NTNU
Norwegian University of
Science and Technology

Contributing Factors of Immersion in Interactive Digital Storytelling

Evaluation of Heart Rate as a Parameter for Immersion in QoE assessment of Interactive Storytelling

Martin Ervik

Master of Science in Communication Technology
Submission date: June 2017
Supervisor: Andrew Perkis, IES
Co-supervisor: Sebastian Arndt, IES

Norwegian University of Science and Technology
Department of Electronic Systems

Summary

Technology has improved to the point where the focus area of quality assessment is shifting from the delivery focused Quality of Service (QoS), to a more user-focused assessment field, Quality of Experience (QoE). This enables us to use physiological measurements of users to assess the QoE. Although widely accepted and used, there is little research on using physiological measurements for assessing QoE in the new display technology, e.g. telling an interactive digital story by using head-mounted-displays (HMD) in Virtual Reality (VR).

The goal of this thesis was to test if changes in physiological signals can be used as a parameter in QoE assessments when using HMD in VR. This is to evaluate the emotional response of a user in terms of the level of immersion, in addition to self-assessments. It aims to investigate if the QoE can be quantified through physiological measurements only, hence introducing the possibility to measure the QoE without intruding the user.

A test hypothesis was formulated: "The level of immersion is related to the measured heart rate in a way that the heart rate increases when a participant's level of immersion increases.". This report describes the design and implementation of a proposed system suitable for conducting an experiment trying to test this thesis.

In the experiment, the participants played through a high-quality digital story named *Everest VR*. It consisted of five scenes with different content. During the scenes, participants had their heart rate measured by a wearable health tracker. After each scene, the participants answered questionnaires regarding their experience of that particular scene.

23 people participated in the experiment. Self-reported measures of immersion show that the content itself may not necessarily need to be highly arousing in order to make the participant feel more immersed. When it comes to physiological measures, we can see that a more interactive scene leads to a higher level of heart rate. Therefore it is more important that the user of a VR system has the impression of being in control of the experience in order to enjoy it more.

There was found no statistical correlation between the physiological and subjective measurements, hence discarding our test hypothesis. However, to better estimate the emotional state, measures such as heart-rate variability or skin conductance may be more appropriate and needs further investigations. It is also necessary to develop a standardized experimental setup for conducting similar research.

Sammendrag

Teknologien har forbedret seg til et punkt hvor fokusområdet for kvalitetsvurdering har skiftet fra det leveringsfokuserte Quality of Service (QoS), til et mer brukerfokusert vurderingsfelt, Quality of Experience (QoE). Dette åpner for å bruke fysiologiske målinger til å vurdere QoE. Selv om det er allment akseptert og brukt, er det lite forskning på bruk av fysiologiske målinger for å vurdere QoE i ny skjermteknologi, for eksempel gjennom å fortelle en interaktiv digital historie ved bruk av head-mounted-displays(HMD) i virtual reality (VR).

Målet med denne oppgaven var å teste om endringer i fysiologiske signaler kan brukes som parameter i QoE vurderinger når HMD i VR blir brukt til å fortelle en interaktiv digital historie. Dette er for å evaluere den følelsesmessig responsen en bruker får av opplevelsen, sett med tanke på hvor oppslukt (immersed) brukeren føler seg, i tillegg til subjektive tilbakemeldinger. Oppgaven tar sikte på å undersøke om QoE kan kvantifiseres gjennom fysiologiske målinger utelukkende, og dermed introdusere muligheten for å måle QoE uten å avbryte opplevelsen til brukeren.

En testhypotese ble formulert: "Nivået av oppslukthet (immersion) er relatert til den målte hjertefrekvensen på en måte slik at hjertefrekvensen øker når en brukers følelse av oppslukthet (immersion) øker.". Denne rapporten beskriver utformingen og gjennomføringen av et foreslått system som er egnet for å gjennomføre et forsøk som prøver å teste denne hypotesen. I eksperimentet spilte deltakerne gjennom en interaktiv digital historie av høy kvalitet kalt *Everest VR*. Den besto av fem scener med forskjellig innhold. Under scenene fikk deltakerne hjertefrekvensen sin målt av en smartklokke. Etter hver scene besvarte deltakerne spørreskjemaer om opplevelsen de hadde i den aktuelle scenen.

23 personer deltok i forsøket. Subjektive tilbakemeldinger om nivået av oppslukthet (immersion) viser at innholdet i seg selv ikke nødvendigvis må være svært spennende for å få deltakeren til å føle seg mer oppslukt(immersed) i opplevelsen. Når det gjelder de fysiologiske målingene, kan vi se at en mer interaktiv scene fører til at hjertefrekvensen øker. Derfor er det viktigere at brukeren av et VR-system har inntrykk av å være i kontroll over opplevelsen for å kunne nyte den mer.

Det ble ikke funnet noen statistisk sammenheng mellom de fysiologiske og subjektive målingene, og dermed ble testhypotesen forkastet. For å bedre estimere en brukers emosjonelle tilstand kan tiltak som hjertefrekvensvariabilitet eller hudkonduktans være mer hensiktsmessig, noe som trenger videre undersøkelser. Det er også nødvendig å utvikle et standardisert eksperimentelt oppsett for å utføre lignende undersøkelser.

Preface

This thesis is submitted to fulfill the requirements for receiving a Master of Science (MSc) degree at the Norwegian University of Science and Technology (NTNU). The work presented is conducted over a five-month period from January to June 2017 and was supervised by Professor Andrew Perkis and Post.Doc. Sebastian Arndt.

I would like to thank both of my supervisors for very good supervision, advice along the way and feedback on my work. Arndt, S. deserve extra thanks for his priceless contribution in everything from how to conduct an experiment, as well as deep knowledge on the topics of Quality of Experience (QoE) and physiological signal processing. MSc student Kristoffer Larsen deserves acknowledgment as well since he provided us with the necessary code needed for the analysis.

I am also proud to say that my supervisors used the results in a paper submitted to the *International Conference on Interactive Digital Storytelling* (ICIDS 2017). ICIDS 2017 is the premier annual venue that gathers researchers, developers, practitioners, and theorists to present and share the latest innovations, insights, and techniques in the expanding field of interactive storytelling and the technologies that support it. The conference takes place between the 14th-17th of November in Funchal, Madeira, Portugal.

Also, thanks to Icelandic Sólfar Studios for providing us with the *Everest VR* content and technical support when needed.

Table of Contents

Summary i

Sammendrag ii

Preface iii

Abbreviations vii

1 Introduction 1
1.1 Motivation . . . 1
1.2 Thesis Outline . . . 2
1.3 Research Questions . . . 2

2 Quality Assessment of Interactive Digital Storytelling 3
2.1 Virtual Reality . . . 3
2.1.1 What is it? . . . 3
2.1.2 Head-Mounted-Display . . . 3
2.1.3 Sensors in Virtual Reality Systems . . . 4
2.1.4 Other Important Aspects . . . 4
2.2 Digital Storytelling . . . 4
2.2.1 What is it? . . . 4
2.2.2 Interactive Storytelling in Virtual Reality . . . 5
2.3 Quality of Experience . . . 5
2.3.1 What is it? . . . 6
2.3.2 Measuring the QoE . . . 7
2.3.3 Emotions in QoE Assessment . . . 8
2.4 Physiological Signal Responses . . . 8
2.4.1 What is it? . . . 8
2.4.2 Measuring Physiological Signals . . . 9
2.4.3 Physiological Quality Assessment . . . 10
2.5 Immersion . . . 10
2.5.1 Difference Between Immersion and Presence . . . 11
2.5.2 Quality Assessment of Immersion in Storytelling . . . 11
2.6 Related Work . . . 12

3 Experiment System Design 13
3.1 Measurement and Analyze Procedure . . . 13
3.2 Devices - Hardware and Software . . . 14
3.2.1 Virtual Reality System - HTC Vive VR . . . 14
3.2.2 Wireless Health Tracker - Fitbit Charge 2 . . . 16
3.2.3 Web-based Questionnaires - Google Forms . . . 19
3.3 Stimulus - The Digital Stories . . . 19

3.3.1 Training Session - Trondheim 360 20
3.3.2 Main Session - Everest VR 20

4 Methodology **25**
4.1 Standards and Recommendations 25
4.2 Research Protocol 26
4.2.1 Hypothesis 26
4.2.2 Resources 26
4.2.3 The Experiment 27

5 Results **31**
5.1 Subjective Measurements 31
5.2 Physiological Measurements 34
5.3 Correlation Test 35

6 Discussion **37**
6.1 Subjective Measurements 37
6.2 Physiological Measurements 38
6.3 Critics 38
6.4 Further Work 39

7 Conclusion **41**

Bibliography **43**

A Research Protocol Template **47**

B Research Protocol for QoE in Digital Storytelling Experiment **49**

C The Questionnaire **59**

D Fitbit Web-API **62**

E Python Program for Data Conversion **63**

F Matlab Program for Data Analysis **64**

G Questionnaire Results: MOS-ratings **65**

H Questionnaire Results: ANOVA test **74**

I Physiological Results: Time Plots **75**

J Physiological Results: Average Time spent in different HR-Zones **86**

K Physiological Results: Pairwise T-test Between HR-zones **89**

Abbreviations

QoE	=	Quality of Experience
QoS	=	Quality of Service
HMD	=	Head Mounted Display
VR	=	Virtual Reality
AR	=	Augmented Reality
3D	=	Three-Dimensional
ECG	=	Electrocardiography
EDA	=	Electrodermal Activity
HR	=	Heart Rate
SAM	=	Self-Assessment Manikin
MOS	=	Mean Opinion Score

1 Introduction

1.1 Motivation

Storytelling has played an important part through the ages in preserving the constant urges we have as humans. It does not matter which culture you origin from, storytelling has been, and always will be, the main tool to entertain, educate, and socialize ourselves.

Storytelling probably originated with a group of people sitting around a campfire in a cave. Here, they combined spoken words with wall paintings to create engaging stories. As our predecessors gained knowledge, storytelling started appearing in new formats. With the 20th century came modern technology such as TV, computers etc. This has introduced new platforms for storytelling, turning it digital. In Norway analog TV- and radio signals will be turned off for good in 2017, being replaced with a new digital network. At the same time, 3D- and Virtual Reality technology[1][2] are being used in new and creative ways every single day, reforming the use of digital storytelling in sectors like education and entertainment. [3][4][5]

With recent years advance in technology, systems using head-mounted-displays(HMD) in the fields of Virtual Reality(VR) and Augmented Reality(AR) allows aspects of immersion [6][7][8] and engagement to be considered when producing content for new and interactive, digital narratives. Here, immersion may be defined as how our senses and emotions are deeply involved in an experience[9][10], making us feel as a part of what is happening in the story. Even though systems using HMD are becoming cheaper and more accessible, we still have not seen the 'killer' application or story creating the definite 'must-have' excitement from consumers. To create new and exciting content, producers need ways to gain insight in what consumers experience as good quality content. To do so, we need to assess the users Quality of Experience(QoE) [11][12][13].

With the level of immersion being an important aspect[14] of interactive storytelling, it takes role as a predominant candidate when it comes to quantifying the QoE. Traditionally, the QoE has mainly been assessed by subjective questionnaires [15][16][17], but new concepts also use physiological signal responses[18] (e.g. heart rate variability) from the user as a parameter in quality assessment. Since QoE can be closely related to the 'sensory experience' term[19][20], it can be possible for QoE assessments to use the variation in physiological signals [21][22] to evaluate the emotional response from a user in terms of the level of immersion [23][24][25][26].

Related work on the topic already exists, and are widely accepted for standard displays [27][28][29]. Recent experiments [30][31][32] have been conducted on new display technology(e.g. HMD in VR) as well, with a focus on immersive VR experiences. However, more research is needed to get a better understanding of the relationship between physiological and subjective parameters, and how this can be used in QoE assessment of interactive storytelling. The motivation for this report is to contribute to this research field.

1.2 Thesis Outline

The focus of this thesis is to contribute to the field of QoE assessment of interactive storytelling (when using HMD in VR). The main goal is to provide knowledge of the relationship between users self-reported level of immersion and measures of physiological signals, and how this can be used in QoE assessment.

This report describes the design and implementation of a system suitable for conducting an experiment investigating the level of immersion in VR storytelling content. This is manifested in a test hypothesis:"*The level of immersion is related to the measured heart rate in a way that the heart rate increases when a participant's level of immersion increases.*". The system consists of a HTC Vive VR , Steam VR Engine and a Fitbit Charge 2 health tracker. To test the hypothesis the system is used in an experiment on 23 participants, playing through two different digital stories: Trondheim 360, used as a training session, and Everest VR, the main digital story. The latter consisting of 5 scenes with different content. While playing, participants heart rate is measured. After each scene, the participants answer questionnaires regarding their experience of that particular scene. The data gathered will then be analyzed in Matlab and IBM's statistical software SPSS.

