mm for the age range 26-40. 19-25 year olds in the TV group had an average interpupillary distance of 63.33mm, with no one over 25. Compared to this, female participants in the VR group had a value of 62.75mm for ages 19-25 and 65mm for over 25. All of our average values are higher than the ones from the Paper but are within the standard deviation as well. So all in all the interpupillary distance of our subjects was typical for age and gender.
However at this point it must be said that the interpupillary distance was measured by hand with the help of a cardboard pupil distance meter (cf. Optik Peter Meyer, n.d.). Therefore a small deviation of delta max. = 5mm is possible but not influential. The scatter can be seen in the box plot charts below (Figure 13, Figure 14).

Additional note: The callibration of the VR headset was not possible for one subject with very small IPD. It failed several times in a row. Test has been cancelled for this subject.

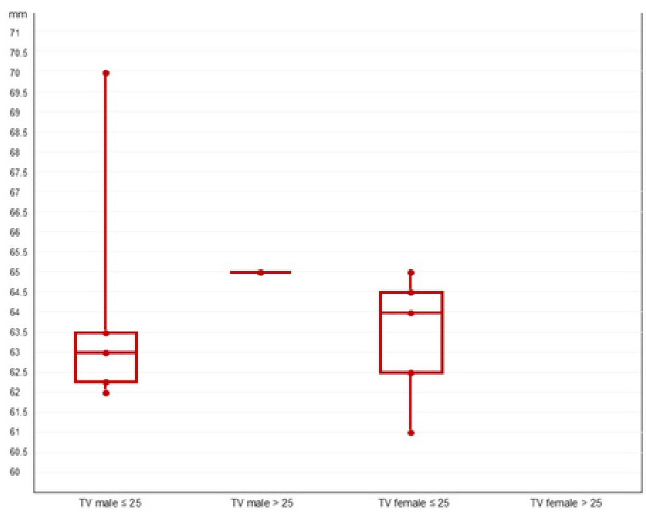


Figure 13: Interpupillary distance box plot TV (Alka et al. 2022)

7 RESULTS & DISCUSSION

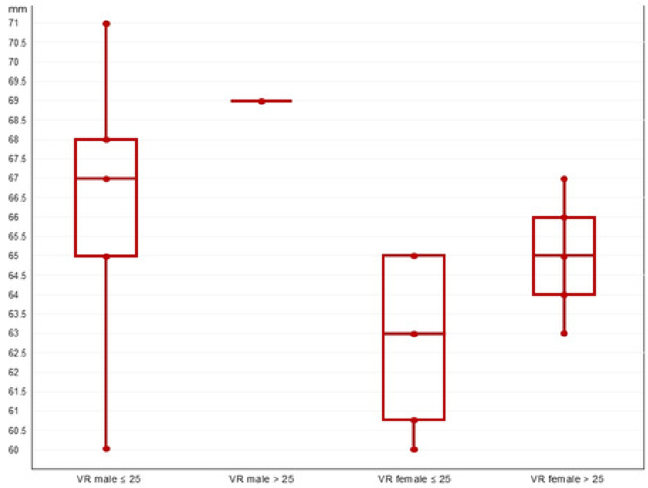


Figure 14: Interpupillary distance box plot VR (Alka et al. 2022)

Stereotest

To ensure that the subjects had sufficient stereoscopic vision, a stereoscopic Randot test was used. This tests both the depth perception of the test subject and normal stereo vision. Each subject was first shown 8, then another 10 Randot stereo images. The 8 images each contained a shape or nothing. The 10 images each contained 3 circles with one standing out stereoscopic from the background. In each case the number of the last circle recognized was entered as the value.

“Normal young adults (20 to 36 years of age) will achieve a stereoacuity [(ability to perceive depth)] of 60 seconds of arc” (Stereo Optical Company Inc., 2017).

According to this statement and looking at the table 4, our subjects have to get at least to the 6th circle to fulfill the 60 seconds of arc. Figure 15 shows the statistical distribution, all of them achieved this goal (Lenscan Medical Inc., 2012).

Scoring Key		Seconds of arc at 16 in.
1	L	400
2	R	200
3	L	140
4	M	100
5	R	70
6	M	50
7	L	40
8	R	30
9	M	25
10	R	20

Table 4: Subject’s stereo scoring key and seconds of arc (Lenscan Medical Inc., 2012)

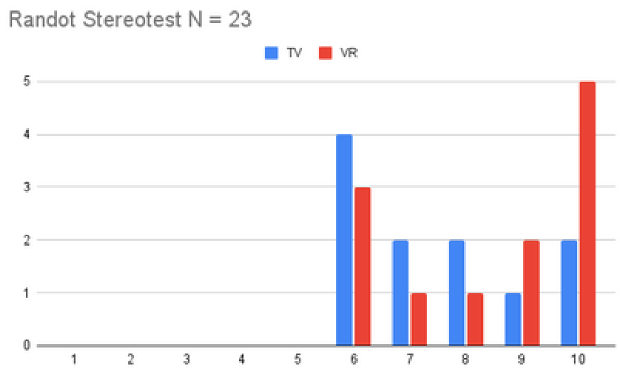


Figure 15: Randot Stereotest (Alka et al. 2022)

Conclusion:

Compared to values from other sources, the interpupillary distance of the VR subjects was typical for age and gender and within the technical IPD capabilities of the VR headset. All subjects had suffucient stereoscopic capabilities. Thus, hypothesis A2 can be accepted.

7 RESULTS & DISCUSSION

NOTICE: Although knowing that if the samples are too small, the power of our tests may not be sufficient to detect a precise difference between the means, we decided to perform an experimental statistical analysis of the pilot study subject data.

7.1.1 Questionnaires

7.1.1.2 Semantic differentials

To measure the affective reaction we choosed semantic differentials (SD). It uses a set of bipolar scales to evaluate a participant's subjective perception of the displayed scene (Snider et al., 1969). All 23 participants were advised to take the SD twice, first before and then after experiencing the item. The goal was to determine the individual changes related to the VR scene. The VR group consisted of VR experienced (5) and inexperienced (7) participants. As stated above the group consists of 14 male and 9 female subjects (TV: 8 male, 3 female & VR: 6 male, 6 female) with an age span of 18-34 years.

TV N=11

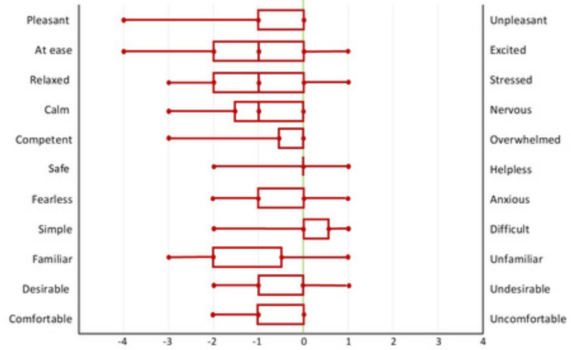


Figure 16: Semantic differentials box plot TV (Alka et al. 2022)

Figure 16 shows the difference in the TV participants' summed up answers before and after being exposed to the item/scene. The green line marks the zero, indicating no change in the participant's answer.

An expansion towards the left means that the subject's answer was more positive after watching the scene. The positive numbers right from the zero represent a more negative answer compared to the first attempt before being exposed to the scene.

Thus an increase in overall positive states is notable, since most of the participants felt at least one or more points better. The only exception is the *simple - difficult* scale, due to three subjects giving one more point to the difficult side of the scale after the scene.

VR N=12

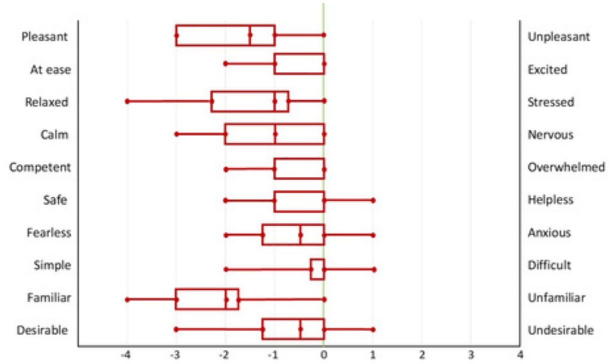


Figure 17: Semantic differentials box plot VR (Alka et al. 2022)

Figure 17 shows the answers for the VR group. They also took the questionnaire twice similar to the TV group the subjects felt overall more positive since there is an increase into the negative numbers in the diagram.

To provide statistical evidence for these trends, we apply the two-sample t-test with two independent samples. To determine whether the equal or unequal variance t-test should be used the difference between the sample variances was tested by the Levene's test for each bipolar scale. To determine whether the equal or unequal variance t-test should be used we used the Levene's test. The therefore needed f-value can be seen in the corresponding F-Table (Attachment 2). We did this for every of the 11 scales. A t-test with unequal variance is per-formed with the first four, and a t-test with equal variance for the other seven. The hypothesis concerning semantic differentials is: The semantic differentials were more negative after VR tests and unchanged after TV tests.

To accept or reject this hypothesis we applied the t-test to each of the before and after responses of each group to see if there was a significant difference between the responses. The null hypothesis for the TV test is: In the TV group, there is no difference between first and second time responses. The alternative hypothesis is: In the TV group, there is a difference between first and second time responses. We make the assumption that there is no significant difference between the responses. So we consider the one-sided test. The t-test tables from the data of the TV group (Attachment 3) show that the p-value one-tail is greater than 0.05 for the following five scales: *Competent - overwhelmed, safe - helpless, fearless - anxious, simple - difficult* and *desirable - undesirable*. Consequently there is no significant difference and we can accept the null hypothesis for them. Between the first and second answers of *pleasant - unpleasant, at ease - excited, relaxed - stressed, calm - nervous, familiar - unfamiliar and comfortable -uncomfortable* there is a significant difference.

Thus we have to reject the null hypothesis and accept the alternate hypothesis here. To track in which direction the values have changed, we look at the boxplot (Figure 16). All six bipolar scales that show a significant difference have better scores after the second round than after the first. If we consider all but *fearless - fearful*, the five remaining scales have 100% better and equal scores the second time. Knowing that only one subject selected a higher value afterwards than before, which is only 9.09 % and thus 90.91 % are at 0 and below, we get that 91.67% gave a lower value the second time or the same value as the first time. Therefore, we can reject our hypothesis, since nearly all of the subjects rated their emotional state better or the same the second time.

The null hypothesis for the VR test is: there is no difference between first and second time responses. The alternative hypothesis is: In the VR group, there is a difference between first and second time responses. The t-test tables we got from the data of the VR group (Attachment 4) show that the p-value one-tail is only at four of the 11 scales higher than 0.05. Between the first and second answers of *safe - helpless, simple - difficult, desirable -undesirable* and *comfortable - uncomfortable* is no evidence of a significant difference. We accept the null hypothesis for them.

7 RESULTS & DISCUSSION

Between the first and second answers of *pleasant - unpleasant, at ease - excited, relaxed - stressed, calm - nervous, competent - overwhelmed, fearless - anxious and familiar - unfamiliar* there is a significant difference. Thus we have to reject the null hypothesis and accept the alternate hypothesis here. The box plot diagram (Figure 17) shows in which direction the values have changed. All of the seven bipolar scales which show a significant difference have better values after the second time filling than after the first time. If we take out *fearless - anxious*, the six remaining scales have 100% better and equal values the second time. We know that only one subject selected a higher value afterwards than before, which is only 8.33%. So 91.67% are at 0 and below. This means that 91.67% gave a lower value the second time or the same value as the first time. This is also evident in the box plot diagram.

Therefore, we can reject our hypothesis B1, since nearly all of the subjects rated their emotional state better or the same the second time.

The values rejects the hypothesis A4. The box plot charts (Figure 16, Figure 17) show large differences between the first and second values for several cases.

The rejection of A4 also supports the rejection of B1. The subjects felt more relaxed after the experiment than before, so they did not start the test in a state of relaxation or low emotional arousal.

7 RESULTS & DISCUSSION

7.1.1 Questionnaires

7.1.1.3 VRSQ

The VRSQ (Virtual Reality Sickness Questionnaire) was developed based on the SSQ (simulator sickess questionnaire) and aims to determine any motion sickness or feeling unwell during the usage of VR (Kim et al., 2018).

Since our focus lies on equipment related stress in VR, our goal here was to exclude stress related to motion sickness. Therefore we chose the VRSQ to determine the subjects physiological wellbeing.

The box plot charts below show how the responses changed in the second phase compared to the first (Figure 18, Figure 19). They show that most of the subjects have neither given a higher nor a lower number as an answer. They picked the same number the first and the second time they filled in the VRSQ questionnaire.

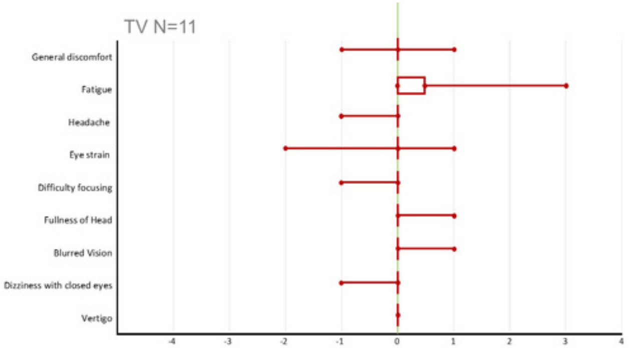


Figure 18: VRSQ box plot TV (Alka et al. 2022)

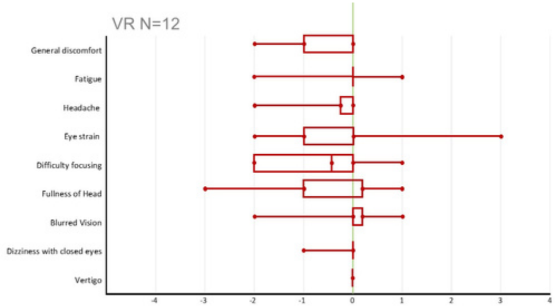


Figure 19: VRSQ box plot VR (Alka et al. 2022)

Again we used the Levene's test to determine whether the equal or unequal variance t-test should be used. We did this for every of the nine disciplines. All but one (*eye strain*) led to an equal variance test.

The hypothesis says that the results of the VRSQ are more negative after the VR tests and unchanged after the TV tests. To accept or reject this hypothesis we applied the t-test to each of the before and after responses of each group to see if there was a significant difference between the responses. The null hypothesis for the TV test is: In the TV group, there is no difference between first and second time responses. The alternative hypothesis is: In the TV group, there is a difference between first and second phase responses.

We make the assumption that there is no significant difference between the re-sponses. So we consider the one-sided test. The t-test tables created from the data of the TV group (Attachment 5) show that the p-value one-tail is greater than 0.05 in all disciplines. There is no significant difference between the two samples, so we can accept the null hypothesis.

7 RESULTS & DISCUSSION

Examining the answers of the VR group, equal variances are for the disciplines *general discomfort, headache, difficulty focusing* and *blurred vision,* the other ones are analyzed with unequal variances. Here the null hypothesis for the VR test is: In the VR group, there is no difference between first and second time responses. The alternative hypothesis is: In the VR group, there is a difference between first and second time responses. We assume that there is no significant difference between the responses. So again we consider the one-sided test.

The t-test tables created from the data of the VR group (Attachment 6) show that the p-value one-tail is greater than 0.05 in all disciplines except for difficulty focusing. Thus there is no significant difference between the two samples for the other disciplines, we can accept the null hypothesis for them. Our hypothesis is rejected for these. But looking at *difficulty focusing* where the p-value is 0.036, we have to reject the null hypothesis and accept the alternate one. This says that there is a significant difference between the answers. Consulting the box plot chart we can find out in which direction this difference goes. We know that only one respondent gave a higher value afterwards than before, which is only 8.33%. So 91.67% are at 0 and below. This means that 91.67% gave a lower value the second time or the same value as the first time. This coincides with the box plot diagram and confirmes our statements according to the optical resolution of the headset.

Therefore we can reject our hypothesis B2 as well, because the majority of subjects rated their concentration problems better or the same at the second time.

7.1.1 Questionnaires

7.1.1.4 Raw-TLX

The Raw-TLX captures workload at multiple levels: *Mental Demands, Physical Demands, Time Demands, Performance, Effort,* and *Frustration*. Subjects were asked to fill out the questionnaire after watching the relaxation scene, to give us an idea of how stressful they found the experiment. The charts Figure 19 and Figure 20 below show their answers in the form of a box plot.

TV N=11

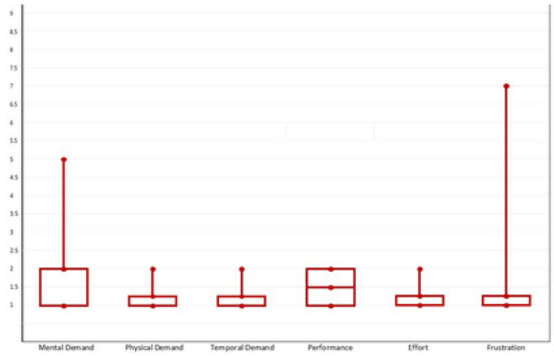


Figure 20: Raw-TLX box plot TV (Alka et al. 2022)

VR N = 12

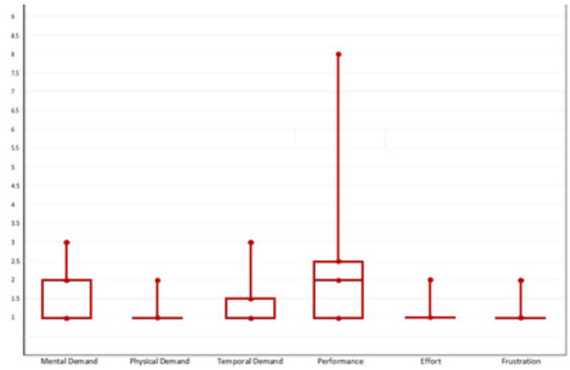


Figure 21: Raw-TLX box plot VR (Alka et al. 2022)

For the statistical analysis of the two samples the two-sample t-test was used. Since we have two independent samples this is the correct choice. To determine whether the equal or unequal variance t-test should be used, the difference between the sample variances was tested by the Levene's test for each level (mental demand, physical demand, etc.). The therefore needed f-value is 2.85. This leads to the t-Test with equal variances for the levels *Mental Demand, Physical Demand,* and *Effort* and to the t-Test with unequal variances for the other three levels (Attachment 6).

The null hypothesis in the t-test assumes that there is no difference between the responses of the TV group and the VR group. The alternative hypothesis is accordingly: There is a difference between the answers of the TV group and the VR group. This is also in line with our hypothesis that the VR subjects felt more stressed after the test than TV subjects. So we go into the test with the assumption that the VR group is more stressed than the TV group. Therefore, we test one-sided.

7 RESULTS & DISCUSSION

Looking at the p-value of the first level Mental Effort, which is 0.26 and comparing it with the significance level (0.05), we see that it is higher. There is no significant difference between the two samples, so we can accept the null hypothesis. This can be substantiated by looking at the value of t-statistics and the one-tail critical t-value. Since the t-statistic value is smaller than the critical t-value, then no difference is detectable at the 0.05 error level. The p-value of the other 5 levels is also above 0.05 and t-Stat is also below the critical t-value in each case. Here, too, the null hypothesis can be accepted.