The report consist of several parts. First, some theory, definitions, and basic knowledge needed to understand quality assessment of interactive digital storytelling is presented. Basic introduction to physiological signals and immersion is presented as well. Next, the experimental system design and methodology is described alongside a research protocol developed to answer the research questions below. Finally, the results are analyzed and discussed before being round up in a conclusion.

1.3 Research Questions

Q1: Which factors are contributing to the level of immersion in interactive storytelling?

Q2: Can physiological signals, e.g. heart rate, be used to evaluate the QoE of interactive storytelling (i.e. when using HMD in VR)?

2 Quality Assessment of Interactive Digital Storytelling

This chapter will present theory needed to understand quality assessment of interactive digital storytelling, and how to evaluate immersive experiences from users. It also includes background knowledge needed to design and perform the experiment described in chapter 3 and 4. The theory includes aspects regarding interactive digital storytelling by using HMD in VR, QoE, immersion, physiological signals, as well as related work in this research field.

2.1 Virtual Reality

In order to understand how digital stories can create immersive experiences in VR, some background theory on the topics of VR, HMD systems and how they work are needed.

2.1.1 What is it?

VR is a computer simulated environment that often tries to replicate the real world. It opens up the possibility for creating quite immersive interactive stories. With recent VR systems such as the HTC Vive[1], it is easy to create free-view productions, with the possibility to interact with the content, as well as participate in a story or game.

In addition to VR, or some might say in opposition to, there also exist Augmented Reality (AR) systems and productions. Contrary to VR, where you place a real world object into a virtual environment, AR takes virtual objects and place them in the real world.

When placed in a VR environment, in order to stimulate our senses in the most correct way possible, it is important to have good technology simulating them. HMDs, binaural techniques, haptic technologies and other sensory effects is crucial for a VR system to deliver fully immersed experiences.

2.1.2 Head-Mounted-Display

Head-mounted-displays(HMD) are used to display 3D graphics, making it possible with free-view applications in VR systems. For a HMD to work properly as a VR component, it needs to include some important features to provide immersive experiences:

- Low delay between head movement and display update, as well as accurate tracking of head movements.
- For the virtual environment to look as real as possible, the resolution of the display needs to be high.
- It needs to cover the whole field of view as it would be seen from a human perspective.

[1] https://www.vive.com/eu/product/

- It needs to be designed in a way such that the weight or fit does not become bothersome for the user. The user should forget that he/she is wearing a HMD.

In addition to the HMD, some VR systems come with hand controllers(for interactivity) and tracking sensors(for user location and head-tracking in space) as described in the next section.

2.1.3 Sensors in Virtual Reality Systems

Sensors are essential when creating the interface a user will need to interact and communicate with a VR system. The sensors could be microphones to do speech recognition and detection, tactile sensors like hand-controllers to control objects and make decisions, and sensors capable of doing motion tracking. Motion tracking is essential for a VR system to retrieve the real-time position of the user, mapping it to the virtual environment. The most common types of motion tracking uses infrared or visual cameras.

2.1.4 Other Important Aspects

Haptic Technologies recreates the sense of touch for the user by applying forces, vibrations or motions through the system in use. This mechanical stimulation can be used to assist in the creation of virtual objects in a computer simulation, to control such virtual objects[1].

Binaural Techniques refers to the techniques used to recreate sound from a 3D-perspective. It is normal to use *head-related transfer functions* (HRTF) for this purpose. This function is unique for each person, containing filter coefficients reflecting head-size, shape etc. Filtering sound with this function can simulate sound from a given point in space, thus represent sound in virtual environments[2].

Other ensory effects includes machines and artifacts added to a system to evoke sensory emotions and feelings from aspects such as heat, wind, odour etc.

2.2 Digital Storytelling

2.2.1 What is it?

Digital Storytelling is a relative new and broad term, where it can cover everything from digital techniques used to tell a narrative(interactive stories in VR, web-based stories, narrative computer games etc.), to describe how everyday people use digital tools to tell their story(YouTube, blogs etc), and as a modern extension to the traditional and ancient art of storytelling[3][4]. The term is also widely used to describe digital learning tools for educational purposes[5].

This thesis will use the first description of the term, using digital techniques and tools to tell a narrative. Or more specific; using HMD in VR to tell a story. This can include 360degree tours in an environment of real world still pictures, or experiences in virtual environments created by game engines like Steam VR.

2.2.2 Interactive Storytelling in Virtual Reality

Interactive storytelling is a form of digital storytelling in which the storyline is not predetermined. In an interactive story, the user experiences a unique story based on their interactions with the environment. As this thesis focus is on telling interactive digital stories by HMD in VR, we divide the options for interactivity in two:

- **Passive use of HMD/VR system** - The user of the HMD/VR system sits or stands still in the real world, and can only observe and look around in the virtual environment.

- **Active use of HMD/VR system** - The user of the HMD/VR system can walk around in the real world, with the movements being reflected in the virtual environment. Controllers or gloves may also be used to interact with content.

The latter gives rise to a more immersive experience, with the user being able to simulate more real world features. A system like HTC Vive VR can provide such an environment.

2.3 Quality of Experience

Quality assessment of technology has the last 10-15 years start to switch from the delivery-focused area Quality of Service(QoS), to a more user-focused assessment field, namely the QoE. QoS is defined as the characteristics of a communication system to satisfy the users stated and implied needs[11]. QoE considerably extends this term (see Fig. 2.1), adding additional factors affecting a user's perception of quality. This can occur in situations where a system is difficult to use, learn, or the content is not applying or relevant for the user. In cases like this, the QoS can be high, but the end user might experience low QoE.

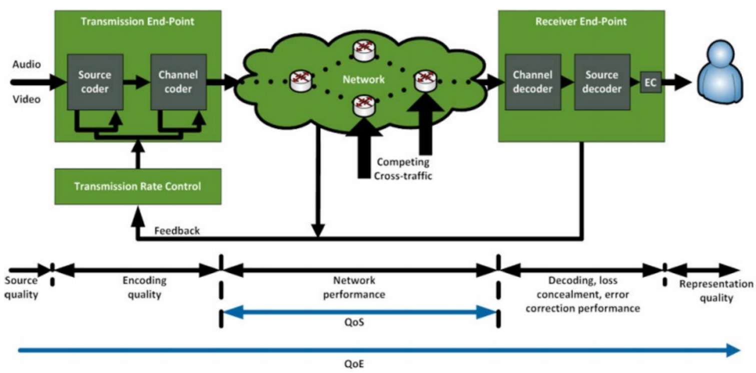


Figure 2.1: The relationship between QoS and QoE in a networked media handling system. [13]

2.3.1 What is it?

QoE is a fast emerging multidisciplinary field based on social psychology, cognitive science, economics, and engineering science, focused on understanding overall human quality requirements [12]. One applying definition of QoE was given by Qualinet[2]:

”**QoE:** *The degree of delight or annoyance of the user of an application or service. It results from the fulfillment of his or her expectations with respect to the utility and or enjoyment of the application or service in the light of the user's personality and current state.”*

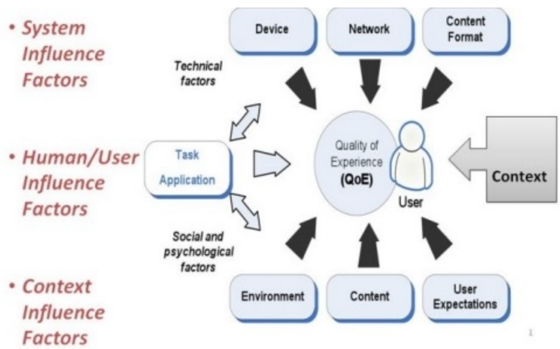


Figure 2.2: Factors influencing the QoE. [13]

In other words QoE is a subjective opinion about a system's performance, or the level of satisfaction a user gets from an application or service. This means that it is influenced by many factors, often split into human, system and context factors [11](as seen in Fig. 2.2).

Human influencing factors is any variant or invariant property or characteristic of a human user. They are factors decribing demographic and socio-economic background, the physical and mental state of the user, or the user's emotional state. These factors may influence how a user process, analyze or judge the experience.

System influencing factors refer to properties and characteristics that determine the technically produced quality of an application or service. They are factors like coding, transmission, storage, sampling etc. These factors are typical for measurement of QoS, which indicates that QoS is indeed included in QoE. These factors influence the user perception of a system.

Context influencing factors are factors describing how the setting or situational property influences a user's environment in terms of which it gives the user certain expectations. These factors include cost of product, comparison to similar systems, frequency of use, time of day and task. These factors also influence how a user process, analyze or

[2]European Network on Quality of Experience in Multimedia Systems and Services [11]

judge the experience.

Fig. 2.3 shows a detailed overview of the formation process in how we create QoE in our human minds. Notice how the different influence factors are influencing the expectations of a user, how a user perceives the experience and finally how the experience is judged.

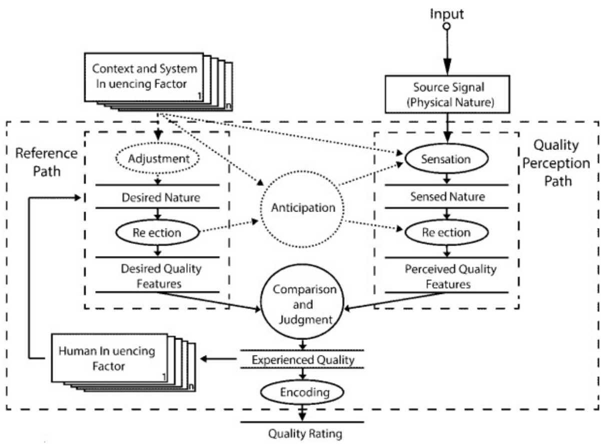


Figure 2.3: Quality formation process. [13]

2.3.2 Measuring the QoE

QoE is an important factor in modern technology systems and applications. Thus, it is important to evaluate and measure it. Traditionally, QoE has mainly been assessed by using post-test questionnaires gathering users subjectivity through their comments, opinions and thoughts[15]. These are normally quantified with the use of Mean Opinion Scores (MOS)[16] using a Likert scale[17]. When using a Likert Scale, the user rates each question on a five point scale where 5 = strongly agree, 4 = agree, 3 = neutral, 2 = disagree, 1 = strongly disagree. Other odd numbered n-point scales like 5 = excellent to 1 = bad can also be used.

When taking into account the nowadays extended use of affordable new display technology such as HMD in VR, new concepts and methods needs to be considered in addition to the subjective assessment methods. This is needed as new types of content and interaction systems are being created. One of the key challenges on this topic is how to assess and objectify the QoE, without directly interfering with the experiencing part.

In addition to the subjective assessment of QoE, there is an emerging trend to objectify QoE by capturing human physiological and cognitive information [18]. Since QoE can be closely related to the *sensory experience* term[19], new concepts in the research field of QoE use physiological signal responses for quality assessment. In QoE assessments, next to self-assessment by questionnaires, measurement of changes and variation in physiological data can then be used to evaluate the emotional response of a user.

2.3.3 Emotions in QoE Assessment

Most people have little problem recognizing and identifying when we are having an emotion. However, emotion is a very difficult concept in psychology to define, hence it exists several definitions of it. However, one dimension often used to structure emotion in is in the terms of *quality*[20], ranging from pleasure to displeasure. Emotional impact of multimedia content being experienced by a user can then be used as a measure of *quality*.

When considering QoE assessment, the factors influencing QoE are closely related to the user's emotions. It is mentioned in [20] that there exist at least two ways emotions are affecting the QoE:

- First, a stimulus may evoke(among other things) an emotion in the recipient, and the QoE researcher intends to assess this emotional impact.

- Second, a stimulus may cause an emotion in the recipient due to its meaning, making the emotion influence other relevant parameters. The QoE researcher needs to figure if the stimuli can be considered as emotionally neutral, or measure its emotional connotation for later analyses.

When quantifying the level of emotional stimuli in an end user, it needs to be captured. As mentioned in section 2.3.2, physiological signal responses can be used for this purpose, with the use of parameters like the variation in heart rate.

2.4 Physiological Signal Responses

In order to understand how physiological signal responses are related to, and can be used to quantify our emotional responses more objectively, we need some basic knowledge regarding the topic and how we measure it.

2.4.1 What is it?

Physiological signal responses are measurements of bio-physiological responses gathered from sensors placed on the human body. The sensors convert the physiological data from analog to digital presentation. The data is often transformed to an alternate domain, such as the time-frequency domain, where important info can be extracted and analyzed. Variations in the physiological signals will occur when our bodies experience emotional situations. In these situations, the body would react through sweat, increased heart rate, blushing, facial expressions and so on.

Physiological signal responses from the human body can be represented as several different parameters. Some of these are shown in Fig. 2.4.

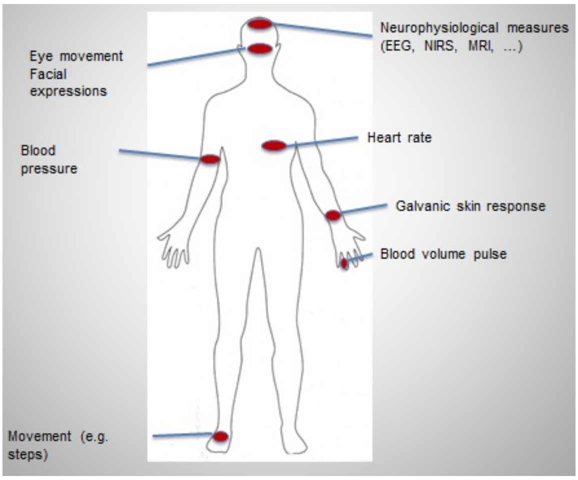


Figure 2.4: Measurable parameters extracted from physiological signal responses. [21]

2.4.2 Measuring Physiological Signals

In order to measure physiological signals, we need good methods, equipment and technology.

Biophysiological methods to capture parameters from physiological signals exists in several ways. Neuro-physiological measures(ECG[3], NIRS[4]), as well as measures of the peripheral autonomic nervous system(skin conductance, heart rate variability etc.) are among the most common. The latter is what is of most relevance for this thesis since it provides the connection between internal/external stimuli and the central nervous system. This is what allows the body to respond to its environment. [22]

Wearable sensors are technology equipment attached to the human body used to gather physiological data. This can be either medical apparatus or wearable consumer products like smart watches or health bands.

When it comes to wearable consumer products, HR is the most common parameter. From HR data we can obtain information regarding the level of stress, physical activity, and an estimation of how well trained someone is.

Optical tracking are used in commercial smart watches like those available from Fitbit[5]. This makes it possible to track HR without using additional equipment like a ECG-belt on the torso. When the heart beats, the capillaries expand and contract based on blood volume changes. LED lights (usually green light) on the tracker reflect onto the skin to detect blood volume changes. Finely tuned algorithms use this to measure HR.

[3] ECG=Electrocardiogram
[4] Near Infrared Spectroscopy
[5] https://www.fitbit.com/no/store

2.4.3 Physiological Quality Assessment

The most widely used quality assessment methods today uses subjective quality evaluation which is retrospective and intrusive, and signal based quality estimation which give no insights to the user. By using physiological quality assessment instead, these problems can be avoided. Using signals directly from the human body becomes less intrusive with the user.

Directly measurable physiological signals used in quality assessment can be ECG and Electrodermal Activity(EDA). These are good candidates since both heart rate(ECG) and stress level(EDA) increases with highly arousing content[24][25].

Peripheral measurable physiological signals can also be used in quality assessment by using a Self-Assessment-Manikin(SAM) scale[26]. This is a nine point scale using manikin figures with different expressions, where the user rates 3 questions in terms of arousal, valence and dominance. This is however rather similar to a post-test questionnaire, and are not eliminating the intrusive part of the assessment problem.

Limitations of physiological parameters are several. Measuring them are time-consuming and expensive, while the signal is often difficult to analyze. The amount of data often becomes huge and noisy and a lot of sensors is not robust against movement, making measures like EDA or EEG demanding the user to stand or sit still. A lot of apparatus are heavy and impractical as well, making the level of intrusiveness impact user QoE. One exception is smart watches using optical tracking.

However, the future looks bright as a lot of the limitations seems to be solved, with a lot of new, wearable medical and consumer products on the way[33].

2.5 Immersion

If physiological signal responses are to be quantified and used as a parameter for QoE, it needs to be verified through a measurable subjective parameter reflecting the same matter. Users perceived *level of immersion* can be used for this purpose.

In order to understand how physiological signal responses are related to the level of immersion, we need to know more about the definition of immersion in technology terms and how to use it as a measure of QoE.

Just like the earlier mentioned term of *emotion*, *immersion* is a term everyone knows what is, but finds it hard to define. Immersion has long been used as a describing term in culture areas such as museums[6], theme parks[7], and art[8]. But it does not have a well agreed upon definition to be used for technology in general. Immersion is also often confused with the notion of *presence*, so to define the concept of immersion as it can be described and used in technology we first need to distinguish it from the presence term.

2.5.1 Difference Between Immersion and Presence

Both McMahan[10] and Slater[9] among others have proposed definitions of immersion distinguishing it from presence, both concluding that *presence is a human reaction to immersion*. For the purpose of this thesis, Slaters definition is a good starting point:

”**Immersion:** *What the technology delivers from an objective point of view. The more that a system delivers displays(in all sensory modalities) and tracking that preserves fidelity in relation to their equivalent real-world sensory modalities, the more that is is immersive.*”

But how is this connected with emotions? With the above definition in mind, one might say that presence is more about *form*, and distinguished from emotional responses evoked by immersive experiences which has to do with the *content* itself. For instance, an expression like ”The VR environment was very real and made me feel like I was actually *there*.” is a sign of presence. An expression like ”The VR environment was very real, but it was really boring and uninteresting in a way I lost interest after a while.” is more a emotional response to how immersive the content was. The latter creates a link between emotions and immersion which is of great interest when wanting to capture and measure the QoE of a digital story, as emotions are closely related to physiological responses.

2.5.2 Quality Assessment of Immersion in Storytelling

When it comes to immersive storytelling, Zhang, Hoel and Perkis[14] defines being immersed in a story as how one or more features of the story are engaging to the extent that you are mentally absorbed into that virtual environment. Having in mind that presence is a human reaction to immersion, Zhang, Hoel and Perkis also defines being immersed as:

”*Being immersed means that we are mentally preoccupied by that story experience with a sense of realness and sense of presence, the very duality embedded in the definition of ’immersive experience’.*”

With this in mind is it clear that immersion is a major quality factor when it comes to determine the success of the experience and story. A goal would be to contextualize and quantify the quality of immersive experience in processes where users interact with a digital story, e.g. when using HMD in VR. Fig. 2.5 shows a framework for measuring the quality of immersive experience in storytelling.

One of the human factors highlighted in Fig.2.5 is physiological parameters. As Singer[34] have established a link between physiological arousal and emotional states, we can then define and describe content in terms of how immersive and arousal it can be for users. By combining variations in physiological measurements (e.g. changes in heart rate or EDA) with questionnaires capturing system and design factors, one might be able to quantify and measure the QoE of different content in an interactive digital story. If these combined measures correlates in terms of how immersive and arousal different content is described, then there exists a good chance that QoE of an interactive digital story can be assessed through an experiment covering these factors.

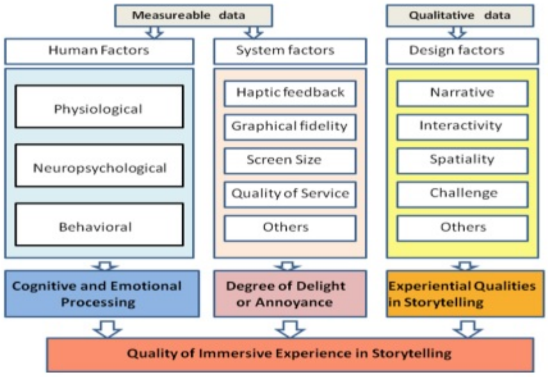


Figure 2.5: Framework of Quality of Immersive Experience in Storytelling. [14]

2.6 Related Work

Physiological QoE assessment is already widely accepted. However, previous research on this topic has focused on how physiological measurements can be used to assess QoE using standard displays, and how the addition of several sensory effects increases this(fan blowing wind, a dog sneezes you in the face etc)[27][28][29]. However, the research has a gap[30] when it comes to using physiological measurements for assessing QoE in new display technology, such as HMD in immersive VR environments.

However, there has been recent experiments[31][32] in this research field. An experiment on evaluation of heart rate and electrodermal activity as immerse parameters for QoE in VR environments when using HMD has been conducted, though only with a passive system (see section 2.2.2) with lack of interactivity. A main obstacle has been the lack of good affordable and wearable sensors that are robust to activity and movement. In the two experiments, participants had to sit still and not move their hands for the sensors to work properly. This will of course limit the level of immersion to some point. Adding the possibility to move around can increase the level of immersion, hence give researchers a better data basis.

To the best of the authors' knowledge, this is some of the very first work that uses physiological parameters to compare user QoE in different content of interactive storytelling, particularly with the focus on using HMD in VR.

3 Experiment System Design

This chapter describes and discusses the implementation of a system suitable for conducting an experiment that aims to answer the research questions given in chapter 1. Fig. 3.1 sketches out how such a system can be designed.

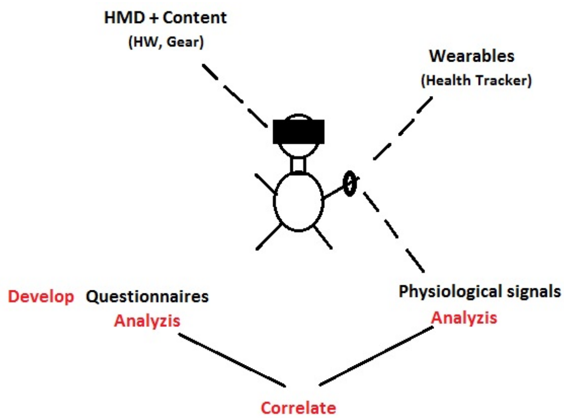


Figure 3.1: Illustration of an experiment system design used to assess QoE in digital storytelling by using HMD in VR.

3.1 Measurement and Analyze Procedure

The experiment process and procedure illustrated in Fig. 3.1 can be described in the following way:

1. Combine HMD + High quality VR content.
2. Track physiological signal responses from participants with the use of wearable sensors.
3. Develop subjective post-test questionnaires to be answered by participants.
4. Analyze raw data from physiological and subjective data.
5. Identify content evoking increased level of immersion in participants.
6. Identify possible correlation between physiological and subjective data.

With the following in mind, it is clear that a state-of-the-art HMD/VR system delivering high quality content is needed. Wearable sensors, as well as thorough developed questionnaires, are also essential to conduct an experiment of this form.

3.2 Devices - Hardware and Software

The following subsections will describe the technology, devices and software needed to conduct an experiment as described in the section above, and how to gather physiological and subjective data from participants taking part in it.

3.2.1 Virtual Reality System - HTC Vive VR

It was desired that the content to be used in the experiment should be presented in VR. Therefore, it is best to use a HMD and system made specifically for it. There is a lot of displays on the market satisfying all the important features of HMD given in section 2.1.2, where the leading ones are Oculus Rift and HTC Vive VR. They often come out equal in tests on the terms of amount of content available, or resolution/graphics[35]. The difference is that Vive VR comes with infrared sensors giving the opportunity to use the whole experiment room as an interactive arena/environment. Being able to move around and interact with the content is most certainly important in creating highly immersive experiences, hence making the HTC Vive the best choice for an experiment like this. The area used for this experiment covered approximately a space of 3x4m.

Vive VR is a collaboration between HTC and the Valve Corporation, most famously known for the Steam[1] software. Steam features Valve's SteamVR technology, allowing developers to create VR projects with Unreal Engine 4[2] for the HTC Vive.

System parts:

- Pair of Hand-controllers used for interaction and navigation. Also used to simulate hand-gestures.
- Pair of base stations (infrared sensors) follows with the system, hence enabling 360 degree room-scale motion-tracking. They also include 2 power cables and adapters, 2 mounting kits and 1 sync cable.
- HMD with audio-headset enabling 360 degree free-field view and 3D-sound. Also includes attached 3-in-1 tether and link Box.

In addition to the system parts that follows with the HTC Vive VR System, a PC with an additional screen is required to run the system. The PC needs to fulfill some specific requirements.

System requirements:

- GPU: Nvidia GeForce GTX 970, AMD Radeon R9 290 equivalent or better
- CPU: Intel i5-4590, AMD FX 8350 equivalent or better

[1]a digital distribution platform that offers digital rights management (DRM), multi-player gaming, video streaming and social networking services.