In summary, our hypothesis B3 'The Raw-TLX results indicate that VR subjects felt more stressed after the test than TV subjects.' can be rejected.

7.1.1 Questionnaires

7.1.1.5 Qualitative questions

The goal here was to create an opportunity for subjects to mention aspects about the shown scene, that were unnoticed by the research team. Therefore, the open questions allow people to elaborate on various auditive and visual aspects, as well as their own perception. This questionnaire helps to determine the mental stress of subjects and eliminate possibilities of non equipment-related stress.

We chose a qualitative approach rather than quantitative here, because we wanted to immerse deeper into the subject's perception and obtain more precise information on the motivation, mindset and attitude of the interviewees, which is why free answers were also permitted and (only) this questionnaire was asked via an interview in order to get better feedback. It was important for us to find out what the respondents paid attention to, how they felt about the test itself, and what they specifically noticed about our scene or what irritated them.

1. Is this your first experiment of this kind?

TV N=11

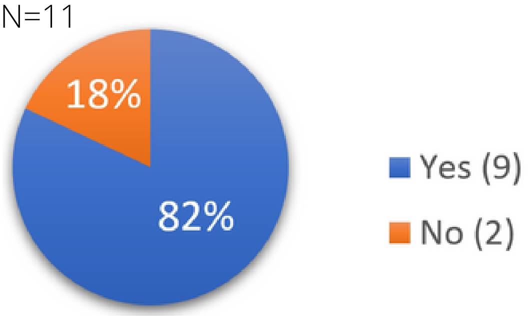


Figure 22: TV group answers question 1 (Alka et al. 2022)

With 82% most of the participants in the TV group have never done an experiment like that before. 18% have taken part in an analogical experiment before. Since it was the first experience of that kind for most subjects, this data can explain initial nervousness or stress observed in the participants.

VR N=12

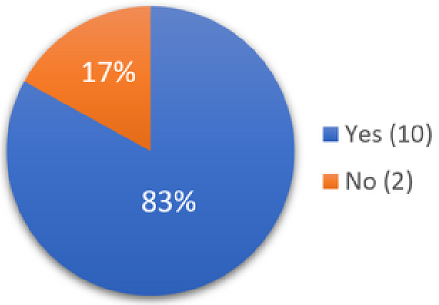


Figure 23: VR group answers question 1 (Alka et al. 2022)

The VR group answered almost identical with two answering "no" and ten stating "yes". Similar to the TV group this data can explain initial nervousness or stress observed in the participants.

2. Was this your first experience with Virtual Reality?

VR N=12

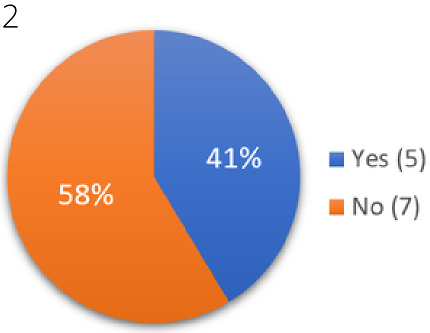


Figure 24: VR group answers question 2 (Alka et al. 2022)

7 RESULTS & DISCUSSION

The VR group is evenly distributed between experienced VR users and non experienced VR users. This mix makes the entire VR group representative of both, experienced users and newcomers.

3. How familiar are you with current VR technologies?

VR N=12

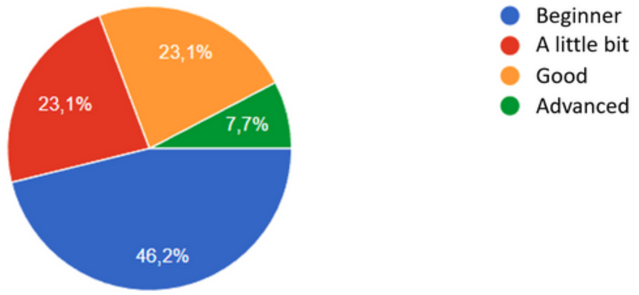


Figure 25: VR group answers question 3 (Alka et al. 2022)

This questions shows the level of experience withing the VR group. That distribution can also be seen as representative, since it covers various experience levels. Almost half are inexperienced, while the other half have previous experience (a little bit - advanced).

4. What kinds of VR have you used in the past?

VR N=9

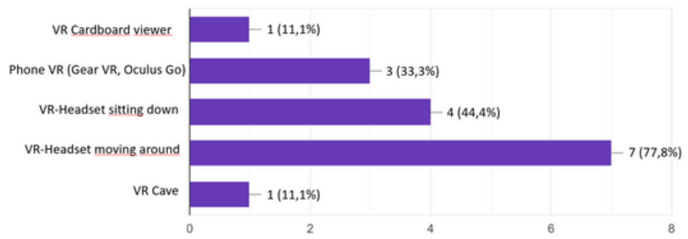


Figure 26: VR group answers question 4 (Alka et al. 2022)

Since three participants never used any kind of VR, only 9 of the 12 VR group members answered this question, as shown in the chart (figure 25).

All of the suggested options were used at least by one subject.

The classic headset with the option to move around freely was the most common used VR kind, followed by wearing a headset but sitting down, Phone VR and last cardboard/VR cave.

This shows an actual trend towards classic VR headset usage for people who seek an immersive VR experience.

This trend could also be explained by the Gaming industry that makes the use of VR Headsets more attractive than the other virtual reality options given above.

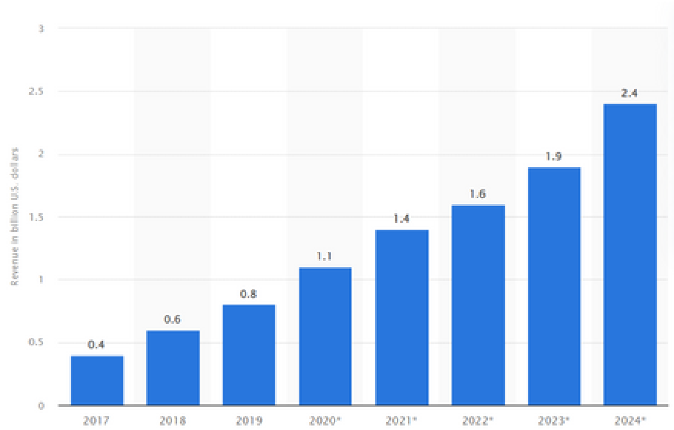


Figure 27: Global VR gaming revenue worldwide from 2017 to 2024 (Statista, J.Clement, 2022)

5. How many hours do you spend on VR in a typical week?

VR N=12

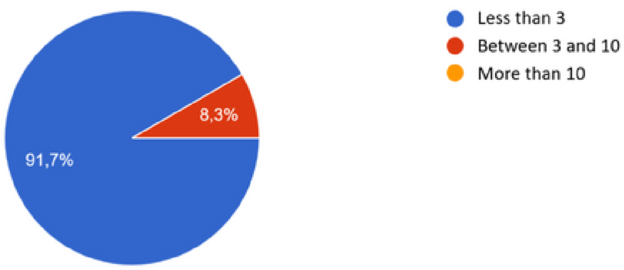


Figure 28: VR group answers question 5 (Alka et al. 2022)

Most subjects said they spend less than three hours on VR in a typical week, while only 8,3% of our subjects spend between 3 to 10 hours a week on VR (Figure 27). No subject spent more than 10 hours on VR, which either shows, that VR is not established in the daily routine, or other activities in real life fill their weekly plans. That again could indicate, that VR is mostly used in private terms, not for studies or during work.

6. How did this experience compare to other VR experiences you've tried? (N=9)

Keywords here have been relaxing, beautiful, and chill.
All participants stated that they had a relaxing and positive experience compared to what they have experienced before.
Thus we can conclude that VR has not been used before for relaxation or relaxation exercises. Alternatively previous use was not relaxing compared to the shown contents.

7. Were you able to relax during the experience? Why?

TV N=11

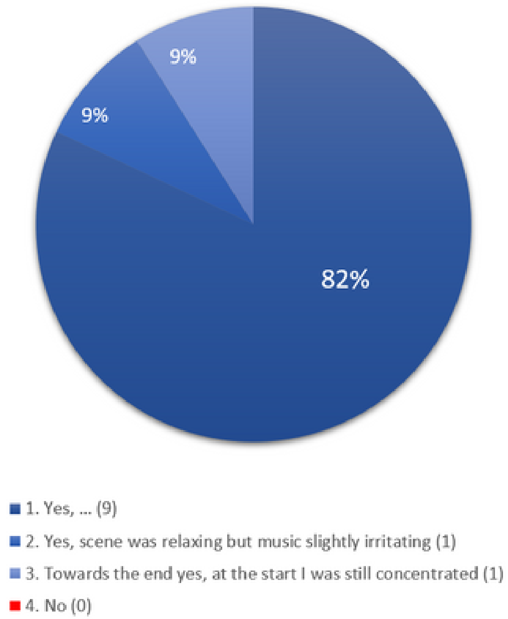


Figure 29: VR group answers question 7 (Alka et al. 2022)

Most said yes (9) but did add a unique elaboration of what they found particularly relaxing (see next question for details). Two participants said it still was relaxing but one was slightly irritated by the music and the other just very concentrated at the beginning.
This indicates a personal and individual sensation each subject has and could either mark a trend if more people would state a certain factor irritating (must be looked on in a test series with more subjects) or it could also be just be personal preferences that vary randomly in between the population.

All 12 participants of the VR group answered this question with yes and gave an elaboration on what they found relaxing. This shows a tendency towards an even more positive outcome and less negative experiences while using VR compared to TV. This claim supports hypothesis A5 “the experimental setup had no negative influence on the subjects”. For details on what relaxing aspects exactly were named, see next question.

8. Have you found any visuals to be particularly relaxing?

Subjects were allowed to name more than one aspect.

TV N=11, VR N=12

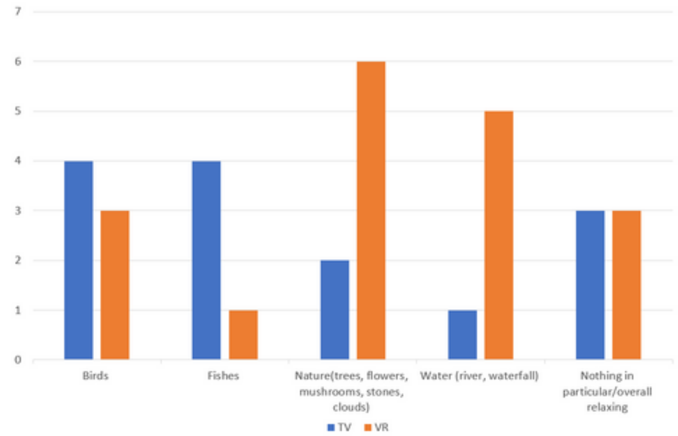


Figure 30: VR and TV group answers question 8 (Alka et al. 2022)

Most said yes (9) but did add an unique elaboration of what they found particularly relaxing (see next question for details).

Two participants said the experience still was relaxing but one was slightly irritated by the music and the other just very concentrated at the beginning.

VR N=12

All 12 participants of the VR group answered this question with yes and gave an elaboration on what they found relaxing. This shows a tendency towards an even more positive outcome and less negative experiences while using VR compared to TV. For details on what relaxing aspects exactly were named, see next question.

9. Were there any visual impressions that irritated you?

While some participants named the birds and the, water quite relaxing, others felt slightly irritated by those. For the birds this might be caused by their flight paths. The birds followed random paths causing them to fly side-or backwards sometimes, which caused some subjects to perceive them as interesting to look at, thus relaxing, while others found that “irritating”.

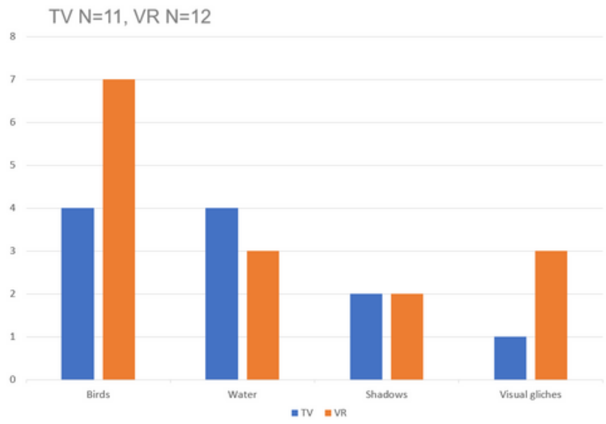


Figure 31: VR and TV group answers question 9 (Alka et al. 2022)

Most of the subjects added that the “irritation” was minor, it was rather something that they noticed. Therefore suggesting that severe irritations from this aspect can be excluded.

10. Was the listening experience relaxing?

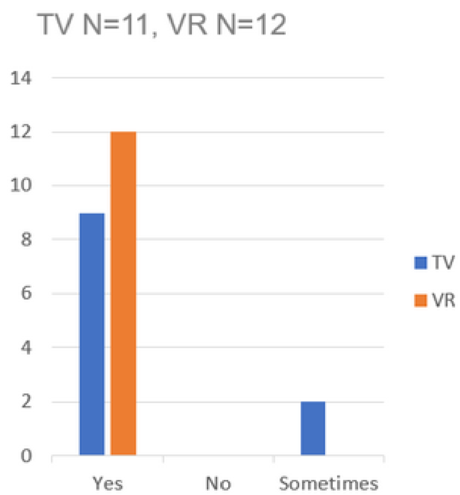


Figure 32: VR and TV group answers question 10 (Alka et al. 2022)

Two participants from the TV group stated “relaxing, but only while the music was playing” and “not disturbing but also not particularly relaxing”. All other subjects in both groups felt relaxed by the audio.

11. Did you find any sound particularly disturbing?

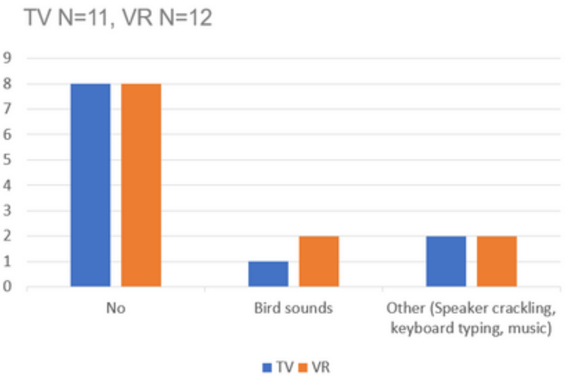


Figure 33: VR and TV group answers question 11 (Alka et al. 2022)

12. Did you understand what you had to do during the experiment?

All of the TV group (N=11) participants did understand what they had to do and answered with “yes”.
In the VR group (N=12) nine said “yes” and three said “no”. This might have been caused by the fact that the subjects didn’t have to perform a task, they were advised to just relax and watch the scene.

13. Were you aware which measuring devices were used?

All participants from both groups (VR and TV) answered with “yes”. During the instruction, the investigator explained all measuring devices and answered subject's questions about the device if they had any. This is reflected in this question's answers.

Furthermore the investigator tried not to give any negative or stressful impacts (for example by never using words such as stress, talking about relaxation instead) on possible answers correlating the subject's stress sensation.

14. Was the duration of the experience appropriate?

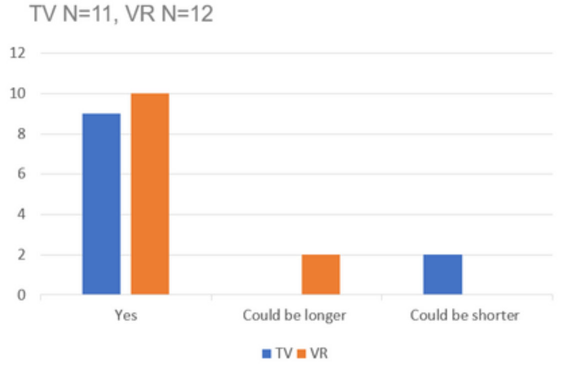


Figure 34: VR and TV group answers question 14 (Alka et al. 2022)

While most of the participants were pleased with the duration, two of the VR group said it could be even longer. It was the opposite for the TV group with two subjects saying it could be shorter. This suggests a trend of the VR group enjoying the experience more, hence wanting to stay in VR for longer time.

15. Were the speakers pleasant?

All participants from both groups (VR and TV) answered with "yes". This indicates that they did not feel disturbed by the speakers presence or position. For the TV group the speaker devices were placed directly left and right of their head in correct, individual height. For the VR subjects the integrated VR headset speakers were used. Therefore the investigator had to observe the subjects, if the speakers were not accidentally flipped up and harming the audio experience. The answers indicate that no such problem did occur.

16. Was the VR headset comfortable?

This question was only answered by the VR group (TV group did not have a headset on). Two of the 12 participants stated that it felt unfamiliar/too warm. The rest (10) stated that they felt comfortable.

Subjective sensation may differ and as shown here, there are cases of unease due to the apparatus. Nevertheless most of our subjects did not have any problem with the VR Headset.

17. Was the test procedure relaxing for you?

All participants answered with yes. This data suggests that content-related stress was successfully avoided from the subjective view of the participants. It also indicates that no stress was sensed during the use of a VR headset from the device, at least sensed subjective.

18. Did you get bored during the experiment?

Both groups had one participant saying yes, all others did not get bored during the experiment.

19. Was the seating position comfortable?

One subject in the VR group said no. All others from both groups felt comfortable. This could be a hint of individual sensation on what type of chair is comfortable.

20. Were there any measuring devices that bothered you?

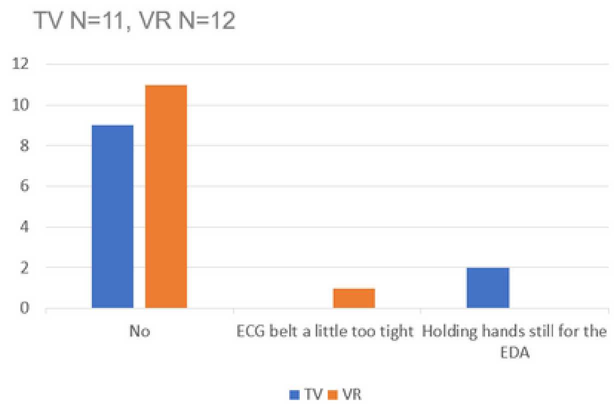


Figure 35: VR and TV group answers question 20 (Alka et al. 2022)

Most of the TV group subjects (9) were not bothered by any measuring devices. Only two stated that they had to focus on holding their hands still for the EDA measurements. Also most of the VR group participants (11) were not bothered by the devices, only one person said they put the ECG belt on a bit too tight. The subjects answered overly positive and were not disturbed by the devices. This could show that they felt no stress caused by the measuring devices.