[2]The Unreal Engine is a game engine developed by Epic Games, where Unreal Engine 4 is the most recent software released and used for VR systems.

Figure 3.2: Overview of system parts included in the HTC Vive VR. [36]

- RAM: 4 GB or more
- Video Output: HDMI 1.4, DisplayPort 1.2 or newer
- USB Port: 1x USB 2.0 or better port
- Operating System: Windows 7 SP1, Windows 8.1 or later, Windows 10

In addition to include the required specifications mentioned above, the PC need to install the SteamVR software. This will be used for setup and calibration of the Vive VR system parts, as well as it is used to buy and download VR content and productions. Downloaded content will also run through this software.

Setup of HTC Vive through SteamVR:

1. Navigate to `https://support.steampowered.com/kb_article.php?ref=2001-UXCM-4439#room-pick` and follow the thorough installation instructions. The following points summarize the most important parts.
2. Pick a suitable room or space for Room Scale VR. It needs to be at least 2m x 1.5m and clear of any obstacles. The maximum distance supported between base stations is 5m.
3. Place the base stations at a high position in the corners of the room or space used to define the limits for real world movements in the virtual environment. The sensors need to ”see” each other. Cable sync is also possible.
4. Connect the HMD with the computer through the link-box and attached tether with HDMI- and USB-cables.
5. Install and download Steam and SteamVR through `http://store.steampowered.com/about/`. An user account is required. Once SteamVR is installed, set it to SteamVR Beta and run the software.
6. Install base stations and HMD as described in the installation manual.

7. Connect the controllers. The controllers will automatically pair to the HMD when turned on for the first time.

8. Finally, use the PC desktop menu to run Room Setup and SteamVR Tutorial as shown in Fig. 3.3. This will provide a guide on how to calibrate the room in use.

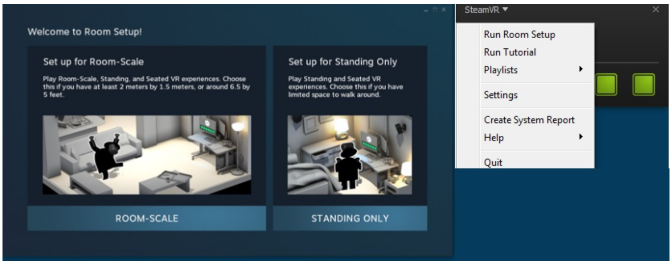


Figure 3.3: Screenshot of Room Setup instructions in SteamVR

When the HTC Vive VR is up and running, VR content like the digital story *Everest VR* can be downloaded from Steam Store.

3.2.2 Wireless Health Tracker - Fitbit Charge 2

Suitable sensors are needed to track and measure physiological data. The sensors will have to be robust against movement and interactivity. There are several wearable consumer products available satisfying this, though most of them only track heart rate. Also, few of these give easy access to raw data as time series of measured heart rate. The health tracker used in an experiment like this will then need to fulfill this description: Track a user's heart rate, with the possibility of open and easy access to the measured data.

Fitbit Charge 2 (see Fig. 3.4) is a good choice, as it is rather cheap, and it has an open web-API[3] which makes it possible to retrieve heart rate data. However, this web-API requires the user to hold knowledge about web programming. The device measures the heart rate optically using green light.

[3] API = Application Programming Interface

Figure 3.4: Fitbit Charge 2 smart watch. [37]

System parts:

- The watch: The main component and used as a display for heart rate, time etc. Gives access to the user interface, with the option of choosing work-out or run mode which tracks the heart rate second by second.
- Mini-USB cable: Used to update the watch, as well as charge the battery between usage. The watch should be connected to the cable between experiment use.
- Android or iOS mobile application: Needs to be downloaded on a smartphone. Required to synchronize measured data from the watch to the web.

Setup of Fitbit Charge 2 for experiment use can be performed in this order:

1. Unbox the watch and charge it fully the first time.
2. Navigate to `https://www.fitbit.com/no/setup`
3. Download the Android or iOS mobile application and run through the setup instructions.
4. Remember to update the watch to the most recent software. Should be done automatically, but can be done manually in the mobile application.
5. Attach the watch to the wrist and make sure that it reads the heart rate.
6. Navigate to ”Run Mode”(see Fig. 3.5) by tapping the watch. Press and hold the button on the side to initiate run mode. This tells the watch to track the heart rate every second(it tracks randomly otherwise).
7. The watch is now ready to be used for heart rate data tracking in an experiment. Next paragraph describes how to retrieve the data.

Figure 3.5: Illustration photo of run mode option in Fitbit Charge 2.

Connection to Fitbit web-API is needed to retrieve data on a format it can be analyzed. It is available through what is called authorized HTTP-requests[4] to the company's data servers. Note that this part requires insight on how a web application works, and how to use HTTP-requests to communicate with it.

Heart rate data from a unique watch can be retrieved by following this procedure:

1. Navigate to Fitbit's web-API documentation `https://dev.fitbit.com/docs/`. Review the basics on how the Fitbit Web API works, and follow the quick start manual.

2. Register your application by following the manual to get API client credentials(essential to be authorized). You will need a Fitbit account (free) to register an app. When registered, you'll get an unique authorization key(Access Token).

3. To get access to the data, you'll need to program and implement an OAuth 2.0[5] authorization flow for user authorization and API authentication. This require knowledge about web programming, javascript and JSON. More on this step can be found here `https://dev.fitbit.com/docs/oauth2/`. Code managing this can be found in Appendix D.

4. When the authorization flow is running and connected to the server, it is time to send HTTP-requests to the web-API. The easiest way to do so is by using a program like Postman, which is a tool program for API developers. Download it from here: `https://www.getpostman.com/`

5. Navigate to Fitbit's documentation for developers[6] on how to send requests regarding time specific heart rate data.

6. Open Postman, and run time specific requests for heart rate data as done in Fig. 3.6.

[4]HTTP-requests are the type of requests which your computer sends to the web server containing all sorts of (potentially) interesting information.

[5]OAuth 2.0 is the industry-standard protocol for authorization. `https://oauth.net/2/`

[6]`https://dev.fitbit.com/docs/heart-rate/#get-heart-rate-intraday-time-series`

7. The heart rate time series are now available for download on a JSON[7] format. When analyzing the data in a program like Matlab, it needs to be converted from JSON format to Matlab-objects. Code managing this can be found in appendix E.

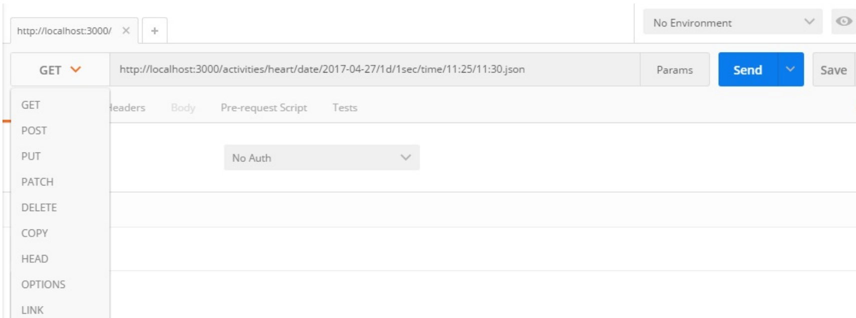


Figure 3.6: Screenshot of HTTP-request to Fitbit web-API in Postman.

3.2.3 Web-based Questionnaires - Google Forms

Subjective data needs to be gathered from post-test questionnaires. Google Forms can be used to create and manage questions in a questionnaire, as well as it is a neat way to handle the data from the results. Register and follow the instructions at `https://www.google.com/forms/about/` on how to use it.

Registered answers can be downloaded as CSV-files[8]. When analyzing questionnaire data, it also needs to be converted from CSV-format to Matlab-objects. Code managing this can be found in appendix E.

3.3 Stimulus - The Digital Stories

In addition to the specified technology, software and devices, the experiment will need immersive content from high quality VR productions. SteamVR can provide different type of content through Steam Store. The content need to satisfy the description of digital storytelling presented in section 2.2.1:

”**Digital Storytelling:** *is the way of using digital techniques and tools to tell a narrative. Or more specific; using HMD in VR to tell a story. This can include 360degree tours in an environment of real world still pictures, or experiences in virtual environments created by game engines like Steam VR.*”

Two VR productions are to be used in this experiment. The participants are set to run through *Everest VR*, an experience/digital story told through several immersive scenes.

[7] JSON (JavaScript Object Notation) is a lightweight format that is used for data interchanging. It is based on a subset of JavaScript language

[8] A Comma-Separated Values (CSV) file stores tabular data (numbers and text) in plain text. Each line of the file is a data record consisting of one or more fields separated by commas.

Before doing this, a training session with a production called *Trondheim 360* must be conducted. Both productions are played through SteamVR, with the use of the HTC Vive VR system. The different content will be described in terms of interactivity and immersion in the following sections. .

3.3.1 Training Session - Trondheim 360

Participants will need to conduct a training session with a follow-up questionnaire. The procedure will be similar to the one used in the main part of the experiment. This will help participants get comfortable with experiment setup and equipment, as well as removing influences from first time impressions of post-test questionnaires. A production called *Trondheim 360* made by Dr. engineer Kryzstof Orleanski at the Department of Electronic Systems, NTNU, serves as content for the training session.

In *Trondheim 360*, the user stands on a platform overlooking and navigating through a set of 360 degree still images of Trondheim city, Norwegian nature, and other cultural sites from the region around Trondheim. Not expected to be very immersive, as the pictures and people in it stands still. There is also no other interactivity than the option of looking and moving around on the platform.

3.3.2 Main Session - Everest VR

Everest VR is used in the main session of the experiment, serving as the high quality content to be evaluated. *Everest VR* is a digital interactive story produced by Icelandic Sólfar Studios. It is an experience that allows the user to get a feeling on how it is like to climb Mount Everest. It do so in a series of first person challenges displayed in 5 iconic scenes, an experience lasting for approx. 60-90 minutes depending on the user.

The scenes will be used as the different type of content to be evaluated in terms of the level of immersion, as measured by variations in physiological measurements from participants. In the listing below, each scene will be described in terms of how immersive(as immersion is described in the theory section) the content is.

- **Scene 1 - Base Camp:**
  Base Camp is where the journey begins. After a quick demo on how to use the controllers, the participant ”wakes” up in front of an altar in which Sherpas give offerings to mountain gods ahead of the perilous journey. A table with all the equipment needed for the journey is also present.

  As this is an introduction scene, it is not very exciting in terms of interactivity and tasks. There is stuff laying around that you are able to pick up with the controllers, as well as walking around between the altar, table and tent. A participant will be able to place items on the altar as an offering for good luck on the journey ahead. This is expected to be experienced as one of the least immersive scenes.

- **Scene 2 - Khumbu Icefall**
  The scene involves a treacherous trek across the Khumbu Icefall glacier(see Fig. 3.7). A participant will have to cross large crevices with ladders the Sherpa have

placed. Once safely on the other side, a ladder needs to be climbed to get off the glacier.

This can be described as a very exciting scene. People with fear of heights will definitely feel their heart beating. The participants will have to balance across a ladder only with the safety of grabbing on to some loose ropes. Looking down, there is what can be seen as an endless abyss. Also, the participants will have to climb a ladder to get up and off the glacier, grabbing one step at the time while large pieces from the glacier falls down on each side. This scene is expected to be experienced as the most immersive and arousing compared to the others.

(a) A climber getting ready to cross the glacier in real life. (Photo: Alan Arnette)

(b) A participant crossing the same glacier in Everest VR. (Screendump: Sólfar Studios)

Figure 3.7: Khumbu Icefall - Real world vs VR

- **Scene 3 - Lhotse Face**

  This scene presents the ascent of Lhotse Face. Rising between Camp 2 and Camp 4, Lhotse Face is a nearly vertical wall of ice 1,500m (5,000 feet) high ending in the infamous Death Zone. In this experience the participant will have to climb from Camp 3 and up to Camp 4 along with a group of climbers.

  This scene is probably the most exhausting for participants, as it is by far the most time consuming. In terms of interactivity and immersion, the participants will have to climb a wall of ice with an ice axe movement for a long amount of time. The level of immersion can be ruined a bit, as the participant is not allowed to proceed in their own pace, as well as the time consume can make it boring after a while. How ever, this can still be seen as more exciting than for instance scene 1 and 5.

Figure 3.8: Climbers ascending the Lhotse Face in Everest VR. (Screen dump: Sólfar Studios)

- **Scene 4 - Hillary Step**
  The scene recreates the ascend of the near-vertical 12m rocky outcrop named Hillary Step which is located on the mountain's southeast ridge, and is the last great challenge before the top.

  In this scene, participants will balance on the ridge while securing themselves. The ascent is done with the assistance of fixed ropes, where the participant will have to secure themselves by attaching a carabiner on and off. Compared to scene 2 this is also exciting in terms of the height and view. Balancing on the ledge can indeed evoke emotional responses in a participant. How ever, this part of the scene is very short in time and in that case becomes less spectacular and exciting than scene 2.

(a) Climbers waiting in line at the Hillary Step. (Photo: STR/AFP/GETTY IMAGES)

(b) Screen shot of Hillary Step scene in Everest VR. (Screen dump: Sólfar Studios)

Figure 3.9: Hillary Step - Real world vs VR

- **Scene 5 - Everest Summit**
  The final scene starts after ascending the Hillary Step, and ends at the summit of Everest. After traversing the last slope, the top is reached.

  Besides planting the Everest/Solfar flag on the summit, the scene does not open for a lot of interactivity. Because of the long range of the slope, one will have to "jump" forward by using the controllers to reach the summit , making the scene quite boring in terms of interactivity. This scene is expected to be perceived in a same matter as scene 1, and less immersive than the three scenes in the middle.

Figure 3.10: Planting the flag on top of the summit in Everest VR. (Screen dump: Sólfar Studios)

From the descriptions above, the scenes level of immersion are expected to be perceived from high to low in this order: Scene 2 - Scene 4 - Scene 3 - Scene 1 - Scene 5, where a big difference is expected to be between scene 1/5 and scene 2. This is evaluated in terms of the interactivity and tasks available in each scene.

4 Methodology

Methods and standards to follow in the experiment will be described and discussed in this chapter, as well as a description of the research protocol used for the experiment. All of this is important in order to get reliable results that answers the research questions presented in section 1.3.

4.1 Standards and Recommendations

Standards and recommendations are methodologies used by researchers to ensure a high quality study with reliable results. Standardization organizations like ITU and ISO provide standards that include a set of rules to follow when conducting experiments. Currently there exists no recommendations for evaluating the use of HMD in VR, or interactive multimedia content in general, suitable for this experiment. However, a combination of recommendations for sensory analysis and evaluation of multimedia content, as well as papers proposing evaluation methods in the same research field could be used. From this, a good fundamental base can be created for the methodology used in this experiment.

As a starting point, it was chosen to use the ITU BT.500[38] and ISO 8589[39] recommendations, as well as a paper on the "*Evaluation of heart rate and EDA as an objective QoE evaluation method for immersive VR environments*"[31]. ITU BT.500 is describing a methodology for the subjective assessment of the quality of television pictures, while ISO 8589 describes the lab setup when conducting an experiment including sensory analysis. The paper mentioned is the closest we come an experiment evaluating immersion in terms of physiological signals as a parameter for QoE of interactive content.

All the different parts of the recommendations will not be given here, but important deviations and features added specially for this experiment is listed below:

- Test of color vision and visual acuity is not needed as the experiment is not evaluating the images directly.
- How the test procedure is arranged. Single -or double stimulus does not make sense in this kind of experiment.
- Randomizing the test order of the content does not make sense, as the different content to be compared and evaluated consists of different scenes in a chronological told story. Including more stories would also make the experiment too time demanding.
- Because of the large time consume each test of the experiment takes, and to have a buffer for the data minimum of 20 participants, 25 was decided as a proper number of test participants.

To determine how the methodology proposed by the combination of the recommendations and paper above will be run and executed, a research protocol has been developed.

4.2 Research Protocol

A research protocol forms the basis for how a study is run. A well written and complete protocol is essential for a high quality study. The protocol is a structured plan following a certain structure. The research protocol can have different levels of detail, but is usually containing an introduction, the objective, a hypothesis, methodology and a schedule. The research protocol structure used in this report is found in appendix A.

Test hypothesis, resources, and the experiment methodology and procedure developed in the research protocol will be presented and described in the following subsections.

4.2.1 Hypothesis

A clear statement of primary and secondary objectives of the study should be summarized in a clearly defined test hypothesis. The hypothesis provides an answer to what conclusion we are expecting from this study.

Main Objective: *Physiological signals can be used as a parameter for the level of immersion when evaluating the Quality of Experience in digital storytelling by using head-mounted displays in virtual reality.*

To test if the main hypothesis holds, we define a secondary objective/hypothesis to be tested during the experiment.

Secondary Objective: *The level of immersion is related to the measured heart rate in a way that the heart rate increases when a participant's level of immersion increases.*

4.2.2 Resources

All the resources needed to conduct the experiment are listed below. All equipment was made available for the experiment at the Cafe Media lab(room B-133), NTNU Gløshaugen.

- 1 x HTC Vive VR System:
  - HMD w/ audio headset
  - 2 x HTC Vive controllers
  - 2 x Base stations(infrared sensors)
  - Linkbox and wire connectivity with PC/monitor
- 1 x HTC Vive ready(see section 3.2.1) PC with the following software:
  - Steam
  - SteamVR engine
  - Everest VR
  - Trondheim 360 environment
- 1 x PC with following web-based questionnaires:

  - Demography/Background
  - Content experience
  - Overall experience

- 2 x Fitbit Charge 2 (one in reserve)
- 1 x Video camera (For time synchronization of Fitbit)
- 25 x Participants
- 25 x Cinema tickets

4.2.3 The Experiment

The different scenes in the Everest VR digital story is used to compare different immersive content in terms of participants' variation in heart rate. For this purpose, both subjective and physiological measurements are employed. Also, a suitable room is needed for the experiment setup, as well as a thorough defined execution order for each test participant to follow.

Subjective Measurements - Questionnaires

Questionnaires will be used to collect subjective data of participants' perceived experience of the *Everest VR* digital story. The data will then be used to evaluate the QoE and amount of immersion in a traditional, subjective matter. The questionnaires are divided into three parts:

- **Demography/background questionnaire(Pre-test)**: Is performed once, before the experiment starts. Maps the participants background (demography, former experience, gamer/non-gamer etc).
- **Experience questionnaire (In-test)**: : Is performed after each experience/scene in the digital story, thus, five times for each test participant. The purpose is to cover the subjective experience. It consists of three parts:
  - **Quality**: Questions covering the perceived audio, video and overall audiovisual quality. Uses the Likert scale described in section 2.3.2 according to ITU when evaluating audiovisual quality in general[40].
  - **SAM**: The self-assessment mannequin (SAM) questionnaire covering a participant's level of valence, arousal and dominance[26]. This is used to evaluate the participant's emotional state..
  - **Usability/Immersion**: The aim of these questions is to evaluate two of the key aspects in user QoE for an immersive environment. These questions have been compiled by evaluating different prior work on immersion, including the paper by [31][32].
- **Final overall questionnaire (Post-test)**:): Is performed after the experiment is ended. It covers the usability/immersion for the whole experience of Everest VR, as well as comments and feedback from participants.

The audio/video questions are standard questions used in quality assessment, while the SAM questions is normally used to evaluate the peripheral physiological signal responses from users of a system. The Usability/Immersion questions are mostly gathered and inspired by the paper mentioned initially in this chapter[31], as well a different paper on ”*Virtual Environment Evaluation Questionnaire*”[41].

The questions and assertions used in the questionnaire can be found and more thorough described at the end of the research protocol, and are listed in Appendix C of this thesis.

Physiological Measurements - Heart Rate

Physiological measurements are to be performed by tracking participant’s heart rate with the use of a Fitbit Charge 2 health tracker. Heart rate data from the Fitbit can be accessed by connecting to the web-API as described in section 3.2.2.

The data is tracked by setting the Fitbit to initiate “Run Mode” when a participant enters a new scene. This allows the Fitbit to track data second by second. When the scene is over, “Run Mode” should be stopped. Time series of heart beats per minute from each scene can then be seen in the raw data as time segments where the data is logged second by second.

For evaluation, the heart rate data from each participant is divided temporally into 5 segments, one for each scene. Each scene segment then is again divided into 4 heart rate zones which are derived on a participants’ basis. Here, first the participant’s maximum and minimum detected heart rate was calculated, and then the range in between was divided into 4 equally large zones. Using these zones, the average time spent by each participant in the different zones can be calculated.

Experiment setup

After all the resources is acquired, and the measurement equipment is ready for use, the experiment can be setup.

Fig. 4.1 shows an illustration of the experiment setup. A large monitor is used by the test leader to mirror what the participants sees in the HMD’s VR environment. On the floor, the HMD attached to the computer lies, with one of the base stations seen on top of the pole. The camera was used to synchronize Fitbit data with the different scenes/stories performed.

Fig. 4.2 shows a participant performing one of the scenes in Everest VR while using the HMD and controllers from the HTC Vive. The test leader is holding the wires, avoiding the participant to trip in the attached wires. The test leader also watches the monitor to identify where the participant is located in the current scene.

Figure 4.1: Illustration photo of lab participant answering a questionnaire in between scenes of Everest VR.

Figure 4.2: Illustration photo of a lab participant performing one of the scenes in Everest VR. (The Fitbit is attached on the participant's right wrist, nearly not visible in this picture)

Execution Order

1. Introduction to the experiment. Participant read and sign introduction and consent form.
2. Answer background/demography questionnaire.
3. Attach Fitbit on participant and initiate "Run Mode". Also equip participant with HMD and controllers.
4. Start video camera (for time synchronization with Fitbit)

5. Perform training session.
6. Stop ”run mode”. Answer questionnaire regarding the experience.
7. Perform scene 1 in Everest VR. Again, initiate ”Run Mode” on Fitbit.
8. Stop ”Run Mode”. Answer questionnaire regarding the experience.
9. Redo step 7 and 8 for scene 2, 3, 4 and 5 in Everest VR.
10. Stop video camera. Answer questionnaire regarding the overall experience.
11. Participant receive cinema ticket as reward.

Each test has been estimated to last 60-75 min(time spent answering questionnaires and performing the different scenes included) depending on the participants speed. After each test participant, the data retrieved from the questionnaires and Fitbit should immediately be downloaded and stored in a secure matter.
multirow array

5 Results

This chapter presents the results from the experiment, obtained from the subjective and physiological measurements.

All bar- and time-plots are generated with the Matlab code listed in appendix F. The rest of the statistical analysis have been employed by using the IBM SPSS software (except the correlation and T-test which was conducted in Matlab).

5.1 Subjective Measurements

Subjective data was collected on different dimensions, as described in 4.2.3: experienced quality of the material, emotional response of the participant, as well as subjective reports on the level of immersion. All dimensions were assessed after the participant completed each of the scenes. Hence, five times. The average MOS-rating is presented as error-bar plots conducted in Matlab with a 95% confidence interval. Repeated measure ANOVA was used to identify statistical significant differences.

All results from this analysis can be found in appendix G and H, while the raw data can be found in the Matlab workspace file found in appendix E.

23 participants conducted the experiment, whereby 20 of the participants are used in the questionnaire and physiological data analysis (11 male, 9 female) with an average age of 24.9. Exclusion of participants' data is due to data from three of the participants being corrupted while converting it from .csv-format. Participants' former experience with HMD in VR can be seen in Fig. 5.1.

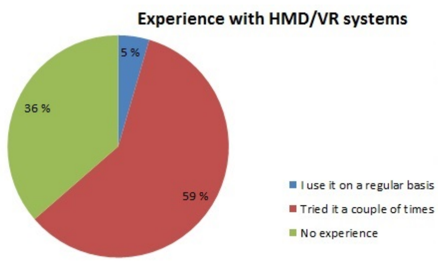


Figure 5.1: Earlier experience with HMD in VR

In terms of quality ratings significant differences between the scenes could be identified for the perceived overall audiovisual (F=2.42, $p \leq 0.1$) as well as for the video quality (F=2.32, $p \leq 0.1$). With scene 1 receiving the lowest scores, and scene 4 and 5 receiving the highest score in both case. No significant differences was found for the audio quality

(F=0.77, n.s.).

Question	F	Sign.
Valence	2,894	0,490
Arousal	3,010	**0,009****
Dominance	2,786	**0,001****
Overall AV	2,423	0,077*
Audio	0,777	0,538
Video	2,317	0,078*

Table 5.1: Statistical analysis of self-reported measurements from Quality/SAM. ** indicate $p \leq 0.05$, and * $p \leq 0.1$.