21. Do you have any other thoughts on this experiment that you would like to share?

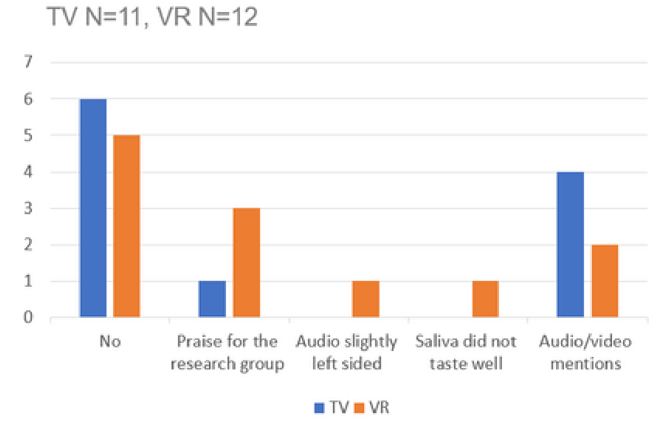


Figure 36: VR and TV group answers question 21 (Alka et al. 2022)

In the TV group 6 subjects did not have any additional thoughts. One person said the research group did well. Two criticized the scene (water texture could be better) and two said they rather would have tried the VR instead of TV.

In the VR group 5 participants did not have anything to add and three had praise for our group. Furthermore one subject mentioned the audio being “left sided” and one that the saliva for the cortisol samples did not taste well. One subject said that the scene was a little “shaky” and another one that the scene would have been more relaxing on TV (those have been summarized under Audio/Video mentions).

7 RESULTS & DISCUSSION

7.1.1 Questionnaires

7.1.1.5 Conclusion subjektive questionnaires

Through the demographic questionnaires we could show an even distribution between gender and age. The VR group had an equal distribution between men and women. TV group contains 72.73% male and 27.27% female subjects.
The age span goes from 19 to 32 and VR experience was evenly distributed.

Furthermore, the interpupillary distance was found to match the average.
The stereoscopic vision of subjects was tested via Randot test and ensured no irregularities in participants' abilities for three-dimensional vison occurred.

The semantic differentials were evaluated through application of the t-test. Thereby we could reject our hypothesis since the semantic differentials were not more negative after the VR scene and even state that stress levels went down more for the VR group compared to the TV group.

Possible motion sickness was captured through the Virtual Reality Sickness Questionnaire. We found the results to be even more positive (participants felt better) after the VR scene, while staying mostly the same for TV. Thus we have to reject the hypothesis of the VRSQ results being more negative after VR tests.

The Raw-TLX captures workload a given subject is experiencing. Subjects were asked to fill out the questionnaire after watching the relaxation scene, to give us an idea of how stressful they found the experiment.

Analyzing the data, we found that the hypothesis "the TLX results suggest that VR subjects felt more stressed after the test than TV subjects" can be rejected since VR subjects did not show signs of higher stress.

The goal of the qualitative questions was to create an opportunity for subjects to mention aspects about the shown scene, that were unnoticed by the research team. Therefore open questions were designed to be held in an interview to encourage even more insights. No significant irritation sourced were found. Also from the interviews a tendency towards VR subjects being more relaxed can be observed.

Overall the participants subjective perception seems to lean towards rating the VR experience more relaxing than TV. Hence the hypothesis B4 "The qualitative questions show negative reactions towards the VR scene but not towards the TV scene" must be rejected.

7.1.2 Cortisol

Cortisol is a low molecular weight steroid that acts as the most potent glucocorticoid in humans and by following its levels daily cyclic patterns, it has become an indicator used to study circadian rhythms. (cf. Seeley, Stephens, & Tate, 2003).

Cortisol levels can be found by analysis of several biological fluids: serum, plasma, urine, and saliva. Salivary analysis offers much flexibility given its noninvasive nature and can be easily performed by non-medical personnel. Salivary cortisol samples are also very stable and can be stored at 5°C for up to 3 months and at -20°C for up to 1 year. In addition, its freezing and/or thawing has no effect on the results (cf. Aardal & Holm, 1995; Garde & Hansen, 2005).

We have considered here the most common known problems to minimize the chances of affecting the accuracy of our research.

Preanalytic factors are directly related to the choices subjects make before the test. For this reason, the selection of experimental subjects was conducted carefully as described in Chapter 7.1. All those factors that could directly influence the correct sampling of the hormone should be excluded as far as possible. Potential sources of analytical error are represented by the sampling stage. Therefore, we used standard cortisol sampling pipettes. They come with a special sterile absorbent cylinder that experimental subjects can place between their cheek and teeth thus activating the salivary glands.

Sources of post-analytical error are mainly related to data management (e.g., data transfer to different file formats) and interpretation of results. In order to have a correct methodology, the tubes were kept stably in a domestic freezer at a temperature of about -18° fine at sedition to the analytical laboratory.

Description of salivary cortisol assay

Each of the 23 subjects were tested by salivary cortisol analysis. An initial detection was carried out at entry of the subjects into the experimental environment, a second test was performed before the medial content was presented, and a third immediately after the end of the media experience. The saliva samples were then stored at a temperature of ca. -18° Celsius until they were shipped to the DRESDEN LAB SERVICE GMBH Analysis laboratory. There the saliva samples were frozen and stored at -20°Celsius until analysis. After thawing, samples were centrifuged at 3,000 rpm for 5 min, which resulted in a clear supernatant of low viscosity. Salivary concentrations were measured using commercially available chemiluminescence immunoassay with high sensitivity (Tecan - IBL International, Hamburg, Germany; catalogue number R62111).

Intra- and inter-test coefficients of variation (CVs) have been used to estimate the accuracy of the salivary cortisol immunoassay (EIA). Both stood at 9% and therefore below the maximum acceptable level.

7 RESULTS & DISCUSSION

88% of the samples were taken in the afternoon between about 13:00 and 17:00. The remainder in the morning between about 09:00 and 12:00 noon.

Sample ID	1 Cortisol nmol/l	2 Cortisol nmol/l	3 Cortisol nmol/l
01VRE	2,14	3,26	2,68
02VRE	6,59	4,95	5,12
03VRE	11,04	13,60	9,37
04VRE	15,49	12,32	9,89
06VRE	19,94	4,18	3,77

Table 5: Absolute cortisol values VRE

Sample ID	1 Cortisol nmol/l	2 Cortisol nmol/l	3 Cortisol nmol/l
01VRU	4,60	4,30	4,13
02VRU	4,06	4,29	3,57
03VRU	3,47	5,80	5,46
04VRU	9,50	6,10	7,53
05VRU	10,53	9,15	9,69
06VRU	3,44	3,25	2,88
07VRU	2,96	4,06	4,26

Table 6: Absolute cortisol values VRU

Sample ID	1 Cortisol nmol/l	2 Cortisol nmol/l	3 Cortisol nmol/l
02TV	4,74	3,22	2,91
03TV	2,90	3,20	2,14
04TV	3,55	3,78	6,38
05TV	5,79	4,40	3,97
06TV	2,20	2,56	2,73
07TV	2,17	2,15	2,20
08TV	9,22	11,14	11,52
09TV	5,28	5,34	5,71
10TV	2,06	1,86	1,93
11TV1	2,07	1,67	1,85

Table 7: Absolute cortisol values TV

Result

After sorting the absolute data obtained from the laboratory we proceeded to calculate the difference between the various measurements. Since the hormone under consideration is subject to many physiological, physical, and psychological variations. At first it is essential to obtain a baseline for each subject and then calculate the variation. (Blair 20217)

The first measurement represents the baseline. The second measurement was taken immediately before the start of media enjoyment, thus after the experimental subjects wore all the measuring devices but without having worn (in the case of the VR groups only) the appropriate headset. The third measurement was taken immediately after the experiment and after the experimental subjects removed their headsets but with the biomedical monitoring equipment still in place. Since the baseline is the first measurement, we performed the second measurement to check and monitor any changes from the first analysis.

Sample ID	1 Cortisol nmol/l	2nd vs 1st	3rd Vs 2nd	3rd Vs 1st
01VRE	2,14	1,12	-0,58	0,54
02VRE	6,59	-1,64	0,17	-1,47
03VRE	11,04	2,56	-4,23	-1,67
04VRE	15,49	-3,17	-2,43	-5,60
06VRE	19,94	-15,76	-0,41	-16,17
Median		-1,64	-0,58	-1,67
Average		-3,38	-1,50	-4,87

Table 8: Differences in cortisol values VRE

Sample ID	1 Cortisol nmol/l	2nd vs 1st	3rd Vs 2nd	3rd Vs 1st
01VRU	4,60	-0,30	-0,17	-0,47
02VRU	4,06	0,23	-0,72	-0,49
03VRU	3,47	2,33	-0,34	1,99
04VRU	9,50	-3,40	1,43	-1,97
05VRU	10,53	-1,38	0,54	-0,84
06VRU	3,44	-0,19	-0,37	-0,56
07VRU	2,96	4,06	0,20	1,30
Median		-0,19	-0,17	-0,49
Average		0,19	0,08	-0,15

Table 9: Differences in cortisol values VRU

Sample ID	1 Cortisol nmol/l	2nd vs 1st	3rd Vs 2nd	3rd Vs 1st
02TV	4,74	-1,52	-0,31	-1,83
03TV	2,90	0,30	-1,06	-0,76
04TV	3,55	0,23	2,60	2,83
05TV	5,79	-1,39	-0,43	-1,82
06TV	2,20	0,36	0,17	0,53
07TV	2,17	-0,02	0,05	0,03
08TV	9,22	1,92	0,38	2,30
09TV	5,28	0,06	0,37	0,43
10TV	2,06	-0,20	0,07	-0,13
11TV1	2,07	-0,40	0,18	-0,22
Median		0,02	0,12	-0,05
Average		-0,07	0,20	0,16

Table 10: Differences in cortisol values TV

7 RESULTS & DISCUSSION

As depicted in the box plot in figure 36 the distribution of the change in cortisol levels between the first and second measurements is not homogeneous in all 3 groups tested. In the VRE group we see a general decrease in cortisol levels with peaks that suggest contamination of the salivary test of the 06VRE experimental subject. In fact, in this case there is an abnormal and rapid decrease of -15.76 nmol/l in a small time span of ca. 14 min. In addition, a change of this magnitude is not found in any other sample.

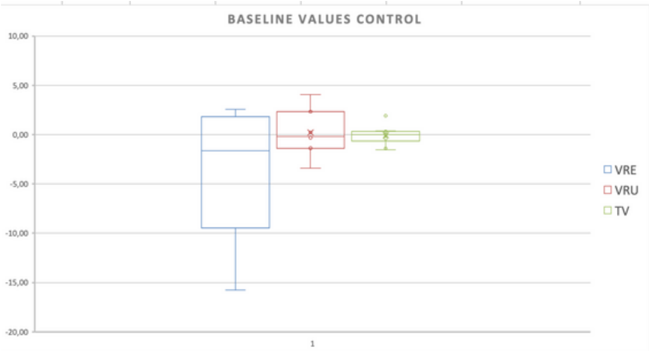


Figure 37: Distribution of differences in cortisol levels

Although changes in cortisol levels can be detected in the VRU and TV groups, they are minimal until approaching 0 in the TV group. In conclusion, we can claim that since the change in cortisol levels from the first measurement to the second measurement is not significant, we proceeded with the statistical analysis of the comparison of variance between the VR and TV groups in order to confute or reject Hypothesis B7: The subjects' cortisol values rose compared to the baseline measurement for VR subjects and remained unchanged for TV subjects.

For statistical evaluation, we used a T-test: Two-sample assuming unequal variances.
In both, P-Vaue in Figure 37 and 38 being far bigger than Alpha 5% established (VRU=69% and VRE=17%) the results are not statistically significant. However, we have some very interesting indications that force us to reflect. In the reference TV group for comparison with the two VR groups we see a slight increase in mean cortisol levels. They are, however, minimal. While analyzing the difference in the averages of variation we can see that there are significant differences in the decreases in cortisol levels in the two VR groups. The VRU group stands at -0.28 while the VRE group stands at -5.01. Here we can see that there is a substantial difference in the decrease of chortisol levels in the groups so selected based on the experience of the experimental subjects with VR technology.

T-test: Two-sample assuming unequal variances

	VRU 3rd Vs 1st	*TV 3rd Vs 1st*	*Difference in the averages of variation*
Media	-0,148571429	0,136	-0,28
Variance	1,809447619	2,320537778	
Observations	7	10	
Difference assumed for the averages	0		
gdl	14		
Stat t	-0,406304184		
P(T<=t) one-sample	0,345330854		
t critical one-sample	1,761310136		
P(T<=t) two-sample	69%		
t critical two-sample	2,144786688		

Figure 38: T-test VRU vs TV

T-test: Two-sample assuming unequal variances

	VRE 3rd Vs 1st	*TV 3rd Vs 1st*	*Difference in the averages of variation*
Media	-4,874	0,136	-5,01
Variance	44,82273	2,320537778	
Observations	5	10	
Difference assumed for the averages	0		
gdl	4		
Stat t	-1,652053405		
P(T<=t) one-sample	0,086933865		
t critical one-sample	2,131846786		
P(T<=t) two-sample	17%		
t critical two-sample	2,776445105		

Figure 39: T-test VRE vs TV

In conclusion we have to reject our hypothesis B7 since even though in the TV group we are in the presence of even minimal changes in average cortisol levels, in the two VR groups we see a decrease in average cortisol levels. Particularly significant is the decrease in levels in the VRE groups. This suggests that the experienced subjects not only do not seem to be adversely affected by the VR equipment but also seem to feel comfortable being able to achieve a relaxed state. The same phenomenon, even though in significantly lesser forms is also observed in the VRU group. Its difference can probably be associated with the experimental subjects' degree of confidence with the technology.

However, to crystallize the results, we invite repeating the experiment with the same number of subjects by implementing the hormone sampling techniques as described below. This would serve to confirm or disprove the phenomenon in this pilot study. However, for more representative results it is necessary to increase the experimental population.

7.1.3 ECG/EDA

This section analyzes the readouts and the analysis of the electrocardiographic (ECG) and electrodermal activity (EDA) sensors which are used in the experiment. The sensors were from the Movisens GmBh, which are well established in the area of biometric research. The ECG and EDA sensors are capable of recording various biometric variables. For this particular experiment the ECG is measuring the heart rate (HR) which is measured in bpm, and heart rate variability (HRV) which is measured in ms. The EDA is measuring the skin conductance level (SCL) in micro siemens. Every test starts with a three minute baseline measurement, which determines the heart rate, heart rate variability and skin conductance level baseline of each subject. After that the subject is guided to the test chair and equiped. Then a 5 minute measurement while viewing the item is performed.

7.1.3.1 ECG

It is expected that the heart rate declines, if the subject relaxes, and raises if the subject is under stress (cf. Taelman et al., 2009).
On the other hand the heart rate variability should be longer if the subject is relaxed and shorter if the subject is under stress. (Young Hwan Lee et al., 2017, p. 235-236)

Heart Rate

In figure 40 the heart rate of subjects of the TV group is shown, in comparison to their baseline, a stabilization of the heart rate can be observed.

7 RESULTS & DISCUSSION

In some cases it could also be observed that the heart rate decreases significantly between baseline and relaxation scene as there is a small timeframe inbetween baseline and relaxation scene where the subject is being prepared for the relaxation scene which is not representable.

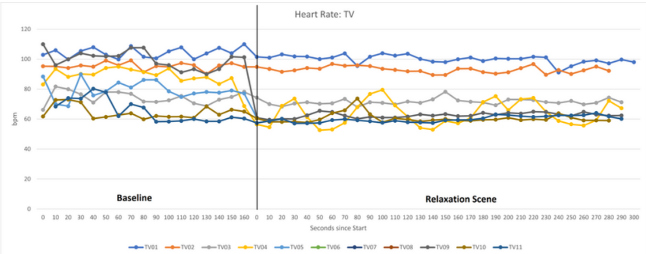


Figure 40: Heart rate: TV group (Alka et al., 2022)

Our results of 4 subjects of the VRE test (figure 41) show that the heart rate of baseline and scene are very similar. The differences are very minor, yet still show that the heart rate of all subjects is slightly lower or the same in the relaxation scene.

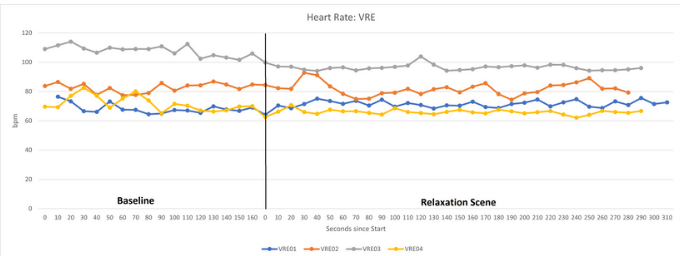


Figure 41: Heart rate: VRE group (Alka et al., 2022)

Our results of 5 subjects of the VRU Test (figure 42) shows similar results as the TV test, the heart rate in the test stabilizes in comparison. to the baseline.

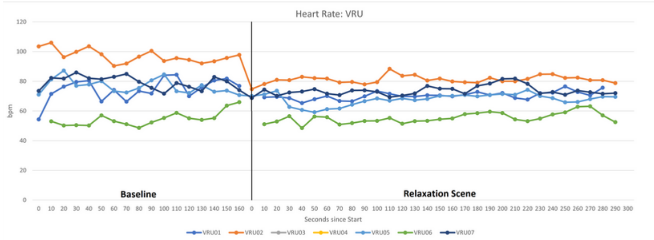


Figure 42: Heart rate: VRU group (Alka et al., 2022)

Notice:
Some EEG/EDA-samples have to be rejected from analysis because the data was obviously incorrect (e.g. by movements of the subjects) or recorded with severe dropouts.

7 RESULTS & DISCUSSION

Heart Rate Variability

As seen in figure 43 it was discovered that the HRV of most subjects stays stable in the TV Test when compared to the baseline. Although some subjects such as 04TV and 02TV show some irregular behavior in their baselines and therefore the baseline of those subjects is not representable compare with the relaxation scene measurement.

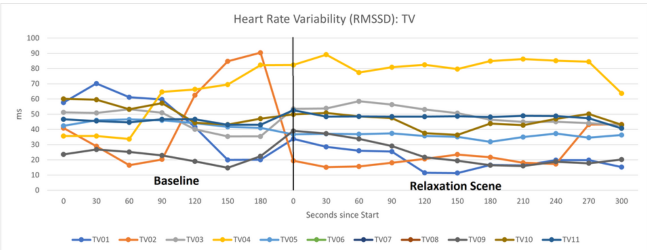


Figure 43: Heart rate variability: TV group (Alka et al., 2022)

As seen on figure 44 the HRV of the majority of the observed VRE subjects is stable along the baseline als well as relaxation scene measurement.

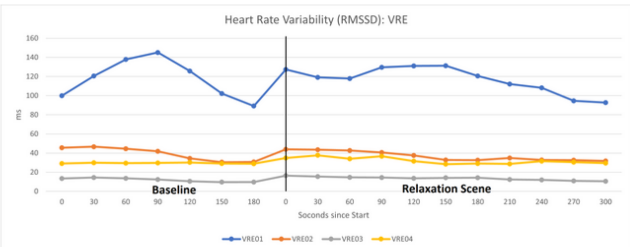


Figure 44: Heart rate variability: VRE group (Alka et al., 2022)

Looking at the VRU data in figure 45, a trend can be observed. The data shows that the HRV of 4 from 5 observed subjects is trending down.