The subjectively rated emotional response of the participants on the SAM scales for valence was rated similar among all scenes, with no significant difference (F=2.89, n.s.). Arousal was rated highest for scene 2, and lowest for scene 1, with a significant overall effect (F=3.01, $p \leq 0.01$). Dominance was rated highest for scenes 4 and 5 and lower for the other scenes (increasing in score from scene 1, 2 and 3) (F=2.79, $p \leq 0.01$) (see Fig. 5.2).

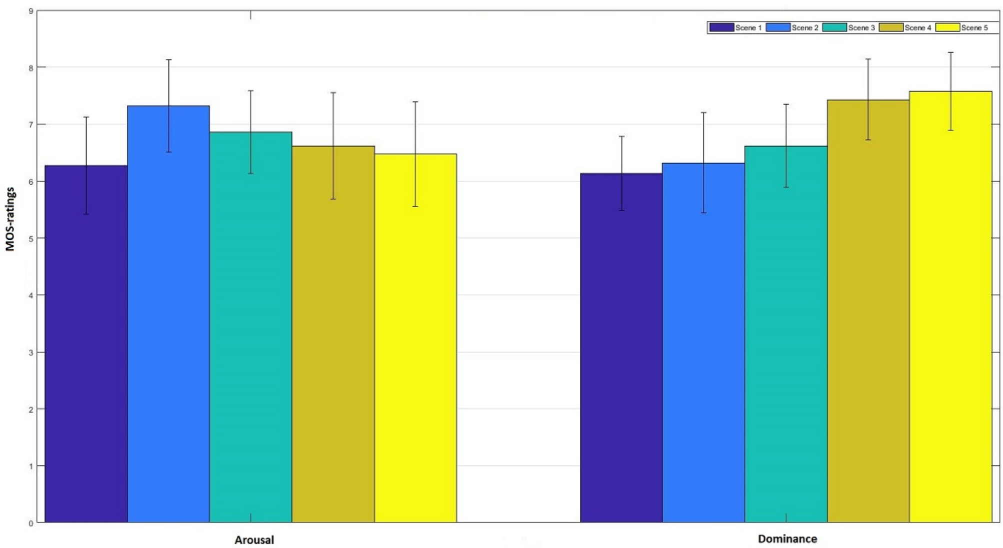


Figure 5.2: Self-assessment mannequin ratings (SAM) for arousal (left) and dominance (right). Bars indicate 95% confidence level.

Analysis of the subjectively obtained measures for immersion reveal differences between scenes for some of the questions. Fig. 5.3 shows a noticeable significance for

Question	F	Sign.
EXP 1	3,604	0,056*
EXP 2	3,312	0,093*
EXP 3	3,804	0,967
EXP 4	3,591	0,341
EXP 5	4,000	0,334
EXP 6	3,119	**0,026****
EXP 7	3,604	0,123
EXP 8	3,627	0,474
EXP 9	4,000	**0,001****
EXP 10	4,000	0,457

Table 5.2: Statistical analysis of self-reported measurements from Immersion/Usability. ** indicate $p \leq 0.05$, and * $p \leq 0.1$.. For more detailed descriptions of the questionnaire items, see appendix C.

question EXP 1 and EXP2, such that scene 3 received lowest ratings. Also, clear statistical significance can be found for questions EXP 6 and EXP 9, such that scene 2 received the highest scores. Details for statistical tests can be found in Table 5.2.

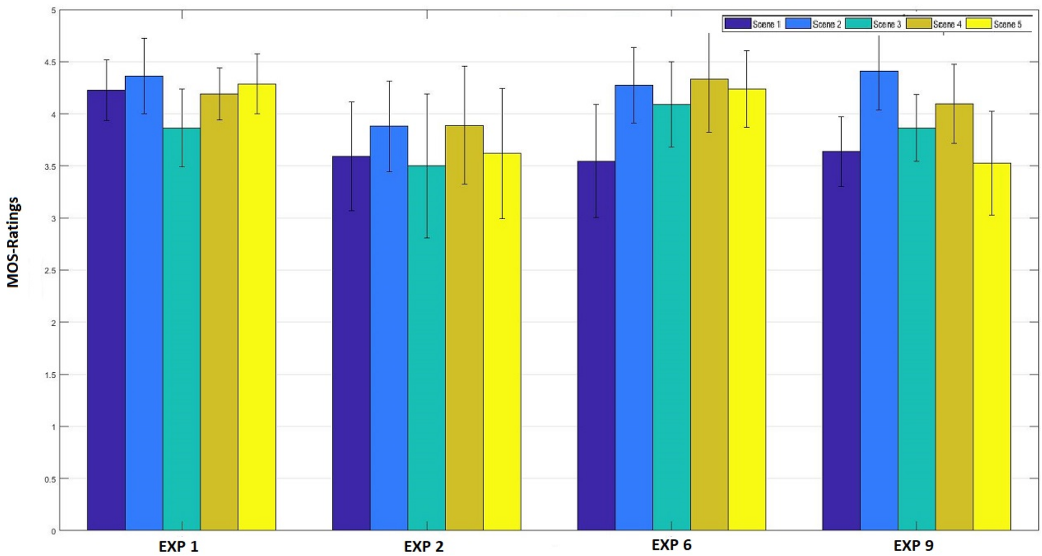


Figure 5.3: Averages of self-reports on EXP 1, EXP 2, EXP 6 and EXP 9. For more detailed descriptions of the questionnaire items, see Appendix C. Bars indicate 95% confidence level.

5.2 Physiological Measurements

The physiological metrics employed during this study was heart rate captured by a Fitbit Charge 2 attached to the participant's wrist. The raw data is presented in appendix I.

The captured data is divided into four equally divided HR-zones (from minimum to maximum registered BPM for each participant), and analyzed in terms of the percentage of time spent in each of these. Such that differences between the different HR-zones in each scene can be analyzed and then identify highly arousing content in terms of increased HR. A pairwise T-test in Matlab was used to conduct the statistical analyzes of the different HR-zones in each scene. Bar plots of time spent in each HR-zone for each scene can be found in appendix J.

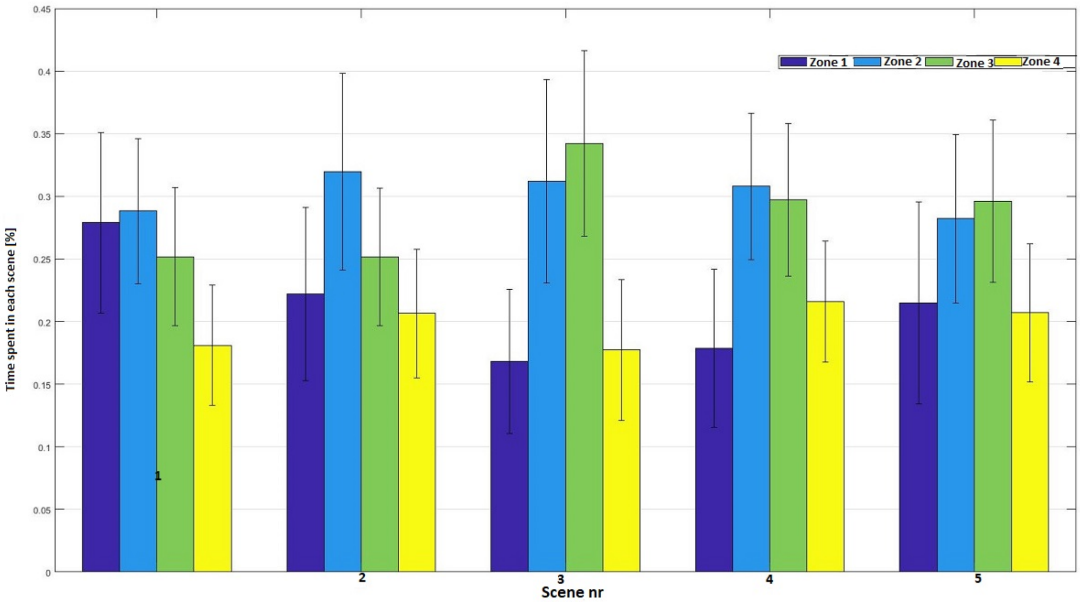


Figure 5.4: Bar- and error plot of percentage of time spent in each HR zone between scenes. Statistics performed with 95% confidence level.

The results of the analysis for the physiological data can be seen in Fig. 5.4 and 5.5. As can be seen participants spent more time in the lower heart rate zones for scene 1 (in particular Zone1). For the rest of the scenes participants spent of most of the time in zone 2 and 3. In scene 1, participants spent significantly less time in zone 4 than the rest of the HR-zones. Scene 2 does not show statistical significance in terms of time spent within the different HR-zones. In scene 3 and 4, most of the time is spent in HR-zone 2 and 3, resulting in statistical significance compared with zone 1 and 4.

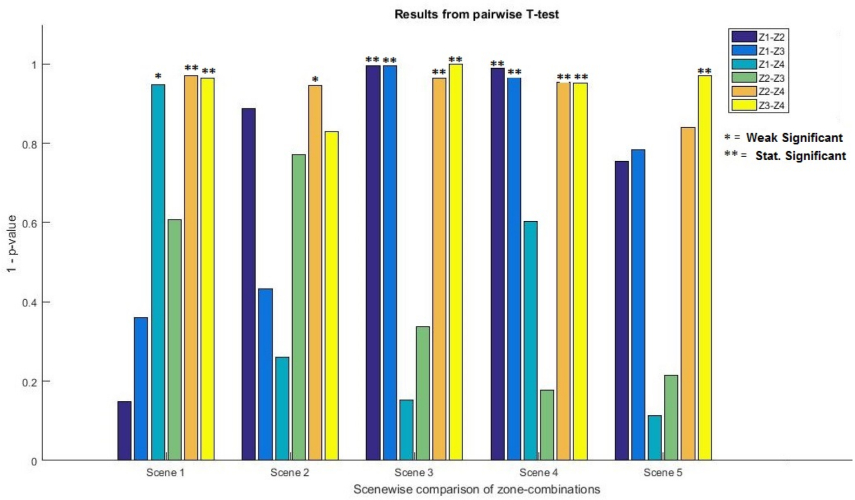


Figure 5.5: Results from pairwise T-test of the different HR-zones in each scene. Statistics performed with 95% confidence level.

5.3 Correlation Test

The strength of the relationship between HR measures and questionnaire experience ratings was examined using bi-variate and Spearman correlation tests. A correlation test between each of the statistical significant questions from section 5.1 was performed with each HR-zone from the physiological results.

The results from the correlation tests showed no statistical correlation between the subjective and physiological data.

6 Discussion

This chapter discusses and evaluates the results presented in chapter 4.2.3. A critics section on the experiment design, setup and methodology is also described, as well as a section on suggestions for further work.

6.1 Subjective Measurements

While looking at the ratings for EXP 1 (see Fig. 5.3) it shows that participants found scene 1 and 5 most immersive, but scene 3 the least immersive. This does not correspond with the results from the arousal ratings (Fig. 5.2), and shows that the level of immersion not necessarily correlates with how arousing the content is. However, scene 2 score the highest in both cases of EXP 1 and arousal. Investigating this phenomenon more in detail:

- Scene 1 is the first encountering with the virtual environment that *Everest VR* provides. Hence, the first impression may play a crucial role in how participants perceive and evaluate the level of immersion, in spite of the content not being perceived that exciting and/or arousing.
- Scene 3 provides highly arousing content (see Fig. 5.2), though it scored lowest on the subjective level of immersion (see Fig. 5.3). One reason for this can be the time length of the scene, since scene 3 lasts significantly longer than the rest with mostly only one task (climbing an ice wall) performed over the entire time span. The lack of varying content over time seems to ruin the immersion factor of the scene. Also, not being able to climb and move in a pace the participants choose themselves can ruin the realness of the situation, hence influencing the perceived level of immersion.
- As for scene 4 and 5, they have the highest score in terms of dominance factor (see Fig. 5.2). They also score well on the level of immersion, but of the lowest in the level of arousal. This shows that the participants feel more immersed when they are more in control of the VR environment and system, even though the task or content may not be very exciting.

The results presented in EXP 6, concerning how much participants needed to learn, (Fig. 5.3) are not very surprising, since the variation in difficulty and amount of tasks is different between the scenes. Scene 2 and scene 4 both include grabbing hold of a rope in a specific matter, hence giving almost equal results.

Results concerning how interesting participants found one particular task (EXP9 in Fig. 5.3) are very interesting when comparing it with the results from arousal ratings and ratings of immersion. The tasks in scene 2 and 4 were found by the participants to be the most interesting, which is also the same scenes scoring well in both exciting/arousing and immersive content. When looking at these scenes compared to the others they have more life-like interactivity not bounded by area limits in the VR environment, and with the freedom of doing things in a pace the participants choose themselves.

6.2 Physiological Measurements

The heart rate data in Fig. 5.4 shows some very important and interesting artifacts. During the exploration of scene 1 participants spent most time in the lower heart rate zones, meaning that the content was not very arousing in terms of increased HR. This is not very surprising, as the participants self-reported feedback shows that this is the least interesting and exciting/arousing scene. That again corresponds well with the results from the arousal ratings and EXP9, but again contradicts the results from EXP1, indicating that not necessarily a correlation is present between the level of arousal and level of immersion.

During scene 2, time spent in each heart rate zone is more equally spread. This shows that the level of HR is generally more increased in this scene, as the higher zones are more equaled out with the rest compared to scene 1.

In the case of scene 3, where the participants had to do the same physical activity over a long time span which resulted in more time spent in zones 2 and 3. The type of movement, hatching with an ice axe over and over, will indeed increase the HR to a higher zone over a longer time, but as noted in the results from EXP1 it does not necessarily indicate a higher level of immersion.

The results from the correlation tests showed no statistical correlation between the subjective and physiological data, hence throwing the test hypothesis: ”*The level of immersion is related to the measured heart rate in a way that the heart rate increases when a participant's level of immersion increases.*”. This does not necessarily mean that it has to be thrown away for good, since there are indicators in the results that correlation exists in some sense. To investigate this further, criticism of the data analysis and experiment design should be done, as well as mapping of further work on the topic.

6.3 Critics

It is important to criticize the method in order to recognize and understand the pit falls, weaknesses and possible sources of error in the experiment.

Immersion is a complex term to work with. This thesis tried to identify a correlation between the level of immersion and HR to be used in QoE assessment for interactive storytelling, and uses the description of immersion as described in chapter 2 for this. However this does not necessarily mean that this definition is the correct one to use for this purpose. Research on immersion in multimedia technology is a large topic of its own, and measuring and quantifying it properly can probably not be done until the term is well defined and standardized.

When it comes to the experiment design and methodology there are several sources of error and weaknesses that needs to be addressed. Especially since there up to this date does not exist any specific standards or recommendations for this type of experiment.

- The age and occupation distribution of the participants should ideally have been more wider and spread. Participants was mainly students in the age range from 20 to 25 years old. Also, the amount of participants should be larger. The gender distribution however was 55/45 %, which is quite good.
- The VR experience distribution should ideally have contained participants with a bigger variety in experience. Most of the participants had none or little experience. It would have been of interest to see if the amount of experience influences the variation in physiological signals.
- The experiment did not include any form of randomness to eliminate any advantages the following scenes has due to the fact that the participants will have some experience with the story from the first scene. However, due to the severe length of *Everest VR* and the fact that each scene needs to be performed in a chronological order, it was hard to include other stories to the experiment and randomizing the scenes.
- Computer related problems will influence the results. This is because sometimes the computers have too much to process, leading to delays in motion tracking and HMD update. Lagging also occurred in some scenes for some participants, making the data corrupt.
- The training session, *Trondheim 360*, should in retrospect been a digital story more like *Everest VR*. The main reason for this is that *Trondheim 360* consisted of real world still images, while *Everest VR* is a pure virtual game engine environment.

Due to the lack of good consumer wear for measuring physiological data, and especially wearable sensors that also gives easy access to the raw data, one should be critical to the results from the sensor data.

- The Fitbit Charge 2 can at best only provide data second by second(though it does not even do this all the time either). Also, the wristband is not a very good fit for small wrists. Long measuring breaks can occur since wristband sits too loose, making the data resolution bad.
- HBPM is not ”raw” data in the sense of HR. HR variability, which HBPM is an estimate from, would probably be a better choice for data analysis. Also, more physiological data like EDA should be included. Hopefully wearable sensors for EDA in interactive purposes will be available in near future.
- Other factors such as participants level of movement and gesture behaviour was not taken into consideration. There is no doubt that variation in HR is related to the pattern movement of a participant.

6.4 Further Work

The results from chapter 5 indicates that the test hypothesis must be thrown. But, as it was mentioned, there was some indicators in the results from 5.1 and 5.2 that there may be some correlation after all. To investigate this further, a good standardized experimental setup should be defined before conducting a new experiment. Other physiological parameters and better developed questionnaires should be included and investigated, as well as more research on the topics of immersive experiences in storytelling.

Based on the work and results from [31], [32] and this thesis, the following suggestions to further work is provided:

- Analyze the data obtained by this experiment more thorough by looking at the actual change of HR, and comparing them to the videos captured.
- Investigate how the combination of variations in physiological signals, eye-tracking and hand gestures/body movement can be related to the level of immersion.
- Evaluate which physiological parameter(HR variability, EDA, etc.) that is most suitable when relating measurements of physiological signals to the level of immersion.
- Investigate how the two paragraphs above can be used for QoE assessment in digital storytelling.
- Develop a questionnaire for evaluating the quality of immersive experience in digital storytelling.

7 Conclusion

The motivation for this report was to contribute to the field of QoE assessment, by providing knowledge of the relationship between contributing factors of immersion in interactive digital storytelling.

The report describes the design and implementation of a system suitable for conducting an experiment investigating the level of immersion in VR storytelling content. It aims to evaluate the correlation between users physiological signals and level of immersion, and if it can be used to quantify the QoE.

A test hypothesis was developed and tested through an experiment by using a self-developed research protocol, HTC Vive VR system, and a Fitbit Charge 2 health tracker. All participants in the experiment, 23 in total, played through an interactive digital story named *Everest VR*. *Everest VR* consisted of 5 scenes with different content. During the scenes, their heart rate was measured. After each scene, the participants had to answer questionnaires regarding their experience of that particular scene.

Self-reported measures of immersion show that the content itself may not necessarily need to be highly arousing in order to feel most immersed. It is, therefore, more important that the user of a VR-system has the impression of being in control of the experience, hence enjoying it more. When it comes to physiological measures of the VR-user, we can see that a more interactive scene leads to a higher level of heart rate. However, to better estimate the emotional state, measures such as heart-rate variability or skin conductance may be more appropriate and need further investigations.

No statistical correlation was found between the physiological and subjective measurements, hence the test hypothesis was discarded. Since the experiment included several sources of error and the fact that this is a very early work on the topic, more research is needed before discarding it for good. The experiment needs better-developed questionnaires, as well as a defined and standardized experimental setup for similar research.

The results from this report contribute to the field of QoE assessment. We can show that different subjective dimensions are contributing to the level of immersion experienced and that heart rate is reflecting the level of interactivity of an interactive digital story.

It also provides a better understanding of further research and work needed, in order to obtain more knowledge on the relationship between subjective and physiological parameters.

Bibliography

[1] G. Robles-De-La-Torre, "International society for haptics: Haptic technology, an animated explanation.," *Isfh. org*, no. 1, 2010.

[2] B. Xie, *Head-Related Transfer Function and Virtual Auditory Display*, ch. 1, pp. 20–33. J. Ross Publishing, 2 ed., 2013.

[3] "*Digital Storytelling*." Accessed: 2017-05-27 [Online], Available: `https://en.wikipedia.org/wiki/Digital_storytelling`.

[4] "*The Power of Storytelling*." Accessed: 2017-05-27 [Online], Available: `http://www.tellingdigitalstories.com/2015/02/01/power-storytelling/`.

[5] B. R. Robin, "Digital storytelling: A powerful technology tool for the 21st century classroom.," *Theory into practice*, vol. 47, no. 3, pp. 220–228, 2008.

[6] M. Roussou, "Immersive interactive virtual reality in the museum," *Proc. of TiLE (Trends in Leisure Entertainment)*, 2001.

[7] S. Lukas, *The immersive worlds handbook: designing theme parks and consumer spaces.*

[8] O. Grau, *Virtual Art: from illusion to immersion.*

[9] M. Slater, "A note on presence terminology.," *Presence connect*, vol. 3, no. 3, pp. 1–5, 2003.

[10] A. McMahan, "Immersion, engagement and presence.," *The video game theory reader*, vol. 67, no. 86, 2003.

[11] P. Callet, S. Moller, and A. Perkis, "Qualinet white paper on definitions of quality of experience output from the fifth qualinet meeting," 2012.

[12] K. Laghari and K. Connelly, "Toward total quality of experience: A qoe model in a communication ecosystem," *IEEE Communications Magazine*, vol. 50, no. 4, 2012.

[13] A. Perkis, "Quality of experience (QoE) in multimedia applications," *SPIE Newsroom*, 2013.

[14] C. Zhang, A. S. Hoel, and A. Perkis, "Quality of immersive experience in storytelling: A framework," in *2016 Eighth International Workshop on Quality of Multimedia Experience (QoMEX), Lisbon*, 2016.

[15] I. Hupont, J. Gracia, L. Sanagustin, and M. A. Gracia, "How do new visual immersive systems influence gaming qoe? a use case of serious gaming with oculus rift.," *Quality of Multimedia Experience (QoMEX), 2015 Seventh International Workshop*, pp. 1–6, 2015.

[16] I. T. Union, "Itu-t recommendation p. 800: Methods for subjective determination of transmission quality..," 1996.

[17] "*Likert Scale*." Accessed: 2017-05-27 [Online], Available: `https://en.wikipedia.org/wiki/Likert_scale`.

[18] C. Timmerer, T. Ebrahimi, and F. Pereira, "Toward a new assessment of quality.," *Computer*, vol. 48, no. 3, pp. 108–110, 2015.

[19] J. Donley, C. Ritz, and M. Shujau, "Analysing the quality of experience of multisensory media from measurements of physiological responses," in *Sixth International Workshop on Quality of Multimedia Experience, QoMEX 2014, Singapore*, 2014.

[20] S. Möller and A. Raake, *Quality of Experience*, ch. 9, pp. 121–123. Springer, 1 ed., 2013.

[21] S. Arndt, "Using physiology in qoe," *Lecture slides in course TTT4135 - Multimedia Signal processing, NTNU*, 2017.

[22] A. A. Boulton, "Neurophysiological techniques: basic methods and concepts.," 1990.

[23] M. M. Bradley and P. J. Lang, *Handbook of psychophysiology 2*, ch. 22, pp. 602–642. Cambridge University Press, 2 ed., 2000. Chapter on "Emotion and Motivation".

[24] M. Barreda-Ángeles, R. Pépion, E. Bosc, P. Le Callet, and A. Pereda-Baños, "Exploring the effects of 3d visual discomfort on viewers' emotions.," *In Image Processing (ICIP), 2014 IEEE International Conference*, pp. 753–757, 2014.

[25] G. M. Wilson and M. A. Sasse, "Do users always know what's good for them? utilising physiological responses to assess media quality.," *In People and Computers XIV—Usability or Else!*, pp. 327–339, 2000.

[26] M. M. Bradley and P. J. Lang, "Measuring emotion: the self-assessment manikin and the semantic differential.," *Journal of behavior therapy and experimental psychiatry*, vol. 25, no. 1, pp. 49–59, 1994.

[27] A. Perrin, H. Xu, , E. Kroupi, M. Rerábek, and T. Ebrahimi, "Multimodal dataset for assessment of quality of experience in immersive multimedia.," *In Proceedings of the 23rd ACM international conference on Multimedia.*, pp. 1007–1010, 2015.

[28] J. Donley, C. Ritz, and M. Shujau, "Analysing the quality of experience of multisensory media from measurements of physiological signals.," *Sixth International Workshop on Quality of Multimedia Experience (QoMEX).*, 2014.

[29] G. Ghinea, C. Timmerer, W. Lin, and S. R. Gulliver, "Mulsemedia: State of the art, perspectives, and challenges.," *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, vol. 11, no. 1, p. 17, 2014.

[30] U. Engelke, D. P. Darcy, G. H. Mulliken, S. Bosse, M. G. Martini, S. Arndt, J. Antons, N. Chan, K.Y. abd Ramzan, and K. Brunnstrom, "Psychophysiology-based qoe assessment: A survey.," *JSTSP special issue on measuring QoE for advanced media technologies and services*, 2016.

[31] D. Egan, S. Brennan, J. Barrett, Y. Qiao, C. Timmerer, and N. Murray, "An evaluation of heart rate and electrodermal activity as an objective qoe evaluation method for immersive virtual reality environments.," *In Quality of Multimedia Experience (QoMEX), 2016 Eighth International Conference.*, pp. 1–6, 2016.