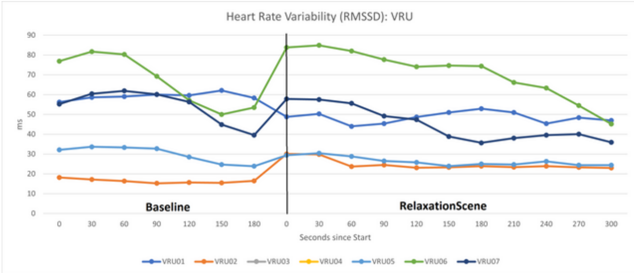


Figure 45: Heart rate variability: VRU group (Alka et al., 2022)

7.1.3.2 EDA

The EDA measures the skin conductance level (SCL). The SCL increases if the skin begins to sweat, which in itself can correlate to a signal of stress. Therefore an increase of the SCL correlates to Stress buildup. (cf. Bach et al., 2010)

Skin Conductance Level

In figure 46 we see the TV test. It shows a steady decline of the SCL on all observed subjects, it decreases below the baseline.

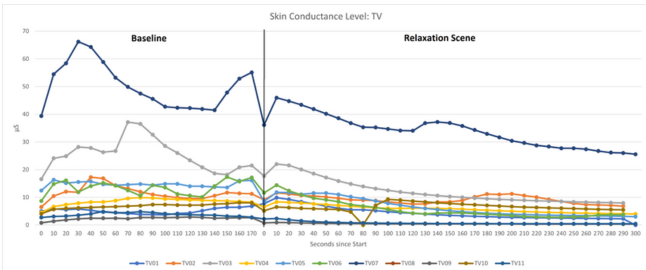


Figure 46: Skin conductance level: TV group (Alka et al., 2022)

7 RESULTS & DISCUSSION

In figure 47 a pattern can be observed on all observed subjects except 04VRE. It demonstrates a spike in between the baseline test and the VRE test. This phenomenon could be explained by the initial sweat production of the skin while the head gets embraced by the VR headset. Therefore it could be an artifact of these causes. 20 seconds after the start of the relaxation scene, the SCL starts to stabilize for most of the subjects and later falls below the baseline.

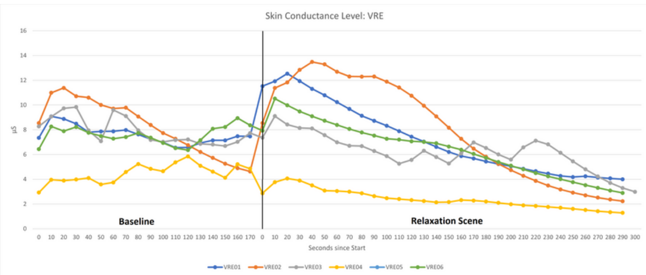


Figure 47: Skin conductance level: VRE group (Alka et al., 2022)

Like previously, a similar pattern between the end of the baseline and the start of the relaxation scene can be seen in figure 48 of the VRU test. It also should be noted that the pause in between baseline and relaxation scene also could be responsible for those drastic changes in the measurements therefore the first couple seconds of each relaxation scene should not be evaluable. But looking at the measurement of the relaxation scene again a trend can be seen, the SCL goes overall below the baseline.

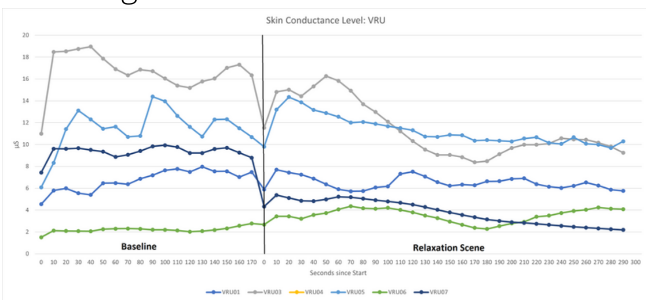


Figure 48: Skin conductance level: VRU group (Alka et al., 2022)

Heart Rate		TV01	TV02	TV03	TV04	TV05	TV09	TV10	TV11	VRU01	VRU02	VRU05	VRU06	VRU07	VRE01	VRE02	VRE03	VRE04
Baseline	Mean	104,35	95,45	74,40	88,34	79,50	99,72	64,47	65,49	74,61	97,04	76,57	54,65	79,26	68,33	82,59	107,90	71,82
	Standard Deviation	3,33	2,27	4,21	6,21	5,95	5,89	4,34	7,65	7,61	4,45	4,85	4,82	4,40	3,33	3,15	3,55	5,02
	Variance	10,46	4,86	16,65	36,26	33,34	32,64	17,75	54,90	54,57	18,62	22,15	21,79	18,22	10,38	9,36	11,83	23,74
	MEDIAN	103,98	95,36	74,75	89,30	78,53	101,25	62,70	61,57	75,90	96,24	75,44	53,59	79,99	67,38	83,71	109,01	69,95
Relaxation Scene	Mean	100,01	92,99	71,54	63,73	62,55	62,55	60,49	59,94	70,47	81,14	67,91	55,29	73,66	71,36	81,84	96,47	65,87
	Standard Deviation	2,73	2,12	2,01	8,17	1,60	1,60	3,21	2,07	2,54	2,57	3,78	3,47	3,26	2,43	4,43	2,03	1,68
	Variance	7,23	4,33	3,89	64,55	2,47	2,47	9,92	4,14	6,25	6,41	13,83	11,63	10,30	5,71	18,95	3,98	2,71
	Median	100,30	92,90	71,25	61,46	62,38	62,38	59,57	59,51	70,30	80,80	68,58	54,85	72,85	71,47	81,81	96,18	66,01
Baseline Corrected Mean		-4,34	-2,47	-2,86	-24,60	-16,95	-37,17	-3,98	-5,55	-4,15	-15,90	-8,65	0,64	-5,60	3,03	-0,75	-11,43	-5,94

Skin Conductance Level		TV01	TV02	TV03	TV04	TV05	TV06	TV07	TV09	TV10	TV11	VRU01	VRU02	VRU03	VRU05	VRU06	VRU07	VRE01	VRE02	VRE03	VRE04	VRE06
Baseline	Mean	5,09	11,67	25,50	8,46	14,74	13,47	50,24	2,38	6,83	3,78	6,73	12,13	16,64	11,44	2,19	9,32	7,58	8,21	7,91	4,48	7,58
	Standard Deviation	0,99	2,55	5,83	1,28	1,00	2,51	8,16	0,53	0,96	0,67	0,97	1,02	1,82	1,91	0,25	0,57	0,71	2,23	1,10	0,74	0,72
	Variance	0,93	6,16	32,13	1,54	0,94	5,93	62,96	0,26	0,87	0,43	0,89	0,98	3,12	3,44	0,06	0,31	0,48	4,69	1,14	0,52	0,49
	Median	5,05	11,36	25,44	8,76	14,76	14,05	48,85	2,49	7,06	3,66	6,95	12,17	16,78	11,55	2,18	9,45	7,48	8,47	7,48	4,60	7,62
Relaxation Scene	Mean	4,41	9,19	12,21	5,78	6,44	5,77	33,91	0,48	6,45	0,78	6,49	0,60	11,38	11,23	3,50	3,81	7,25	7,98	6,19	2,3927	6,4151
	Standard Deviation	2,26	1,42	4,32	1,30	3,19	3,25	5,73	0,16	1,67	0,52	0,56	0,07	2,48	1,27	0,61	1,10	2,86	4,04	1,43	0,7542	2,1539
	Variance	3,53	9,05	10,47	5,88	5,18	4,31	34,33	0,41	6,33	0,54	6,35	0,59	10,38	10,71	3,53	3,91	6,41	8,36	6,28	2,2643	6,7759
	Median	5,11	2,02	18,65	1,68	10,20	10,55	32,84	0,02	2,79	0,27	0,32	0,00	6,14	1,60	0,38	1,21	8,18	16,34	2,05	0,57	4,64
Baseline Corrected Mean		-0,68	-2,48	-13,30	-2,68	-8,30	-7,70	-16,33	-1,90	-0,39	-3,00	-0,24	-11,53	-5,26	-0,21	1,31	-5,51	-0,33	-0,23	-1,72	-2,087	-1,166

Heart Rate Variability		TV01	TV02	TV03	TV04	TV05	TV09	TV10	TV11	VRU01	VRU02	VRU05	VRU06	VRU07	VRE01	VRE02	VRE03	VRE04
Baseline	Mean	47,18	49,21	45,32	55,40	43,98	22,10	52,13	45,20	59,14	16,36	29,85	66,97	54,06	117,20	39,19	11,97	29,51
	Standard Deviation	20,41	30,35	7,98	19,90	2,29	4,03	7,22	1,64	1,79	1,03	4,19	13,32	8,58	20,80	7,08	2,01	0,49
	Variance	416,40	921,03	63,62	396,19	5,25	16,28	52,11	2,71	3,22	1,07	17,57	177,36	73,61	432,82	50,14	4,03	0,24
	Median	57,69	40,94	50,81	64,71	44,18	22,93	53,23	45,53	59,02	16,34	32,13	69,28	56,33	120,54	41,93	12,42	29,56
Relaxation Scene	Mean	20,42	23,29	49,95	81,51	35,82	24,53	45,22	48,11	48,43	24,71	26,11	70,97	45,06	114,18	36,93	13,58	32,09
	Standard Deviation	7,22	10,18	5,47	6,73	1,63	8,64	4,96	2,77	2,78	2,61	2,31	12,56	8,77	15,87	4,91	1,86	3,26
	Variance	52,07	103,54	29,90	45,32	2,65	74,68	24,58	7,69	7,74	6,83	5,32	157,87	76,88	251,91	24,14	3,46	10,66
	Median	19,79	19,39	50,73	82,46	36,29	20,30	46,90	48,47	48,70	23,65	25,38	74,34	40,00	118,48	34,97	14,17	31,55
Baseline Corrected Mean		-26,76	-25,92	4,63	26,11	-8,16	2,43	-6,91	2,91	-10,71	8,35	-3,74	4,00	-9,01	-3,02	-2,26	1,60	2,58

Table 11: Overview over all EDA/ECG data (Alka et al., 2022)

7.1.3.3 Overview EDA & EKG Data

Let's take a look at the baseline corrected averages which were calculated by subtracting the mean of the baseline from the mean of the relaxation scene. The correction allows us to compare the observed subjects amongst themselves and get an overview about the trends. By looking at table 11, figure 49, and the previous data from the tests it is clear that the SCL is trending down in correlation to the baseline therefore the hypothesis B6 seems to be false.

In figure 50, the previous data shows a similar picture: the heart rate of the baseline was higher than the mean of the relaxation scene therefore it's also trending down, which confirms that the hypothesis B5 is false. But when looking at the HRV in figure 51 the overall picture isn't completely clear as the values are overall balanced out as half of each group has 50% negative and 50% positive HRV differentials. But since no clear trend could be analyzed the conclusion would be that the HRV stays the same and therefore the hypothesis B5 would be also false.

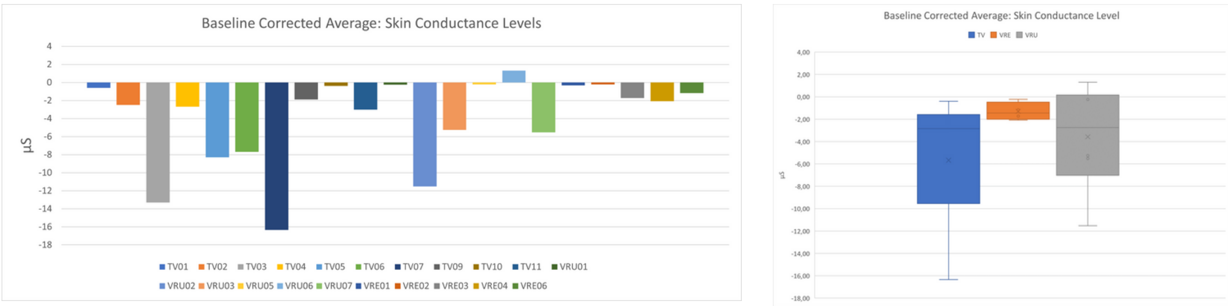


Figure 49: Baseline corrected average of skin conductance levels (Alka et al., 2022)

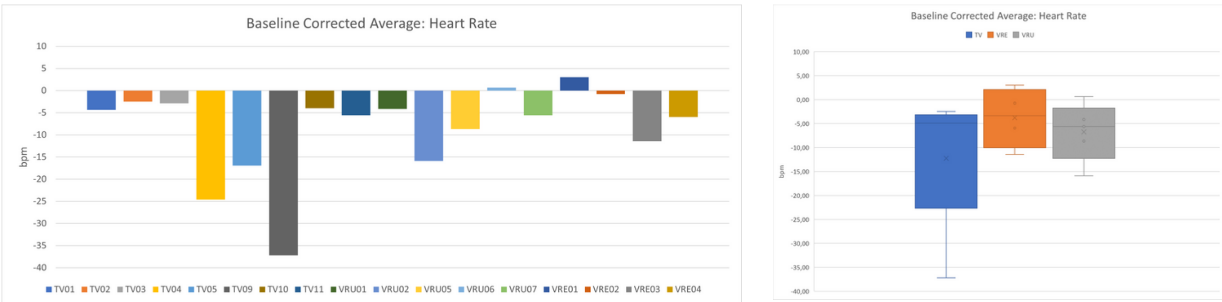


Figure 50: Baseline corrected average of heart rates (Alka et al., 2022)

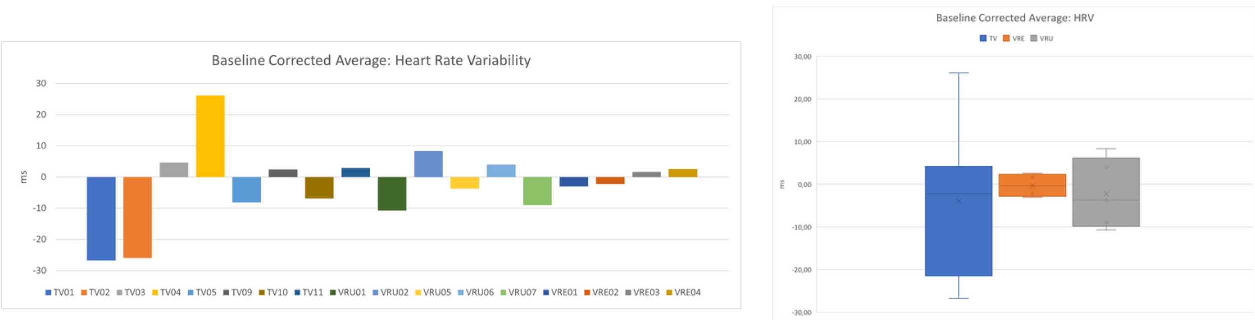


Figure 51: Baseline corrected average of heart rate variability (rMSSD) (Alka et al., 2022)

7.1.4 Headset sensors

This section presents an analysis of the data collected by the headsets (HP Reverb G2 Omnicept Edition) built-in sensors. These sensors include a heart rate sensor positioned at the wearer's forehead. While this data will not be used to evaluate the hypotheses, since the heart rate was already measured by an ECG, it was interesting to see for future projects if the headset's data compares to the ECG's data and if the headset's sensor could replace the cheststrap-ECG that was used for this experiment. Unfortunately there were several subjects where the headset did not measure any heart rate data at all. Additionally the data of some subjects contained too many errors to be usable (Figure 52, Figure 53). These errors could be caused by improper wearing of the headset (loose straps). For reliable data, an actual ECG seems to be still recommended.

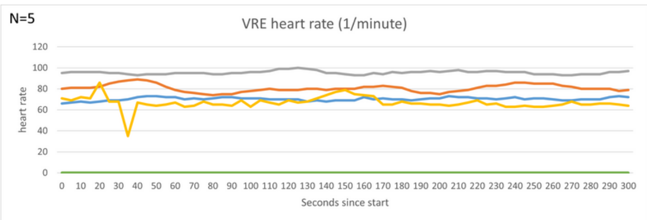


Figure 52: VRE heart rate (headset sensor) (Alka et al. 2022)

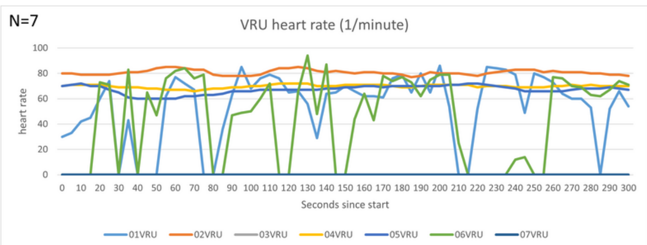


Figure 53: VRU heart rate (headset sensor) (Alka et al., 2022)

The cognitive load data the headset provided did not lead to new findings. The hypothesis was that subjects would start the experiment with a low cognitive load and that the stress caused by the VR scene would be made visible by a rising value. Instead the measured values ranged from low to high but not in a rising manner (see Figure 54, Figure 55). The issue with our experimental design and the cognitive load measurement is that it was developed for the purpose of evaluating the difficulty of a task while our subjects just passively watched the scene. What is also notable is that the cognitive load value is calculated based on data measured by different headset-sensors but as seen above, those did not always provide errorfree results and means that the cognitive load values of these subjects are incorrectly evaluated. For this reason all subjects where the measured heart rate was constantly zero (06VRE, 03VRU, 07VRU) or showed severe errors (04VRE, 01VRU, 06VRU) had to be excluded when analyzing cognitive load.

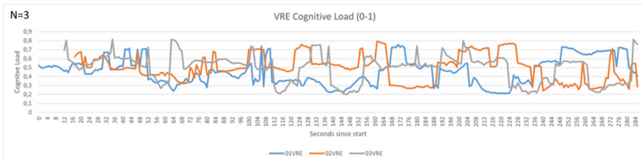


Figure 54: VRE cognitive load (Alka et al., 2022

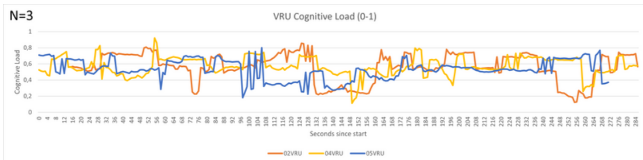


Figure 55: VRU cognitive load (Aslka et al., 2022)

7 RESULTS & DISCUSSION

When taking a look at the box plot diagrams (Figure 56) for each subject goint through the cognitive load analysis, the median values are all close to 0.5, with a slightly higher average median for the 3 subjects without prior VR experience.

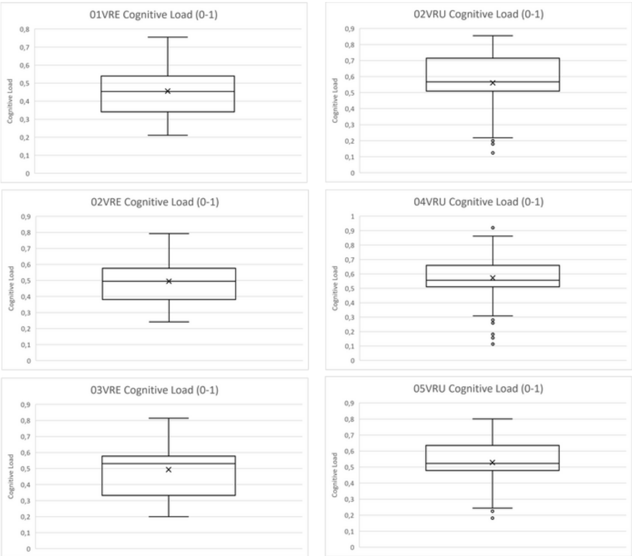


Figure 56: Cognitive load box plot (Alka et al., 2022)

While the median values point towards a cognitive load comparable to a task with medium difficulty, the data's trend lines do not show a rise in value, as stated in our hypotheses. Instead 5 of six subjects show a more or less decrease in cognitive load, suggesting that they became relaxed over time. (Figure 57)

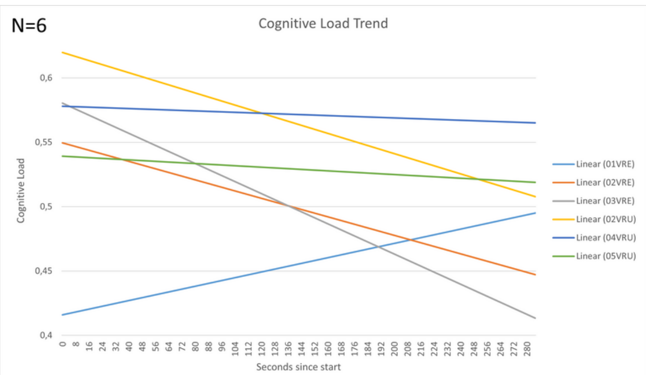


Figure 57: Cognitive load trend (Alka et al., 2022)

The pupil diameter data provided by the headset's built-in eye trackers did not show results that verify the hypothesis that the subjects' pupil diameters increased in size during the VR scene. As seen in the graphics below (Figure 58, Figure 59), the pupil diameters increased and decreased between several millimeters in size during the test. It must be noted that the scene the subjects viewed does not provide a constant brightness, especially for the VR subjects that were able to move their head to view different parts of the scene. This means that changes in pupil diameter are most likely linked to changing light conditions in the scene and not necessarily linked to the subject's stress level (see chapter 5.3)

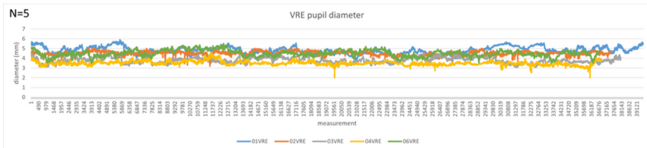


Figure 58: VRE pupil diameter (Alka et al., 2022)

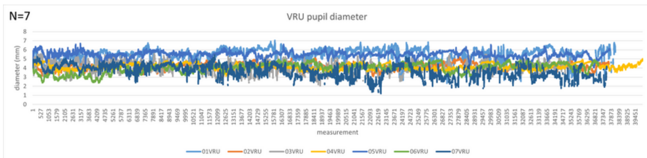


Figure 59: VRU pupil diameter (Alka et al., 2022)

That in mind, when viewing at the data's trend lines (Figure 60, Figure 61), all subjects but two show either a decline in pupil size or only little change. Since the scene does not include a decline in brightness this data does not indicate rising stress levels in the subjects and might even suggest that most became more relaxed over time.

7 RESULTS & DISCUSSION

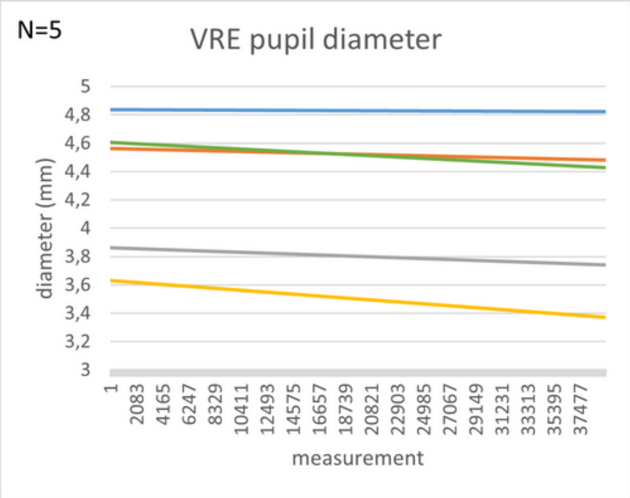


Figure 60: VRE pupil diameter trend (Alka et al., 2022)

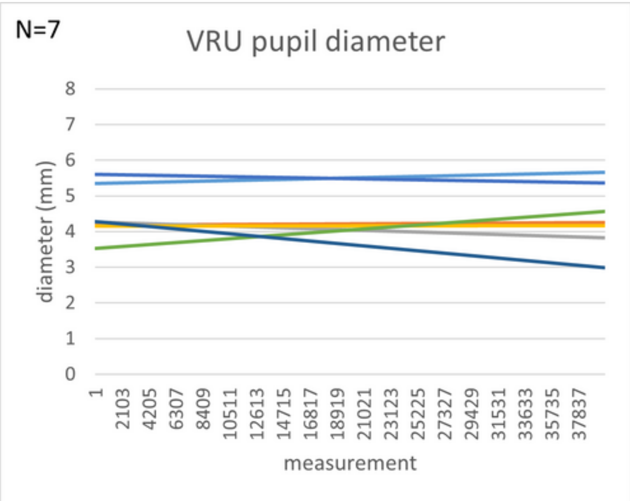


Figure 61: VRU pupil diameter trend (Alka et al., 2022)

When taking a look at the box plot diagrams (Figure 62), most subject's pupil diameters show a relatively small interquartile range. The variations in this range might be caused by changes in scene brightness trough eye- or head-movements.

Several subjects also show a lot of outliers and a large range between minimum and maximum values. Without data on the actual brightness they were exposed to, it is not possible to say how many of these values are errors.

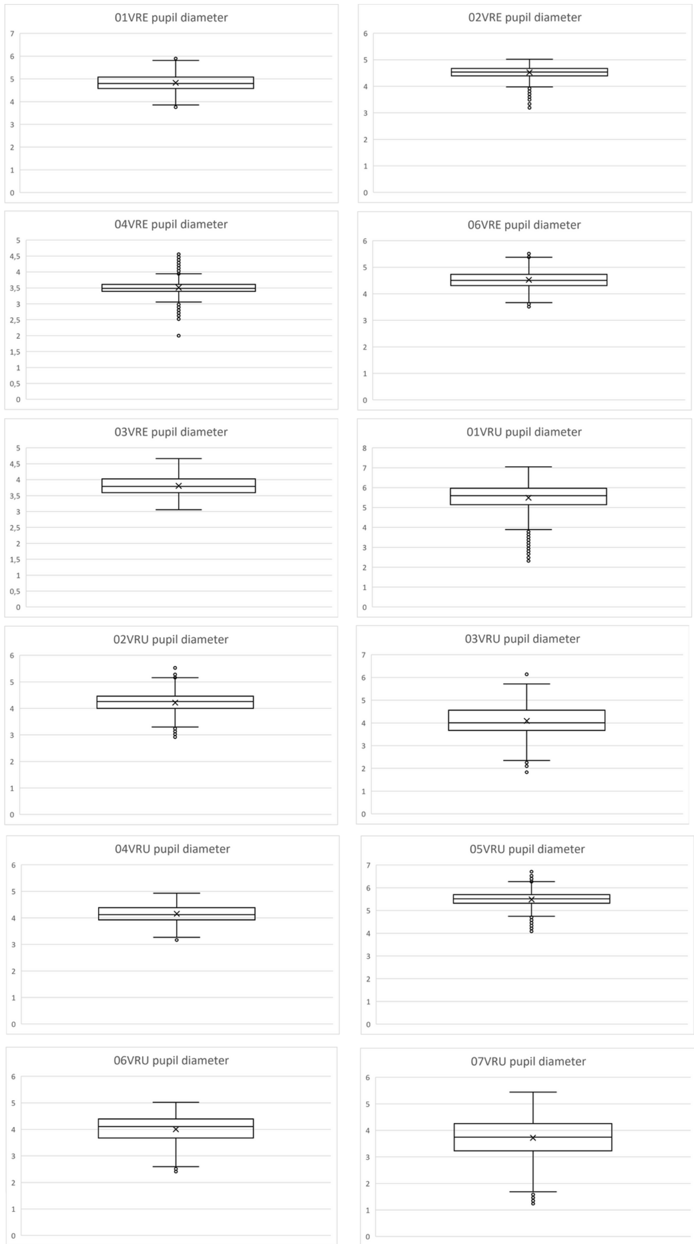


Figure 62: Pupil diameter box plot (Alka et al., 2022)

7 RESULTS & DISCUSSION

7.2 HYPOTHESES SUMMARY

Hypotheses A	Verified?
The subject groups were comparable in terms of age, gender distribution, eyesight and VR experience.	YES
The interpupillary distance (IPD) of the VR test subjects was typical for age and gender and was within the technical IPD possibilities of the VR headset.	YES
The duration of the tests was comparable.	YES
The subjects started the test in a state of relaxation or low emotional arousal.	NO
The experimental setup had no negative influence on the subjects.	YES
It was possible to identify every phase of the test in the collected data.	YES
Conflicts between reality and virtuality could be avoided/ minimized in the experimental setup.	YES
Content-related stress could be excluded from the experimental setup.	YES

Hypotheses B	Verified?
The semantic differentials were more negative after VR tests and unchanged after TV tests.	NO
The VRSQ results were more negative after VR tests and unchanged after TV tests.	NO
The Raw-TLX results indicate that VR subjects felt more stressed after the test than TV subjects.	NO
The qualitative questions show negative reactions towards the VR scene but not towards the TV scene.	NO

7 RESULTS & DISCUSSION

The subjects' heart rates rose compared to the baseline measurement for VR subjects and stayed unchanged for TV subjects.	**NO**
The subject skin conductance level rose compared to the baseline measurement for VR subjects and stayed unchanged for TV subjects.	**NO**
The subjects' cortisol values rose compared to the baseline measurement for VR subjects and stayed unchanged for TV subjects.	**NO**
The subjects heart rate variability declined compared to the baseline measurement for VR subjects and stayed the same for TV subjects.	**NO**
The subjects' pupil diameters increased in size during the VR scene.	**NO**
The subjects' cognitive load values rose during the VR scene.	**NO**

Looking at the tables above it becomes clear that our base hypothesis "Equipment-related stress in VR-headsets does exist" could not be confirmed in our pilot study of N=12 subjects. Some improvements could be made in future to test setup and the number of subjects should be tripled to achieve precise results. But so far we found no indication in this pilot study that one VR- headset by itself could cause stress for shortterm use by young adults.

7.3 MEASURED DATA VS. EXISTING LITERATURE

To verify the accuracy of our accumulated data, the values were compared to bench-marks from scientific literature.

Cortisol levels of the samples we observed were within the ranges identified in a number of previous studies. (cf. Ljubijankić, 2008) In the morning (8 - 9 a.m.) 3.5 - 27.0 nmol/l and in the afternoon (4 - 5 p.m.) 1.3 - 6.0 nmol/l. Since the hormone is subject to many variations throughout the day and susceptible to variations based on psycho-somatic conditions of the subjects and reactions with substances ingested by the subjects, the parameters should be interpreted in a nonrigid manner. As 88% of our samples were carried out between 1 p.m. and 5 p.m. and the remainder between 09 a.m. and 12 noon, we can confirm that the data are within the indicated ranges.

In terms of heart rate the "normal resting heart rate for adults ranges from 60 to 100 beats per minute" (Laskowski, 2020). Since the mean heart rate of all but one TV subject (TV01 with a mean of 103 bpm) is in this range, the results would be considered to be verifiable.

When it comes to the HRV an age focused study shows, that the average HRV (rMSSD) of a 20-29 year old averages around 43 ms with a standard deviation of 19 (cf. Umetani et al., 1998, p.597).

These results match with our findings as the mean of all subjects in all tests in the relaxation scene timeframe is 43,58 ms with a standart deviation of 6,03. The Baseline is higher with 46.16 ms but also with a wider standard deviation 9,01.

The Mean SCL over all subjects were 7,27 µS with an standard deviation of 2. This value is quite low when compared to other studies as the mean SCL in a study about SCL changes in working tasks showed that the mean SCL when recovering or on low demanding Tasks were around 15 to 16 µS with an standard deviation of 7 to 8. Although our findings were far lower than the study mentioned, the task was less demanding, which would explain the lower SCL (cf. Mehler et al., 2012, p.405).

“The normal pupil size in adults varies from 2 to 4 mm in diameter in bright light to 4 to 8 mm in the dark.” (Spector, 1990) The median pupil diameters of our subjects were all between 3 and 5 mm, and while the values never exceeded 8 mm they sometimes dropped below 2 mm, suggesting that these values were probably measuring errors. Considering the brightness of the item used in the experiment, the measured diameters seem appropriate.

8 SUMMARY & OUTLOOK

With the growing importance in the field of virtual reality, the demand for fundamental scientific knowledge in this research area has increased. This also includes research into stress when using VR equipment. Accordingly, the HuFacin_VR project has dedicated itself to the task of creating a basis for stress research on the subject of VR by developing a taxonomy of possibele VR stressors and firstly researching headset-related stress. The VR headset used was an HP Reverb G2 Omnicept Edition.

Measured data and analysis
The experiment was carried out with a total of 23 subjects. 11 of those viewed the same content in a TV seating-box for comparison with the 12 subjects that viewed the same content with a VR headset.

Subjective and objective data has been collected during the tests and analyzed to find indications if our experimental setup caused equipment-related stress in subjects. The subjects were comparable in demographic data (age, gender, etc.). The semantic differentials, VRSQ and Raw-TLX questionnaires did not show statistically relevant differences when comparing the data from before and after a test or between VR and TV subjects. None of the questionnaire-related hypotheses in terms of stress could be verified.

The qualitative questions could not reveal any major irritations regarding aur relaxation scene. Furthermore VR subjects often showed more tendencies of relaxation after the VR scene compared to the TV group.

Alongside quantitative data was collected by measurements. As described in detail in Section 7.1.2 we reject Hypothesis B7 (The subjects' cortisol values rose compared to the baseline measurement for VR subjects and stayed unchanged for TV subjects.). That proves the data compared with the t-test: two samples assuming different variances between VRs and TV groups are not statistically significant, we can record differences in the groups. In both tests the p values are well above the 5% threshold. However, we can record a tendency of decreasing cortisol levels in VR groups. Particularly in the VRE group as a result of what we assume may be a correlation between mastery and frequency of use of the VR equipment (criteria by which experimental subjects were defined as experts) and a relaxed state. The TV group shows results close to 0 tending to increase. The EDA’s skin conductance data and EKG’s heart rate data did not show an increase of stress for the subjects. The data even indicates a trend towards more relaxation. The heart rate variability of observed subjects did not show changes from baseline to relaxation scene measurement and therefore does not confirm the hypothesis.

The headset’s sensor data did not contain any indication that the experiment caused stress in subjects. Data even suggests that subjects became more relaxed over time. A big issue when working with the headset sensors is, that it is not possible to take baseline measurements without having the subject wear the VR headset.

8 SUMMARY & OUTLOOK

When researching the effects of VR headsets on users, this would defeat the purpose of a basline measurement because subjects would already be exposed to the observed factors.

Hurdles

During the project hurdles arose that had to be overcome or solved in order to carry out the test and evaluate measured data. Some of these problems can be concluded from the results of the test series, others give an outlook on what must be considered in future test series in the field of stress measurement in VR. Examples of these hurdles can be seen in the experience of the team members, the time pressure and the location itself. Research into equipment-related stress required good project and time management in order to quickly and intensively work up the missing knowledge about the biological basics, for example the definition and delimitation of the term stress, the rough chemical and technical basics and research-related basics. The technical aspect included operating the measuring devices, reading out the data, handling the VR headset and creating a suitable relaxation scene in Unity. The creation of this scene took a long time, especially since the assets had to be carefully and sensibly integrated. Further time challenges arose from the search for subjects and the associated advertising for the tests, the search for the laboratory and the procurement of the material, e.g. electrodes and measuring devices.

During the subject tests, it became apparent that not all subjects were suitable for the tests, despite the general conditions defined in advance. Some subjects had a too small interpupillary distance for the VR headset's technical limitations, which is why an optimal calibration of the VR headset could not be completed. Due to this fact, a few subjects' data were omitted here, other data were incorrect due to the measuring devices during use, for example the ECG stopped working and could therefore not be used in some tests.

Furthermore, not all of the registered subjects showed up, which meant that appointment slots were occupied, but could not be used. Only rarely was a substitute found for the test subject who did not appear. In general, participation in the experiment was very low, despite sufficient advertising verbally in lectures and at events, by email, in messengers such as WhatsApp or Discord and by posters. There was also a corresponding motivation, an Amazon voucher, as a reward. The subjects who came to the appointment stated that the voucher was partly a motivation to participate. The target group of the subject tests included the students who corresponded to the general conditions. Although the Furtwangen location is the place of study for our target group at Furtwangen University, it can be assumed that the majority of students do not stay in town throughout the semester.

8 SUMMARY & OUTLOOK

For this reason, there are only certain periods of time - outside the lecture-free period and only during the week - when subject tests could be carried out. Another hurdle was the free appointmentfinding tool. It was not sufficiently userfriendly, which led to many emails and personal inquiries and took up time.

Improvements

There are various approaches optimizing the test series that was carried out within this project.

For more meaningful results of the equipment-related stress measurement, more subjects would have to take part in the experiment. This could be achieved through various measures such as a change of location for the trials so that the number of potential participants increases or a change in the experiment period. The experiments tend to be better booked at the beginning of the semester, far away from the exams period and also more in the winter semester - since there are fewer public holidays in Baden-Württemberg in winter. This would also relieve time management and make the evaluation phase easier to plan, so that spontaneous and short-term potential test subjects can still be admitted and taken into account. The laboratory also needs a corresponding amount of time to evaluate cortisol results and communicate them to the project team. This in turn takes time to process and interpret the results.

In addition to the location problem, a more infrastructure- and user-friendly access to register for the test series would also be advantageous. At this point, switching to an appointment tool that meets the following requirements should be considered:

- A description of the test series and the general conditions must be included in the tool.
- The subjects can easily and without problems enter their data in the tool.
- Questions can be sent directly by email to the appointment creator.
- The subjects and the appointment creator must receive a confirmation email when booking or canceling an appointment.
- A reminder email should go out to the subjects shortly before the appointment so that they do not miss it.
- It must be possible to book and cancel the tool exactly for the project period.

Once the above requirements for planning, room reservation, procurement of a set of measuring devices and finding test subjects have been met, it is possible to start with the actual test series. Due to the sensitivity of the measuring devices and the feasibility with meaningful data, there should always be sufficient material on site: For example, enough electrodes, data protection declarations for the subjects, replacement devices if a device fails and material to ensure hygiene (sanitary towels, disinfectant spray,...).

Further outlook

In summary, a further expansion of this research area - the measurement of stress through conflicts between virtuality and reality, content-related and equipment-related stress - offers a good opportunity to fathom human limitations and overcomings physically and psychologically. This newly gained knowledge would be useful for upgrading the VR equipment. It would also allow for better classification of usage time in virtual reality in different workspaces. This way, exercise units in the field of medicine in the training of prospective doctors could be optimized in terms of time during patient consultations. Or the headsets are adapted more dynamically to head shapes or are lighter in weight.

With the help of optimization the HuFacin_VR project experiment, it would be able to carry it out again with a larger group of test subjects. Subsequently, the other two stress-inducing groups - those of content-related stress and stress due to conflicts between virtuality and reality - should also be further examined.

LITERATURE

Acharya, U. R., Joseph, P., Kannathal, N., Lim, C., Suri, J. (2007). Heart rate variability: A review. Medical & biological engineering & computing. 44. 1031-51. 10.1007/s11517-006-0119-0.

Alka, D. M., Kahn, L., Keck, L. D., Nechypurenko, J., Orengo, I., Schürmann, M. (2022). Bericht HuFacin_VR: Human Factors in Virtual Realities. Hochschule Furtwangen University, (p. 21-24).

Annerstedt, M., Jönsson, P., Wallergård, M., Johansson, G., Karlson, B., Grahn, P., Währborg, P. (2013). Inducing physiological stress recovery with sounds of nature in a virtual reality forest—Results from a pilot study. Physiology & behavior, 118, 240-250.

Bach, D. R., Friston, K. J., Dolan, R. J. (2010). Analytic measures for quantification of arousal from spontaneous skin conductance fluctuations, International Journal of Psychophysiology, Volume 76, Issue 1, 2010, https://doi.org/10.1016/j.ijpsycho.2010.01.011.

Barrett, L.F., Russell, J.A. (1999). The structure of current affect: Controversies and emerging consensus. Curr. Dir. Psychol. Sci.

Blair, J., Adaway, J., Keevil, B., & Ross, R. (2017). Salivary cortisol and cortisone in the clinical setting. Current Opinion in Endocrinology, Diabetes and Obesity, 24(3), 161-168.

Choe, E. Y., Jorgensen, A., Sheffield, D. (2020). Simulated natural environments bolster the effectiveness of a mindfulness programme: A comparison with a relaxation-based intervention. Journal of Environmental Psychology, 67, 101382.

Christianson, S. (1992). A. Emotional stress and eyewitness memory: A critical review. Psychol. Bull.

Clement, J. (2022). Global virtual reality gaming revenue 2017-2024. https://www.statista.com/statistics/499714/global-virtual-reality-gaming-sales-revenue/. Accessed 25 July 2022.

Cohen, S., Gianaros, P. J., Manuck, S. B. (2016). A stage model of stress and disease. Perspectives on Psychological Science, 11(4), 456-463.

Collet, C., et al. (2021). Driver Behavior and Performance in an Age of Increasingly Instrumented Vehicles. Frontiers Media SA, Google Books, https://www.google.de/books/edition/Driver_Behavior_and_Performance_in_an_Ag/NFpFEAAQBAJ?hl=de&gbpv=0. Accessed 3 April 2022. (P.147).

D'Amelio, R. (2009). Studienbrief: Entspannungsverfahren. Universitätskliniken des Saarlandes.

Davis, S., Nesbitt, K., Nalivaiko, E. (2014). A systematic review of cybersickness. In Proceedings of the 2014 conference on interactive entertainment (pp. 1-9).

Dennison, M. S., Wisti, A. Z., D'Zmura, M. (2016). Use of physiological signals to predict cybersickness. Displays, 44, 42-52.

Dodgson, Neil. (2004). Variation and extrema of human interpupillary distance. Stereoscopic Displays and Virtual Reality Syst XI. 5291. 36-46. 10.1117/12.529999.

Hanneman, S. K., Cox, C. D., Green, K. E., & Kang, D. H. (2011). Estimating intra-and inter-assay variability in salivary cortisol. Biological research for nursing, 13(3), 243-250.

Hofmann, S. M., Klotzsche, F., Mariola, A., Nikulin, V. V., Villringer, A., Gaebler, M. (2020). Decoding subjective emotional arousal from EEG during an immersive Virtual Reality experience, posted November 10, 2020

Hottong, N. (2021). S3D-Content Produktion WS21/22. Hochschule Furtwangen/ Fakultät Digitale Medien.

Jyotsna, C. & Amudha, J. (2018). Eye Gaze as an Indicator for Stress Level Analysis in Students, 2018 International Conference on Advances in Computing, Communications and Informatics (ICACCI), 2018, pp. 1588-1593, doi: 10.1109/ICACCI.2018.8554715.

LITERATURE

Kemeny, A., Chardonnet, J.-R., Colombet, F. (2020). Getting Rid of Cybersickness in Virtual Reality, Augmented Reality, and Simulators. Arts et Métiers Institute of Technology, Hochschule Frankreich, Springer Nature Switzerland AG.

Kim, H. K., Park, J., Choi, Y., Choe, M. (2018) .Virtual reality sickness questionnaire (VRSQ): Motion sickness measurement index in a virtual reality environment, Applied Ergonomics, Volume 69, Pages 66-73, ISSN 0003-6870, https://doi.org/10.1016/j.apergo.2017.12.016

Laskowski, E. R. (2020). What's a normal resting heart rate?. https://www.mayoclinic.org/healthy-lifestyle/fitness/expert-answers/heart-rate/faq-20057979. Accessed 22 July 2022.

Lenscan Medical Inc. (2012). Randot Stereotests https://lenscanmed.com/wp-content/uploads/2020/11/Randot-stereotests-Manual-1.pdf

Lewis, P.A., Critchley, H.D., Rotshtein, P., Dolan, R.J. (2007). Neural correlates of processing valence and arousal in affective words. Cereb. Cortex.

Ljubijankić, N., Popović-Javorić, R., Šćeta, S., Šapčanin, A., Tahirović, I., & Sofić, E. (2008). Daily fluctuation of cortisol in the saliva and serum of healthy persons. Bosnian journal of basic medical sciences, 8(2), 110.

Mayo Clinic Staff (2021) Stress and high blood pressure: What's the connection?. https://www.mayoclinic.org/diseases-conditions/high-blood-pressure/in-depth/stress-and-high-blood-pressure/art-20044190. Accessed 26 July 2022.

Mehler, B., Reimer, B., Coughlin, J. (2012). Sensitivity of Physiological Measures for Detecting Systematic Variations in Cognitive Demand From a Working Memory Task: An On-Road Study Across Three Age Groups. Human factors. 54. 396-412. 10.1177/0018720812442086. p.405

Mon-Williams M., Warm J. P., Rushton S. (1993). Binocular vision in a virtual world: visual deficits following the wearing of a head-mounted display. Ophthalmic and Physiological Optics 13:387– 391. https://doi.org/10.1111/j.1475-1313.1993.tb00496.x.

Moore, R. K. (2012). A Bayesian explanation of the 'Uncanny Valley' effect and related psychological phenomena. Scientific reports, 2(1), 1-5.

Optik Peter Meyer (n.d.). Anleitung Augenabstand (PD) messen. https://shop.optikmeyer.de/custom/plugins/BeemoOptician/Resources/frontend/_public/anleitung_pd.pdf. Accessed 29 July 2022.

Perogamvros, I., Owen, L. J., Newell-Price, J., Ray, D. W., Trainer, P. J., Keevil, Brian J. (2009). Simultaneous measurement of cortisol and cortisone in human saliva using liquid chromatography-tandem mass spectrometry: application in basal and stimulated conditions, J Chromatogr B Analyt Technol Biomed Life Sci. 2009 Nov 1;877(29):3771-5. Epub 2009 Sep 18.

Pfaff, D. W. (2006). Brain arousal and information theory: Neural and genetic mechanisms. Cambridge, Mass.: Harvard University Press, p. 21

Pointer JS. The far interpupillary distance. A gender-specific variation with advancing age. Ophthalmic Physiol Opt. 1999 Jul;19(4):317-26. doi: 10.1046/j.1475-1313.1999.00441.x. PMID: 10645388.

Pourhoseingholi M. A., Baghestani A. R., Vahedi M. (2012). How to control confounding effects by statistical analysis. Gastroenterol Hepatol Bed Bench. 5(2):79-83. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4017459/.

Renkl A., Paas F., Sweller J. (2016) Cognitive Load Theory: A Special Issue of Educational Psychologist. (2016). Taylor & Francis.

Rodrigues, J., Studer, E., Streuber, S., Meyer, N., & Sandi, C. (2020). Locomotion in virtual environments predicts cardiovascular responsiveness to subsequent stressful challenges. Nature communications, 11(1), 1-11.

LITERATURE

Russell, J. A. (1980). Journal of Personality and Social Psychology – A Circumplex Model of Affect, American Psychological Association, p. 1167

Schandry, R. (2016). Biologische Psychologie: mit Online-Material, 4., überarbeitete Auflage. Beltz, Weinheim Basel.

Seyama, J. I., & Nagayama, R. S. (2007). The uncanny valley: Effect of realism on the impression of artificial human faces. Presence, 16(4), 337-351.

Seyle, H. (1981). Geschichte und Grundzüge des Streßkonzepts. In: Nitsch JR, Allmer H (eds) Stress: Theorien, Untersuchungen, Massnahmen. H. Huber, Bern, pp 163–187.

Sherman, William R., editor (2019). VR Developer Gems. CRC Press. Google Books, https://www.google.de/books/edition/VR_Developer_Gems/zAWcDwAAQBAJ?hl=de&gbpv=. Accessed 5 April 2022.

Siegel, E. H., Wei, J., Gomes, A., Oliviera, M., Sundaramoorthy, P., Smathers, K., ... & Ballagas, R. (2021). HP Omnicept Cognitive Load Database (HPO-CLD)–Developing a Multimodal Inference Engine for Detecting Real-time Mental Workload in VR. Technical report, HP Labs, Palo Alto.

Snider, James G. and Charles E. Osgood (1969). Semantic differential technique : a sourcebook.

So, P., Gore, B. (2020). NASA TLX: TASK LOAD INDEX. https://humansystems.arc.nasa.gov/groups/TLX/. Accessed 26 July 2022.

Spector, R. H. (1990) The Pupils. In: Walker HK, Hall WD, Hurst JW, editors. Clinical Methods: The History, Physical, and Laboratory Examinations. 3rd edition. Boston: Butterworths. Chapter 58. Available from: https://www.ncbi.nlm.nih.gov/books/NBK381/. Accessed 20 July 2022.

Sproll, D., Freiberg, J., Grechkin, T., Riecke, B.E. (2013). Poster: Paving the way into virtual reality - A transition in five stages. 2013 IEEE Symposium on 3D User Interfaces (3DUI), 175-176.

Stereo Optical Company Inc. (2017). Instructions and Answer Key for the Distance Randot Stereotest https://www.stereooptical.com/wp-content/uploads/2017/10/Distance-Randot-Instructions-07-2017.pdf

Stokes, A. F., & Kite, K. (2001). On grasping a nettle and becoming emotional. In P.A. Hancock, & P.A. Desmond (Eds.), Stress, workload, and fatigue. Mahwah, NJ: L. Erlbaum, p. 530

Taelman, J., Vandeput, S., Spaepen, A., Van Huffel, S. (2009). Influence of Mental Stress on Heart Rate and Heart Rate Variability. In: Vander Sloten, J., Verdonck, P., Nyssen, M., Haueisen, J. (eds) 4th European Conference of the International Federation for Medical and Biological Engineering. IFMBE Proceedings, vol 22. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-540-89208-3_324. (p. 1368).

Takai, N., Yamaguchi, M., Aragaki, T., Eto, K., Uchihashi, K., Nishikawa, Y. (2004). Effect of psychological stress on the salivary cortisol and amylase levels in healthy young adults, Archives of Oral Biology, Volume 49, Issue 12, Pages 963-968, ISSN 0003-9969, https://doi.org/10.1016/j.archoralbio.2004.06.007.

Tran, T.-D., Kim, J., Ho, N.-H., Yang, H.-J., Pant, S., Kim, S.-H., Lee, G.-S. (2021). Stress Analysis with Dimensions of Valence and Arousal in the Wild, Applied Science, p.1

Umetani, K., Singer, D. H., McCraty, R., Atkinson, M. (1998). Twenty-Four Hour Time Domain Heart Rate Variability and Heart Rate: Relations to Age and Gender Over Nine Decades,
Journal of the American College of Cardiology, Volume 31, Issue 3, Pages 593-601, ISSN 0735-1097, https://doi.org/10.1016/S0735-1097(97)00554-8. p.597

LITERATURE

University of Nevada (n.d.).Releasing stress through the power of music. https://www.unr.edu/counseling/virtual-relaxation-room/releasing-stress-through-the-power-of-music. Accessed 10 July 2022.

Wang X., Shi Y., Zhang B., Chiang Y. (2019). The influence of forest resting environments on stress using virtual reality.International journal of environmental research and public health 16(18):3263. doi:https://doi.org/10.3390/ijerph16183263.

Wang, Z., He, R., & Chen, K. (2020). Thermal comfort and virtual reality headsets. Applied Ergonomics, 85, 103066.

Wichary, S., Mata, R., Rieskamp, J. (2016). Probabilistic inferences under emotional stress: How arousal affects decision processes. J. Behav. Decis. Mak.

Yamanaka, K. & Kawakami, M. (2009). Convenient Evaluation of Mental Stress With Pupil Diameter, International Journal of Occupational Safety and Ergonomics, 15:4, 447-450, DOI: 10.1080/10803548.2009.11076824, (p. 449).

Yan, Y., Chen, K., Xie, Y., Song, Y., & Liu, Y. (2018). The effects of weight on comfort of virtual reality devices. In International Conference on Applied Human Factors and Ergonomics (pp. 239-248). Springer, Cham.

Yoo , K. & Lee, W. (2011). Mental stress assessment based on pulse photoplethysmography, 2011 IEEE 15th International Symposium on Consumer Electronics (ISCE), pp. 323-326, doi: 10.1109/ISCE.2011.5973841.

Young Hwan Lee, Bon-Hoon Koo, Dai-Seg Bai, Eun-Jin Cheon, & Hye-Geum Kim. (2017). *Stress and Heart Rate Variability: A Meta-Analysis and Review of the Literature*. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5900369/pdf/pi-2017-08-17.pdf, p. 235, 236
Accessed 24 July 2022.

Zheng R. Z. (2017) Cognitive Load Measurement and Application: A Theoretical Framework for Meaningful Research and Practice. Taylor & Francis. United Kingdom

Die Macht der Farben – Farb- und Lichttherapie als Heilmittel. (n.d.). Oberberg Kliniken. https://www.oberbergkliniken.de/artikel/die-macht-der-farben#c5144. Accessed 16 July 2022.

IMAGES

Own Fotos

A lot of images (except own figures and fotos) used in the project report are subject to the Creative Commons License from unsplash.com and can therefore be used for personal and commercial purposes without naming the author.

unsplash.com fotos:

Minh Pham (2020), unsplash.com, creative Common Licence,
https://humansystems.arc.nasa.gov/groups/TLX/. Accessed 26 July 2022.

...

ATTACHMENT 1

SUBJECT TEST MANUAL

TEST START:

WELCOME - STEP 1

*MEASURE ROOM TEMPERATURE
*KEEP A MANUAL LOG READY (NOTE THE TIMES IF YOU ALSO DOCUMENT NOISES OUTSIDE AFTER YOU HAVE PUT ON THE MEASURING DEVICES!)

*A POTENTIAL SUBJECT ENTERS THE TEST ROOM. HE/SHE HAD REGISTERED FOR THE TEST / IS SPONTANEOUSLY INTERESTED TO PARTICIPATE.

INVESTIGATOR
(WELCOMING)

“HELLO AND WELCOME TO THIS SUBJECT TEST. THANK YOU FOR YOUR INTEREST AND YOUR PARTICIPATION. NICE THAT YOU'RE HERE. I'LL TELL YOU THE MOST IMPORTANT THINGS RIGHT NOW ABOUT THE PROCEDURE AND GUIDE YOU THROUGH THE TEST.”

*FIRST INTRODUCE THE TEAM IN THE ROOM

“IF YOU HAVE ANY QUESTIONS, PLEASE FEEL FREE TO ASK. YOU CAN ALWAYS TURN TO ME. FIRST OF ALL, VERY BRIEFLY TO OUR TEST SERIES ITSELF:
WE HAVE A RELAXATION SCENE BELOW PREPARED AND WANT TO KNOW: DOES IT HAVE A RELAXING EFFECT?
WE'LL SHOW YOU THE SCENE LATER. HOWEVER, I HAVE TO ASK YOU A FEW QUESTIONS BEFOREHAND: WHETHER YOU HAVE ONE POOR EYESIGHT OR AN EYE DISEASE, SUCH AS SQUINTING, COLOR BLINDNESS OR PARTIAL BLINDNESS OR WHETHER YOU SUFFER FROM EPILEPSY.”

SUBJECT

YES / NO.

INVESTIGATOR

[OPTION 1] YES:

“THEN UNFORTUNATELY AS THIS IS ONE OF OUR REQUIREMENTS
WE CAN'T HAVE YOU PARTICIPATE IN THE STUDY.
NETHERTHELESS, WE THANK YOU, THAT YOU WERE THERE.
HAVE A GUMMY BEAR AND A NICE DAY.”

[OPTION 2] NO:

“OK, GREAT. THEN I WANT TO TELL YOU A
BIT MORE ABOUT TODAY'S TEST.”

RECONNAISSANCE - STEP 2

INVESTIGATOR

“HOWEVER, BEFORE WE CAN BEGIN THE TEST,
PLEASE READ THE DECLARATION OF CONSENT
AND SIGN IF APPROPRIATE.
I WILL EXPLAIN THIS TO YOU VERBALLY, SO ABOUT
THE COURSE OF THE TEST AND POTENTIAL
RISKS YOU FACE WITH THIS
AGREEMENT TO THE DECLARATION OF CONSENT.
IN THE FOLLOWING TEST WE MEASURE RELAXATION
IN VIRTUAL REALITY.
ALL THE TIME YOU ARE HERE AND WE WILL
RECORD EYE MOVEMENTS AND OTHER ANONYMOUS DATA
WHILE YOU WATCH THE SCENE.”

"JUST A LITTLE REMINDER: WE ARE DURING THE
TESTS ALWAYS NEAR YOU AND
ARE CONTACTABLE AT ANY TIME.
WHEN YOU GET SICK OR YOU FEEL
SOMEHOW UNCOMFORTABLE, YOU CAN
ABORT THE ATTEMPT AT ANY TIME, OF COURSE,BY LETTING US
KNOW OR JUST BY TAKING OFF THE VR HEADSET.
FURTHERMORE, YOU ARE WELCOME TO ASK QUESTIONS AT ANY TIME
EITHER ME OR THE TECHNICAL SUPPORTER. THE TEST ITSELF TAKES
ABOUT 60 MINUTES. IF YOU HAVE TO GO TO THE TOILET FIRST, YOU
MAY GO NOW. IF YOU'RE STILL THIRSTY, WE ALSO HAVE WATER HERE.
ARE THERE ANY QUESTIONS YET? OTHERWISE WE CAN START NOW."

[OPTION 1] THE SUBJECT READS AND SIGNS
THE DECLARATION OF CONSENT NOT…

INVESTIGATOR

“UNFORTUNATELY, THEN YOU CANNOT TAKE THE TEST.
THANKS FOR YOUR INTEREST ANYWAY
AND WE WISH YOU A NICE DAY.”

[OPTION 2] SUBJECT SIGNS THE
LETTER OF ACCEPTANCE.

*ID ASSIGNED

INVESTIGATOR

"PERFECT THANKS, THEN WE CAN NOW
MOVE ON TO THE FIRST PART.”

CORTISOL TEST & EYE TEST STEP 3

“IN ORDER TO MEASURE RELAXATION,
WE NEED 3 CORTISOL SAMPLES,
BY USING THIS LITTLE CONTAINER WHICH
CAN BE EVALUATED IN A LABORATORY. CORTISOL IS AN INDICATOR FOR
STRESS OR IN THIS CASE RELAXATION.

SO NOW COMES THE FIRST CORTISOL TEST
AND THEN THE OTHERS IN THE COURSE.
PLEASE PUT THIS COTTON PAD IN YOUR MOUTH TILL IT'S SOAKED.
IF YOU THINK, IT IS ENOUGH DRENCHED PUT IT IN THE SMALL BOX.
THANKS."

"THEN WE COME TO THE NEXT STEP:
SO THE STEREO EYE TEST.
THIS IS ABOUT YOUR ABILITY
FOR STEREOSCOPIC VISION,
I.E. FOR 3D VISION.
WE ARE NOW RUNNING ONE TEST WHICH IS CALLED RANDOT TEST."

* THE INVESTIGATOR PERFORMS WITH THE
SUBJECTS A RANDOT TEST.
THE SUBJECT PUTS ON THE RANDOT TEST GLASSES AND IS INSTRUCTED
TO EXPLAIN WHAT HE/SHE SEES IN DIFFERENT FIELDS (E.G. CROSS,
CIRCLE,...). HE/SHE GOES THROUGH 10 FIELDS AND MUST TELL,
WHICH CIRCLE SEEMS TO BE HIGHLIGHTED: LEFT, MIDDLE OR RIGHT.
THE SCORE IS RECORDED IN THE QUESTIONNAIRE AND SHOULD AT LEAST
BE 6/10 FOR THE FURTHER TEST.

[OPTION 1] SUBJECT FAILS THE TEST...
"THEN UNFORTUNATELY YOU CAN'T
TAKE THE TEST BECAUSE YOUR 3D VISION
IS NOT APPROPRIATE FOR THE SCENE SHOWN.
BECAUSE THAT CAN CAUSE PROBLEMS WHEN RELAXING LEADING
TO UNSUITABLE VALUES.
NETHERTHELESS THANKS FOR YOUR INTEREST ANYWAY
AND WE WISH YOU A NICE DAY."

[OPTION 2] THE SUBJECT PASSES THE TEST…
"VERY GOOD!
NOW WE NEED TO MEASURE YOUR EYE DISTANCE."

* MEASURE EYE DISTANCE

FIRST QUESTIONNAIRE - STEP 4

INVESTIGATOR:
"WE NEED SOME MORE INFORMATION ABOUT YOU AND
OF YOUR DEMOGRAPHICS.
PLEASE FILL IN AND LET US KNOW
WHEN YOU'RE DONE.”

[SHORT BREAK]

“NEXT COME 2 QUESTIONNAIRES:
SIMULATOR SICKNESS QUESTIONNAIRE AND THE
SEMANTIC DIFFERENTIALS
REPORT BACK WHEN YOU'RE DONE WITH THAT."

[SHORT BREAK]

INVESTIGATOR:
"VERY GOOD.
THEN WE ARE NOW READY FOR THAT
ACTUAL TEST."

PREPARATION & CALIBRATION - STEP 5

INVESTIGATOR:
“WE WILL NOW PROVIDE THE REQUIRED EQUIPMENT
SET UP AND ATTACH IT TO YOUR BODY.”
SHOW ECG CHEST STRAP, EXPLAIN SHORT WHAT IT MEASURES
AND HOW TO PUT IT ON.
SEND SUBJECT TO TOILET AND WAIT TILL HE/SHE IS BACK.

“NOW WE'RE GOING TO USE THESE ELECTRODES
(*SHOW GAUGE) AND ATTACH THEM TO YOUR HAND,
TO MEASURE SKIN RESISTANCE.
THE HAND MUST BE AS FREE AS POSSIBLE DURING THE TEST.
PLEASE LIE STILL AND TRY NOT TO HIT WITH THE ELECTRODES
ON OTHER OBJECTS. WHICH HAND DO YOU PREFER?"

*KEEP A MANUAL LOG READY (NOTE THE TIMES IF THERE ARE
NOISES OUTSIDE, ALSO DOCUMENT THEM AFTER PUTTING ON THE
MEASURING DEVICES!)

“WE WILL NOW MEASURE YOUR VALUES FOR 3 MINUTES,
TO DETERMINE THE BASELINE, BECAUSE EACH INDIVIDUAL
HAS ALSO AN INDIVIDUAL BASELINE.
WE THEN USE THAT TO COMPARE THEM
LATER MEASURED VALUES.”

*THE INVESTIGATOR INSTRUCTS THE SUBJECT
TO TAKE A SEAT ON THE DECKCHAIR AND THIS ONE
PUT IN A RELAXED POSITION. ANEW
POINT OUT THAT THE ELECTRODES SHOULD NOT BE HIT ON
OTHER OBJECTS (RELAXED PERSON, E.G.
ON ARMREST OR LAP).

“NOW COMES THE SECOND CORTISOL TEST.
PLEASE PUT IT AGAIN IN YOUR MOUTH AS LAST TIME.”

* DRAW CURTAINS, DIM BRIGHT LIGHT

*PUT ON VR GLASSES
//IF THERE ARE MORE PEOPLE HELPING ALSO INTRODUCE
THEM HERE!
*IF NECESSARY THE GLASSES STILL HAVE TO BE ALIGNED
SO THAT THE INTEGRATED CAMERAS CAN CORRECTLY CAPTURE THE EYES.
*THE SUBJECT HAS TO SCREW IT TIGHT (CAUTION MEASURING DEVICES)
-> LET THE SUBJECT DO IT HIM-/HERSELF
ASK:
“DOES EVERYTHING FIT PERFECTLY? NOT TOO TIGHT OR TOO LOOSE?”

CALIBRATION - STEP 6

INVESTIGATOR:
“NOW WE CAN START WITH THE CALIBRATION.
THEREFORE A LITTLE CIRCLE WILL BE SHOWN IN THE VR GLASSES IN A
MOMENT. PLEASE FIX IT WITH YOUR EYES, BUT WITHOUT MOVING THE
HEAD. MANY THANKS. LET ME KNOW WHEN YOU'RE READY."

THE CALIBRATION IS STARTED.
THE INVESTIGATOR/ TECH SUPPORT SHOULD ALWAYS CHECK
WHETHER THE PUPILS CAN BE RECOGNIZED CORRECTLY.
SHOULD THIS NOT BE THE CASE,
THE TEST MUST BE RESTARTED, THE SETTINGS
CHECKED AND THE CAMERAS REALIGNED.

INVESTIGATOR:
“SO THE CALIBRATION IS COMPLETE AND WE
SHOW YOU THE RELAXATION SCENE NOW.
IT TAKES ABOUT 5 MINUTES.
LIKE I SAID, IT'S NOT ABOUT TESTING YOU
BUT THE SCENE. YOU MAY ASK QUESTIONS AFTER WATCHING
EXCEPT YOU DECIDE TO ABORT THE TEST."

IMPLEMENTATION - STEP 7

*THE RELAXATION SCENE IS PLAYED
AND VALUES MEASURED.

*AFTER SCENE, TELL SUBJECT THE SCENE IS OVER AND
TAKE OFF THE VR GLASSES

INVESTIGATOR:
“EVERYTHING IS FINE?”
*RESPOND TO ANSWER
“THEN WE NEED THE LAST CORTISOL SAMPLE NOW
FOR TODAY.”

*CORTISOL TEST 3
*OPEN CURTAINS/ LIGHT

INVESTIGATOR:
"WELL, THEN YOU CAN GET UP AND
DID MOST OF IT. ONLY A FEW FINAL
QUESTIONNAIRES ARE MISSING.”

*ACCOMPANY SUBJECT AWAY FROM TEST CHAIR

“OK, VERY GOOD, THEN I'LL RELEASE YOU FROM YOUR
MEASURING DEVICES.”

*REMOVE EDA

“YOU CAN REMOVE THE CHEST STRAP YOURSELF
YOU MAY GO TO THE TOILET AGAIN.”

SECOND QUESTIONNAIRE MARATHON - STEP 8

“NOW COMES THE SSQ QUESTIONNAIRE AGAIN
AND THE SEMANTIC DIFFERENTIALS.
ALSO DO A TLX TEST,
WITH WHICH WE WANT TO CAPTURE THE STRAIN. IF YOU HAVE ANY QUESTIONS YOU CAN ASK THEM ANYTIME.”

“FIRST FILL IN THIS:”
*SHOW

*MEANWHILE TECH. SUPPORT NEEDS TO CHECK DATA (HAVE IT CHECKED), BACKUP DATA

"GREAT LAST BUT NOT LEAST, I’D LIKE TO HAVE A PERSONAL LITTLE INTERVIEW ABOUT VR AND THE SCENE YOU HAVE SEEN.
*ASK QUALITATIVE QUESTIONS

“DO YOU HAVE ANY OTHER QUESTIONS? THEN WE ARE FINISHED HERE.
THANK YOU FOR YOUR PARTICIPATION, YOU HELPED A LOT.
AS A REWARD, THERE ARE GUMMY BEARS HERE”

SANITIZE EQUIPMENT - STEP 9

ATTACHMENT 2

F-TABLES SEMANTIC DIFFERENTIALS

	Subject-ID	ease - excited	pleasant - unpleasant	relaxed - stressed	calm - nervous	competent - overwhelmed	safe - helpless	fearless - anxious	simple - difficult	familiar - unfamiliar	desirable - undesirable	comfortable - uncomfortable
	01VRU1	4	3	4	4	3	3	3	2	5	3	3
	01VRE1	2	2	1	1	1	3	3	1	4	1	3
	02VRE1	2	1	1	1	2	1	1	2	5	4	1
	03VRE1	4	2	4	3	2	1	2	2	4	2	2
	04VRE1	2	3	2	2	4	2	2	1	3	4	3
	02VRU1	1	1	2	1	1	1	1	1	3	1	1
	06VRE	5	2	3	4	2	2	3	1	5	1	1
	03VRU1	2	1	3	4	4	3	4	4	6	4	3
	04VRU1	5	3	4	4	3	3	2	3	5	2	2
	05VRU1	1	1	5	2	2	1	1	1	3	2	1
	06VRU1	3	1	1	1	2	2	1	2	4	1	1
	07VRU1	6	5	6	5	2	2	2	4	5	4	3
Variance		2,810606061	1,537878788	2,727272727	2,242424242	0,96969697	0,727272727	0,992424242	1,272727273	0,96969697	1,71969697	0,909090909
F-value	2,81793047	unequal	unequal	unequal	unequal	equal	equal	equal	equal	equal	equal	equal
Variance ratio		6,288135593	3,625	4,337349398	9,25	1,6	1,142857143	2,220338983	1,365853659	1,09375	2,292929293	2,142857143
	01VRU2	1	1	1	1	2	1	2	2	1	1	1
	01VRE2	1	1	1	1	1	1	1	1	3	1	2
	02VRE2	1	1	1	1	1	1	1	1	2	1	1
	03VRE2	1	1	2	1	2	1	1	2	2	2	1
	04VRE2	1	2	1	1	3	2	1	2	3	2	3
	02VRU2	1	1	1	1	1	1	1	1	1	1	1
	06VRE2	2	1	2	2	1	1	1	1	1	1	2
	03VRU2	2	1	2	2	2	2	2	2	3	3	2
	04VRU2	2	2	3	2	3	3	3	3	4	3	2
	05VRU2	1	1	1	1	1	1	1	1	1	1	1
	06VRU2	1	1	1	1	1	1	1	1	2	2	2
	07VRU2	3	3	3	2	2	3	2	4	3	3	2
	Varianz	0,446969697	0,424242424	0,628787879	0,242424242	0,606060606	0,636363636	0,446969697	0,931818182	1,060606061	0,75	0,424242424

	Subject-ID	ease - excited	pleasant - unpleasant	relaxed - stressed	calm - nervous	competent - overwhelmed	safe - helpless	fearless - anxious	simple - difficult	familiar - unfamiliar	desirable - undesirable	comfortable - uncomfortable
	01TV1	1	1	1	1	1	1	1	1	1	3	2
	02TV1	3	2	2	3	2	1	1	2	5	2	2
	03TV1	2	2	3	3	4	3	3	2	3	4	4
	04TV1	2	1	1	2	1	1	3	3	4	1	2
	05TV1	2	2	1	1	1	1	1	1	1	1	1
	06TV1	6	2	4	5	4	3	3	2	5	2	2
	07TV1	2	2	3	2	1	1	1	3	4	4	3
	08TV1	3	2	2	1	3	1	3	1	2	2	1
	09TV1	4	1	3	3	1	1	2	1	4	3	2
	10TV1	2	5	2	2	2	2	1	2	3	3	2
	11TV1	3	1	3	2	3	2	2	1	5	2	3
Variance		1,818181818	1,290909091	1,018181818	1,418181818	1,490909091	0,672727273	0,890909091	0,618181818	2,254545455	1,072727273	0,763636364
F-value	2,978237016	unequal	unequal	unequal	unequal	equal	equal	equal	equal	unequal	equal	equal
Variance ratio		6,666666667	7,888888889	4,666666667	6,5	1,744680851	1,48	1,884615385	1,36	5,636363636	1,113207547	3
	01TV2	2	1	2	1	1	1	1	1	2	3	2
	02TV2	1	2	1	2	1	2	1	2	3	2	1
	03TV2	2	2	1	1	3	3	2	3	2	3	2
	04TV2	2	1	1	1	1	1	2	1	2	2	2
	05TV2	1	1	1	1	1	1	1	1	1	1	1
	06TV2	2	1	1	2	1	1	1	2	2	1	1
	07TV2	1	1	1	1	1	1	1	1	1	4	1
	08TV2	1	1	1	1	3	1	2	2	2	2	1
	09TV2	2	1	1	1	1	1	1	1	2	1	1
	10TV2	2	1	2	1	2	1	1	2	2	2	1
	11TV2	1	1	2	2	3	2	3	2	3	3	2
Variance		0,272727273	0,163636364	0,218181818	0,218181818	0,854545455	0,454545455	0,472727273	0,454545455	0,4	0,963636364	0,254545455

ATTACHMENT 3

T-TEST TABLES SEMANTIC DIFFERENTIALS TV

Pleasant - unpleasant
t-Test: Two-Sample Assuming Uenqual Variances

	Variable 1	*Variable 2*
Mean	2,72727273	1,54545455
Variance	1,81818182	0,27272727
Observations	11	11
Hypothesized Mean Difference	0	
df	13	
t Stat	2,71068738	
P(T<=t) one-tail	0,00891344	
t Critical one-tail	1,7709334	
P(T<=t) two-tail	0,01782687	
t Critical two-tail	2,16036866	

At ease - excited
t-Test: Two-Sample Assuming Uenqual Variances

	Variable 1	*Variable 2*
Mean	1,90909091	1,18181818
Variance	1,29090909	0,16363636
Observations	11	11
Hypothesized Mean Difference	0	
df	12	
t Stat	2	
P(T<=t) one-tail	0,03432751	
t Critical one-tail	1,78228756	
P(T<=t) two-tail	0,06865501	
t Critical two-tail	2,17881283	

Relaxed - stressed
t-Test: Two-Sample Assuming Uenqual Variances

	Variable 1	*Variable 2*
Mean	2,27272727	1,27272727
Variance	1,01818182	0,21818182
Observations	11	11
Hypothesized Mean Difference	0	
df	14	
t Stat	2,9827938	
P(T<=t) one-tail	0,0049414	
t Critical one-tail	1,76131014	
P(T<=t) two-tail	0,0098828	
t Critical two-tail	2,14478669	

Calm - nervous
t-Test: Two-Sample Assuming Uenqual Variances

	Variable 1	*Variable 2*
Mean	2,27272727	1,27272727
Variance	1,41818182	0,21818182
Observations	11	11
Hypothesized Mean Difference	0	
df	13	
t Stat	2,59272486	
P(T<=t) one-tail	0,0111559	
t Critical one-tail	1,7709334	
P(T<=t) two-tail	0,02231179	
t Critical two-tail	2,16036866	

Competent - overwhelmed
t-Test: Two-Sample Assuming Equal Variances

	Variable 1	*Variable 2*
Mean	2,09090909	1,63636364
Variance	1,49090909	0,85454545
Observations	11	11
Pooled Variance	1,17272727	
Hypothesized Mean Difference	0	
df	20	
t Stat	0,98437404	
P(T<=t) one-tail	0,16834556	
t Critical one-tail	1,72471824	
P(T<=t) two-tail	0,33669112	
t Critical two-tail	2,08596345	

Save -helpless
t-Test: Two-Sample Assuming Equal Variances

	Variable 1	*Variable 2*
Mean	1,54545455	1,36363636
Variance	0,67272727	0,45454545
Observations	11	11
Pooled Variance	0,56363636	
Hypothesized Mean Difference	0	
df	20	
t Stat	0,56796183	
P(T<=t) one-tail	0,28819117	
t Critical one-tail	1,72471824	
P(T<=t) two-tail	0,57638234	
t Critical two-tail	2,08596345	

ATTACHMENT

Fearless - anxious
t-Test: Two-Sample Assuming Equal Variances

	Variable 1	*Variable 2*
Mean	1,90909091	1,45454545
Variance	0,89090909	0,47272727
Observations	11	11
Pooled Variance	0,68181818	
Hypothesized Mean Difference	0	
df	20	
t Stat	1,29099445	
P(T<=t) one-tail	0,1057147	
t Critical one-tail	1,72471824	
P(T<=t) two-tail	0,21142939	
t Critical two-tail	2,08596345	

Simple - difficult
t-Test: Two-Sample Assuming Equal Variances

	Variable 1	*Variable 2*
Mean	1,72727273	1,63636364
Variance	0,61818182	0,45454545
Observations	11	11
Pooled Variance	0,53636364	
Hypothesized Mean Difference	0	
df	20	
t Stat	0,29111125	
P(T<=t) one-tail	0,38698191	
t Critical one-tail	1,72471824	
P(T<=t) two-tail	0,77396382	
t Critical two-tail	2,08596345	

Familiar - unfamiliar
t-Test: Two-Sample Assuming Uenqual Variances

	Variable 1	*Variable 2*
Mean	3,36363636	2
Variance	2,25454545	0,4
Observations	11	11
Hypothesized Mean Difference	0	
df	13	
t Stat	2,77587454	
P(T<=t) one-tail	0,00786988	
t Critical one-tail	1,7709334	
P(T<=t) two-tail	0,01573975	
t Critical two-tail	2,16036866	

Desirable - undesirable
t-Test: Two-Sample Assuming Equal Variances

	Variable 1	*Variable 2*
Mean	2,45454545	2,18181818
Variance	1,07272727	0,96363636
Observations	11	11
Pooled Variance	1,01818182	
Hypothesized Mean Difference	0	
df	20	
t Stat	0,63386569	
P(T<=t) one-tail	0,26667579	
t Critical one-tail	1,72471824	
P(T<=t) two-tail	0,53335158	
t Critical two-tail	2,08596345	

Comfortable - uncomfortable
t-Test: Two-Sample Assuming Uenqual Variances

	Variable 1	*Variable 2*
Mean	2,18181818	1,36363636
Variance	0,76363636	0,25454545
Observations	11	11
Hypothesized Mean Difference	0	
df	16	
t Stat	2,68926437	
P(T<=t) one-tail	0,00806124	
t Critical one-tail	1,74588368	
P(T<=t) two-tail	0,01612249	
t Critical two-tail	2,1199053	

ATTACHMENT 4

T-TEST TABLES SEMANTIC DIFFERENTIALS VR

Pleasant - unpleasant
t-Test: Two-Sample Assuming Uenqual Variances

	Variable 1	*Variable 2*
Mean	2,08333333	1,33333333
Variance	1,53787879	0,42424242
Observations	12	12
Hypothesized Mean Difference	0	
df	17	
t Stat	1,85476533	
P(T<=t) one-tail	0,04052908	
t Critical one-tail	1,73960673	
P(T<=t) two-tail	0,08105815	
t Critical two-tail	2,10981558	

At ease - excited
t-Test: Two-Sample Assuming Uenqual Variances

	Variable 1	*Variable 2*
Mean	3,08333333	1,41666667
Variance	2,81060606	0,4469697
Observations	12	12
Hypothesized Mean Difference	0	
df	14	
t Stat	3,198837	
P(T<=t) one-tail	0,0032177	
t Critical one-tail	1,76131014	
P(T<=t) two-tail	0,00643541	
t Critical two-tail	2,14478669	

Relaxed - stressed
t-Test: Two-Sample Assuming Uenqual Variances

	Variable 1	*Variable 2*
Mean	3	1,58333333
Variance	2,72727273	0,62878788
Observations	12	12
Hypothesized Mean Difference	0	
df	16	
t Stat	2,67881919	
P(T<=t) one-tail	0,00823533	
t Critical one-tail	1,74588368	
P(T<=t) two-tail	0,01647067	
t Critical two-tail	2,1199053	

Calm - nervous
t-Test: Two-Sample Assuming Uenqual Variances

	Variable 1	*Variable 2*
Mean	2,66666667	1,33333333
Variance	2,24242424	0,24242424
Observations	12	12
Hypothesized Mean Difference	0	
df	13	
t Stat	2,9300795	
P(T<=t) one-tail	0,00585579	
t Critical one-tail	1,7709334	
P(T<=t) two-tail	0,01171158	
t Critical two-tail	2,16036866	

Competent - overwhelmed
t-Test: Two-Sample Assuming Equal Variances

	Variable 1	*Variable 2*
Mean	2,33333333	1,66666667
Variance	0,96969697	0,60606061
Observations	12	12
Pooled Variance	0,78787879	
Hypothesized Mean Difference	0	
df	22	
t Stat	1,83973242	
P(T<=t) one-tail	0,03967046	
t Critical one-tail	1,71714437	
P(T<=t) two-tail	0,07934093	
t Critical two-tail	2,07387307	

Save -helpless
t-Test: Two-Sample Assuming Equal Variances

	Variable 1	*Variable 2*
Mean	2	1,5
Variance	0,72727273	0,63636364
Observations	12	12
Pooled Variance	0,68181818	
Hypothesized Mean Difference	0	
df	22	
t Stat	1,4832397	
P(T<=t) one-tail	0,07609808	
t Critical one-tail	1,71714437	
P(T<=t) two-tail	0,15219616	
t Critical two-tail	2,07387307	

ATTACHMENT

Fearless - anxious
t-Test: Two-Sample Assuming Equal Variances

	Variable 1	*Variable 2*
Mean	2,08333333	1,41666667
Variance	0,99242424	0,4469697
Observations	12	12
Pooled Variance	0,71969697	
Hypothesized Mean Difference	0	
df	22	
t Stat	1,92490601	
P(T<=t) one-tail	0,033634	
t Critical one-tail	1,71714437	
P(T<=t) two-tail	0,067268	
t Critical two-tail	2,07387307	

Simple - difficult
t-Test: Two-Sample Assuming Equal Variances

	Variable 1	*Variable 2*
Mean	2	1,75
Variance	1,27272727	0,93181818
Observations	12	12
Pooled Variance	1,10227273	
Hypothesized Mean Difference	0	
df	22	
t Stat	0,58327197	
P(T<=t) one-tail	0,28282111	
t Critical one-tail	1,71714437	
P(T<=t) two-tail	0,56564221	
t Critical two-tail	2,07387307	

Familiar - unfamiliar
t-Test: Two-Sample Assuming Equal Variances

	Variable 1	*Variable 2*
Mean	4,33333333	2,16666667
Variance	0,96969697	1,06060606
Observations	12	12
Pooled Variance	1,01515152	
Hypothesized Mean Difference	0	
df	22	
t Stat	5,2674727	
P(T<=t) one-tail	1,3818E-05	
t Critical one-tail	1,71714437	
P(T<=t) two-tail	2,7635E-05	
t Critical two-tail	2,07387307	

Desirable - undesirable
t-Test: Two-Sample Assuming Equal Variances

	Variable 1	*Variable 2*
Mean	2,41666667	1,75
Variance	1,71969697	0,75
Observations	12	12
Pooled Variance	1,23484848	
Hypothesized Mean Difference	0	
df	22	
t Stat	1,46952686	
P(T<=t) one-tail	0,07792316	
t Critical one-tail	1,71714437	
P(T<=t) two-tail	0,15584632	
t Critical two-tail	2,07387307	

Comfortable - uncomfortable
t-Test: Two-Sample Assuming Equal Variances

	Variable 1	*Variable 2*
Mean	2	1,66666667
Variance	0,90909091	0,42424242
Observations	12	12
Pooled Variance	0,66666667	
Hypothesized Mean Difference	0	
df	22	
t Stat	1	
P(T<=t) one-tail	0,16409163	
t Critical one-tail	1,71714437	
P(T<=t) two-tail	0,32818326	
t Critical two-tail	2,07387307	

ATTACHMENT 5

T-TEST TABLES VRSQ TV

General discomfort

t-Test: Two-Sample Assuming Equal Variances

	Variable 1	*Variable 2*
Mean	0,09090909	0,09090909
Variance	0,09090909	0,09090909
Observations	11	11
Pooled Variance	0,09090909	
Hypothesized Mean Difference	0	
df	20	
t Stat	0	
P(T<=t) one-tail	0,5	
t Critical one-tail	1,72471824	
P(T<=t) two-tail	1	
t Critical two-tail	2,08596345	

Fatigue

t-Test: Two-Sample Assuming Equal Variances

	Variable 1	*Variable 2*
Mean	0,81818182	1,27272727
Variance	0,76363636	0,81818182
Observations	11	11
Pooled Variance	0,79090909	
Hypothesized Mean Difference	0	
df	20	
t Stat	-1,19865825	
P(T<=t) one-tail	0,12233575	
t Critical one-tail	1,72471824	
P(T<=t) two-tail	0,2446715	
t Critical two-tail	2,08596345	

Headache

t-Test: Two-Sample Assuming Equal Variances

	Variable 1	*Variable 2*
Mean	0,18181818	0,09090909
Variance	0,16363636	0,09090909
Observations	11	11
Pooled Variance	0,12727273	
Hypothesized Mean Difference	0	
df	20	
t Stat	0,5976143	
P(T<=t) one-tail	0,27840203	
t Critical one-tail	1,72471824	
P(T<=t) two-tail	0,55680405	
t Critical two-tail	2,08596345	

Eye strain

t-Test: Two-Sample Assuming Uenqual Variances

	Variable 1	*Variable 2*
Mean	0,27272727	0,27272727
Variance	0,81818182	0,21818182
Observations	11	11
Hypothesized Mean Difference	0	
df	15	
t Stat	0	
P(T<=t) one-tail	0,5	
t Critical one-tail	1,75305036	
P(T<=t) two-tail	1	
t Critical two-tail	2,13144955	

Difficulty focusing

t-Test: Two-Sample Assuming Equal Variances

	Variable 1	*Variable 2*
Mittelwert	0,27272727	0,18181818
Varianz	0,21818182	0,16363636
Beobachtungen	11	11
Gepoolte Varianz	0,19090909	
Hypothetische Differenz der Mittelwerte	0	
Freiheitsgrade (df)	20	
t-Statistik	0,48795004	
P(T<=t) einseitig	0,31544584	
Kritischer t-Wert bei einseitigem t-Test	1,72471824	
P(T<=t) zweiseitig	0,63089169	
Kritischer t-Wert bei zweiseitigem t-Test	2,08596345	

Fullness of Head

t-Test: Two-Sample Assuming Equal Variances

	Variable 1	*Variable 2*
Mittelwert	0,09090909	0,09090909
Varianz	0,09090909	0,09090909
Beobachtungen	11	11
Gepoolte Varianz	0,09090909	
Hypothetische Differenz der Mittelwerte	0	
Freiheitsgrade (df)	20	
t-Statistik	0	
P(T<=t) einseitig	0,5	
Kritischer t-Wert bei einseitigem t-Test	1,72471824	
P(T<=t) zweiseitig	1	
Kritischer t-Wert bei zweiseitigem t-Test	2,08596345	

Blurred vision

t-Test: Two-Sample Assuming Equal Variances

	Variable 1	*Variable 2*
Mean	0	0,09090909
Variance	0	0,09090909
Observations	11	11
Pooled Variance	0,04545455	
Hypothesized Mean Difference	0	
df	20	
t Stat	-1	
P(T<=t) one-tail	0,16462829	
t Critical one-tail	1,72471824	
P(T<=t) two-tail	0,32925658	
t Critical two-tail	2,08596345	

Dizziness with closed eyes

t-Test: Two-Sample Assuming Equal Variances

	Variable 1	*Variable 2*
Mean	0,09090909	0
Variance	0,09090909	0
Observations	11	11
Pooled Variance	0,04545455	
Hypothesized Mean Difference	0	
df	20	
t Stat	1	
P(T<=t) one-tail	0,16462829	
t Critical one-tail	1,72471824	
P(T<=t) two-tail	0,32925658	
t Critical two-tail	2,08596345	

Vertigo

t-Test: Two-Sample Assuming Equal Variances

	Variable 1	*Variable 2*
Mean	0	0
Variance	0	0
Observations	11	11
Pooled Variance	0	
Hypothesized Mean Difference	0	
df	20	
t Stat	65535	
P(T<=t) one-tail	0	
t Critical one-tail	1,72471824	
P(T<=t) two-tail	0	
t Critical two-tail	2,08596345	

ATTACHMENT 6

T-TEST TABLES VRSQ VR

General discomfort

t-Test: Two-Sample Assuming Uenqual Variances

	Variable 1	*Variable 2*
Mean	0,91666667	0,33333333
Variance	1,17424242	0,24242424
Observations	12	12
Hypothesized Mean Difference	0	
df	15	
t Stat	1,69774938	
P(T<=t) one-tail	0,05509916	
t Critical one-tail	1,75305036	
P(T<=t) two-tail	0,11019833	
t Critical two-tail	2,13144955	

Fatigue

t-Test: Two-Sample Assuming Equal Variances

	Variable 1	*Variable 2*
Mean	1,25	1,16666667
Variance	1,29545455	1,42424242
Observations	12	12
Pooled Variance	1,35984848	
Hypothesized Mean Difference	0	
df	22	
t Stat	0,17504476	
P(T<=t) one-tail	0,43132274	
t Critical one-tail	1,71714437	
P(T<=t) two-tail	0,86264548	
t Critical two-tail	2,07387307	

Headache

t-Test: Two-Sample Assuming Uenqual Variances

	Variable 1	*Variable 2*
Mean	0,58333333	0,16666667
Variance	1,35606061	0,15151515
Observations	12	12
Hypothesized Mean Difference	0	
df	13	
t Stat	1,17554649	
P(T<=t) one-tail	0,13042854	
t Critical one-tail	1,7709334	
P(T<=t) two-tail	0,26085708	
t Critical two-tail	2,16036866	

Eye strain

t-Test: Two-Sample Assuming Equal Variances

	Variable 1	*Variable 2*
Mean	0,83333333	0,75
Variance	1,24242424	1,29545455
Observations	12	12
Pooled Variance	1,26893939	
Hypothesized Mean Difference	0	
df	22	
t Stat	0,18120657	
P(T<=t) one-tail	0,42893225	
t Critical one-tail	1,71714437	
P(T<=t) two-tail	0,8578645	
t Critical two-tail	2,07387307	

Difficulty focusing

t-Test: Two-Sample Assuming Uenqual Variances

	Variable 1	*Variable 2*
Mean	1,41666667	0,66666667
Variance	1,53787879	0,24242424
Observations	12	12
Hypothesized Mean Difference	0	
df	14	
t Stat	1,94717471	
P(T<=t) one-tail	0,03592932	
t Critical one-tail	1,76131014	
P(T<=t) two-tail	0,07185865	
t Critical two-tail	2,14478669	

Fullness of Head

t-Test: Two-Sample Assuming Equal Variances

	Variable 1	*Variable 2*
Mean	0,91666667	0,5
Variance	1,71969697	0,81818182
Observations	12	12
Pooled Variance	1,26893939	
Hypothesized Mean Difference	0	
df	22	
t Stat	0,90603285	
P(T<=t) one-tail	0,18736757	
t Critical one-tail	1,71714437	
P(T<=t) two-tail	0,37473513	
t Critical two-tail	2,07387307	

ATTACHMENT

Blurred vision
t-Test: Two-Sample Assuming Uenqual Variances

	Variable 1	*Variable 2*
Mean	0,41666667	0,41666667
Variance	0,81060606	0,26515152
Observations	12	12
Hypothesized Mean Difference	0	
df	18	
t Stat	0	
P(T<=t) one-tail	0,5	
t Critical one-tail	1,73406361	
P(T<=t) two-tail	1	
t Critical two-tail	2,10092204	

Dizziness with closed eyes
t-Test: Two-Sample Assuming Equal Variances

	Variable 1	*Variable 2*
Mean	0,08333333	0
Variance	0,08333333	0
Observations	12	12
Pooled Variance	0,04166667	
Hypothesized Mean Difference	0	
df	22	
t Stat	1	
P(T<=t) one-tail	0,16409163	
t Critical one-tail	1,71714437	
P(T<=t) two-tail	0,32818326	
t Critical two-tail	2,07387307	

Vertigo
t-Test: Two-Sample Assuming Equal Variances

	Variable 1	*Variable 2*
Mean	0,08333333	0,08333333
Variance	0,08333333	0,08333333
Observations	12	12
Pooled Variance	0,08333333	
Hypothesized Mean Difference	0	
df	22	
t Stat	0	
P(T<=t) one-tail	0,5	
t Critical one-tail	1,71714437	
P(T<=t) two-tail	1	
t Critical two-tail	2,07387307	

ATTACHMENT 7

T-TEST TABLES RAW TLX

Mental Demand

t-Test: Two-Sample Assuming Equal Variances

	Variable 1	*Variable 2*
Mean	1,63636364	1,91666667
Variance	0,65454545	1,35606061
Observations	11	12
Pooled Variance	1,02200577	
Hypothesized Mean Difference	0	
df	21	
t Stat	-0,66423869	
P(T<=t) one-tail	0,25688274	
t Critical one-tail	1,7207429	
P(T<=t) two-tail	0,51376549	
t Critical two-tail	2,07961384	

Physical Demand

t-Test: Two-Sample Assuming Equal Variances

	Variable 1	*Variable 2*
Mean	1,09090909	1,25
Variance	0,09090909	0,20454545
Observations	11	12
Pooled Variance	0,1504329	
Hypothesized Mean Difference	0	
df	21	
t Stat	-0,98264579	
P(T<=t) one-tail	0,16848413	
t Critical one-tail	1,7207429	
P(T<=t) two-tail	0,33696826	
t Critical two-tail	2,07961384	

Temporal Demand

t-Test: Two-Sample Assuming Uenqual Variances

	Variable 1	*Variable 2*
Mean	1,45454545	1,25
Variance	0,67272727	0,20454545
Observations	11	12
Hypothesized Mean Difference	0	
df	15	
t Stat	0,73144146	
P(T<=t) one-tail	0,23789099	
t Critical one-tail	1,75305036	
P(T<=t) two-tail	0,47578197	
t Critical two-tail	2,13144955	

Performance

t-Test: Two-Sample Assuming Unequal Variances

	Variable 1	*Variable 2*
Mean	2,36363636	1,5
Variance	4,05454545	0,27272727
Observations	11	12
Hypothesized Mean Difference	0	
df	11	
t Stat	1,38058647	
P(T<=t) one-tail	0,0974067	
t Critical one-tail	1,79588482	
P(T<=t) two-tail	0,19481341	
t Critical two-tail	2,20098516	

Effort

t-Test: Two-Sample Assuming Equal Variances

	Variable 1	*Variable 2*
Mean	1,09090909	1,25
Variance	0,09090909	0,20454545
Observations	11	12
Pooled Variance	0,1504329	
Hypothesized Mean Difference	0	
df	21	
t Stat	-0,98264579	
P(T<=t) one-tail	0,16848413	
t Critical one-tail	1,7207429	
P(T<=t) two-tail	0,33696826	
t Critical two-tail	2,07961384	

Frustration

t-Test: Two-Sample Assuming Unequal Variances

	Variable 1	*Variable 2*
Mean	1,18181818	1,75
Variance	0,16363636	3,11363636
Observations	11	12
Hypothesized Mean Difference	0	
df	12	
t Stat	-1,08477141	
P(T<=t) one-tail	0,14966789	
t Critical one-tail	1,78228756	
P(T<=t) two-tail	0,29933578	
t Critical two-tail	2,17881283	

ATTACHMENT 8

RELAXATION SCENE