[32] C. Keighrey, F. Ronan, S. Murray, and N. Murray, "A qoe evaluation of immersive augmented and virtual reality speech language assessment applications.," 2017.

[33] "*31 new digital health tools*." Accessed: 2017-05-29 [Online], Available: `http://www.mobihealthnews.com/content/31-new-digital-health-tools-showcased-ces-2017`.

[34] S. Schachter and J. Singer, "Cognitive, social, and physiological determinants of emotional state.," *Psychological review*, vol. 69, no. 5, p. 379, 1962.

[35] "*Derfor er HTC bedre enn Oculus Rift akkurat nå*." Accessed: 2017-05-25 [Online], Available: `http://itavisen.no/2016/08/12/test-derfor-er-htc-vive-bedre-enn-oculus-rift-akkurat-na/`.

[36] "*HTC Vive System Parts*." Accessed: 2017-05-25 [Online], Available: `https://www.vive.com/eu/`.

[37] "*Fitbit Charge 2*." Accessed: 2017-05-25 [Online], Available: `https://www.fitbit.com/no/charge2`.

[38] I. T. U. R. B. T. Recommendation, ""methodology for the subjective assessment of the quality of television pictures."," *Recommendation ITU-R BT. 500-11. ITU Telecom. Standardization Sector of ITU.*, 2002.

[39] ISO, *International Standard ISO 8589: Sensory Analysis: General Guidance for the Design of Test Rooms.*

[40] I. T. U. T. P. Recommendation, *800, Methods for subjective determination of transmission quality.*

[41] D. Chertoff, B. Goldiez, and J. LaViola, "Virtual experience test: A virtual environment evaluation questionnaire.," *In Virtual Reality Conference (VR)*, pp. 103–110, 2010.

Appendix A - Research Protocol Template

Research protocol

Template

Andrew Perkis, Wendy Ann Mansilla

Version: 6/17/2017

A research protocol forms the basis for how a study is run. A well written and complete protocol is essential for a high quality study. Time spent on writing a detailed protocol will avoid problems during the study, and will make publishing the results easier.

A complete protocol is also essential for the study to be approved by the ethics process and for the study to be compliant with regulations (might be necessary for subjective testing and especially for *Picturing the Brain* and other studies involving the medical community).

1. Synopsis - what do I want to study?

This is similar to an abstract and should be about the same length (250-300 words). It acts as a stand alone summary of the study and should be present in large protocols. It generally consists of 1-2 sentences background then the concise objective or aim followed by a brief outline of description of participants, intervention, methods, outcome measures and proposed analysis.

This section should answer the question: what do I want to study or propose?

2. Introduction / Background

This should include the following:

- Introduction to the topic of interest.
- What is known already - literature review of relevant findings (Brief and focused)
- Highlight area where there is missing information in the literature
- Problem statement
    - State the aims of this trial
    - What the study is going to find out – the problem
    - Formulate one sentence how this is going to be achieved.
- Describe what impact the solution to the problem will have
- A handful of relevant references

The background should not be an exhaustive literature review. At the end the reader should have a clear idea of the research question, an understanding that it is original and relevant, and how this research will help fill the gap in the literature.

3. Hypothesis

Provide a clear statement of primary and secondary objectives of the study. This should be summarized in a clearly defined **hypothesis**.

The hypothesis provides an answer to what conclusion we are expecting from this study.

4. Methodology and design

Describe

- Resources required
    - Logistics and personnel
    - Rooms (testing)
    - Programming language and compilers
    - Data set (testing)
    - Funding
- Which measures will be used to test the hypothesis
    - Assessment methodology (objective or subjective)
        - What do we measure
        - How do we measure it

5. Results (Analysis, discussion)

Details on how the primary and secondary outcomes will be analyzed.

- Statistical analysis
    - Statistical methods to be used
- Hypothesis testing
- Publication plan (e.g. state the conferences or journal the study fits to)

6. Priority and Timetable

State the schedule as to when you want to execute the experiment and the period of writing the results according to your scheduled publication plan.

Appendix B - Research Protocol for Physiological QoE Assessment of Digital Storytelling Experiment

Research Protocol

Evaluation of Heart Rate as a Parameter for Immersion in QoE assessment of Interactive Digital Storytelling

Author: Martin Ervik
Guidance:Sebastian Arndt

Version: 15.04.2017

1.Synopsis - What do I want to study? 2

2. Introduction/Background 2

3. Hypothesis 4
3.1 Main Objective 4
3.2 Secondary Objective 4

4. Methodology and Design 4
4.1 Resources 4
4.2 Stimulus - The Digital Stories 5
4.2.1 Training Session - 'Trondheim 360' 5
4.2.2 Main Task - Everest VR 5
4.3 The Experiment 5
4.4 Questionnaires 7

5. Results (Analysis, Discussion) 7

6. Priority and Timetable 8

7. References 8

A - Project Plan (Milestones) 10

B - Questionnaire 10

1.Synopsis - What do I want to study?

This study tries to design and implement a system suitable for conducting an experiment investigating the level of immersion in Virtual Reality storytelling content. In the study, participants will play through a digital story of high quality using the head mounted display of the HTC Vive and its controllers. The story consists of five scenes with different content that assume a different level of immersion and interactivity.

During the experiment, participants will have their heart rate measured by a wearable health tracker. In addition, the participants will have to answer subjective questionnaires regarding their experience of that particular scene. We will investigate what subjective factors are contributing to experienced immersion and whether the level of immersion is related to the measured heart rate.

2. Introduction/Background

Storytelling has played an important part through the ages in preserving the constant urges we have as humans. It does not matter which culture you origin from, storytelling has been, and always will be, the main tool to entertain, educate, and socialize ourselves. As our predecessors gained knowledge, storytelling started appearing in new formats. With the 20th century came modern technology such as TV, computers etc. This has introduced new platforms for storytelling, turning it more and more digital. Analog TV- and radio signals will be turned of for good, being replaced with a new digital network. At the same time, 3D- and Virtual Reality technology (1) are being used in new and creative ways every single day, reforming the use of digital storytelling in sectors like education and entertainment (2).

In traditional storytelling, the perceiver always follows a certain linear and logical path through the story. This is known as linear storytelling, and there are several digital story productions available for HMD systems in VR today made with a linear story line. The opposite would be non-linear story lines, where a user can interact with and change the content, hence manipulating the line of events.

With recent years advance in technology, systems using Head-Mounted-Displays (HMD) in the fields of Virtual Reality (VR) and Augmented Reality (AR) allows aspects of immersion (3)(4)(5) and engagement to be considered when producing content for new and interactive, digital narratives. Here, immersion may be defined as how our senses and emotions are deeply involved in an experience (6) (7), making us feel as a part of what is happening in the story. Even though systems using HMD are becoming cheaper and more accessible, we still have not seen the ‘killer’ application or story creating the definite ‘must-have’ excitement from consumers. To create new and exciting content, producers need ways to gain insight in what consumers experience as good quality content. To do so, we need to assess the users Quality of Experience (QoE) among others (8).

With the level of immersion being an important aspect in digital storytelling (9), it takes role as a predominant candidate when it comes to quantifying the QoE. Traditionally, the QoE has mainly been assessed by subjective questionnaires (10), but new concepts also use physiological signal responses (11) (e.g. heart-rate-variability) from the user as a parameter in quality assessment. Since QoE can be closely related to the sensory experience term, QoE assessments can use variation in physiological signals to evaluate the emotional response from a user in terms of the level of immersion (12) (13) (14). Immersion has long been used as a describing term in culture areas such as art (15), museums (3) and theme parks (4), but it still does not have a well agreed upon definition to be used for technology in general. Immersion is also often confused with the notion of presence. So to define the concept of immersion as it can be described and used in technology, we first need to distinguish it from the term presence. Both McMahan (16) and Slater (6) among others have proposed definitions of immersion distinguishing it from presence, both concluding that presence is a human reaction to immersion. For the purpose of this paper, we take Slaters definition as a good starting point: "Immersion is what the technology delivers from an objective point of view. The more that a system delivers displays(in all sensory modalities) and tracking that preserves fidelity in relation to their equivalent real-world sensory modalities, the more that is immersive."

With the above definition in mind, one might say that presence is more about form, and distinguished from emotional responses evoked by immersive experiences, which has to do with the content itself. For instance, an expression like "The VR environment was very real and made me feel like I was actually there." is a sign of presence. Whereas an expression like "The VR environment was very real, but it was really boring and uninteresting in a way I lost interest after a while." is more an emotional response to how immersive the content was. The latter creates a link between emotions and immersion, which is of great interest when wanting to capture and measure the experience of a digital story, as emotions are closely related to physiological responses.

One of the human factors highlighted in the framework of quality of immersive experiences in storytelling is physiological parameters (9). As (16) have established a 3 link between physiological arousal and emotional states, we can then define and describe content in terms of how immersive and arousing it can be for users. By combining variations in physiological measurements (e.g. changes in heart rate or skin conductance) with questionnaires capturing system and design factors, one might be able to quantify and measure the QoE of different content in a digital story. If these combined measures correlates in terms of how immersive and arousal different content is described, then there exists a good chance that QoE of a digital story can be assessed through an experiment covering these factors.

Physiological QoE assessment is already widely accepted. However, previous research on this topic has focused on how physiological measurements can be used to assess QoE using standard displays, and how the addition of several sensory effects increases this experience (fan blowing wind, a dog sneezes you in the face etc) (17) (19) (20). However, the research has a gap when it comes to using physiological measurements for assessing QoE in new display technology, such as HMD in immersive VR environments (11). There has been recent work (21) (22) in this research field. An experiment on evaluation of heart rate and electrodermal activity as immerse parameters for QoE in VR environments when using HMD has been conducted, though only with a passive system with lack of interactivity. A main obstacle has been the lack of good affordable and wearable sensors, robust to activity and movement. In the mentioned experiment, the participants had to sit still and not move their hands for the sensors to work properly. This will of course limit the level of immersion and interactivity to some point. Adding the possibility to move around can increase the level of immersion, hence giving a better data basis.

To the best of the authors' knowledge, this is the first work that uses physiological parameters to compare user QoE in different content of interactive storytelling, particularly with the focus on using HMD in VR.

3. Hypothesis

3.1 Main Objective

The main hypothesis of this experiment is:

Physiological signals can be used as a parameter for immersiveness when evaluating the Quality of Experience in digital storytelling by using head-mounted displays in virtual reality.

3.2 Secondary Objective

To test if the main hypothesis holds, we define a secondary objective/hypothesis to be tested during the experiment:

The level of immersion is related to the measured heart rate in a way that the heart rate increases when a participant's level of immersion increases.

4. Methodology and Design

4.1 Resources

All the listed equipment and software to be used in this experiment is available at the Café Media lab(room B-133), NTNU Gløshaugen.

- 1 x HTC Vive VR System
    - HTC Vive headset
    - 2 x HTC Vive controllers
    - 2 x HTC Vive infrared camera sensors (base stations)
    - Linkbox and wires for connectivity with PC

- 1 x computer with following software:
    - Steam
    - Steam VR
    - Everest VR
    - Xbox screen recorder

- 2 x Fitbit Charge 2 (For measurements of heart-rate)
- 30 participants
- 30 cinema tickets

4.2 Stimulus - The Digital Stories

Two VR productions are to be used in this experiment. The participants are set to run through *Everest VR*, an experience/digital story told through several immersive scenes. Before doing this, a training session with a production called Trondheim 360 must be performed. Both productions are played through SteamVR, with the use of the HTC Vive VR system. The different content will be described in terms of interactivity and immersion in the following.

4.2.1 Training Session - 'Trondheim 360'

Participants will need to perform a training session with a follow-up questionnaire. The procedure will be similar to the one used in the main part of the experiment. This will help participants get comfortable with experiment setup and equipment, as well as removing influences from (possible) first time impressions of using head-mounted displays and virtual environments. A production called *Trondheim 360*, internally produced serves as content for the training session. In *Trondheim 360,* the user stands on a platform overlooking and navigating through a set of 360 degree still images of the city, landscape, and other cultural sites from the region around the city. The score of this experience were is not considered for further analysis

4.2.2 Main Task - Everest VR

As a main task, the participants will run through the 'Everest VR' experience. 'Everest' is a highly interactive digital story, and includes different type of scenes with use of both interaction and storytelling. The whole experience lasts for approx. 35-45 minutes, depending on how well each participant adapt to the usability of the controllers.

The different scenes in the experience includes crossing of a glacier, climbing a ladder, and balancing on a mountain ledge with the use of the HTC Vive VR system controllers. Between each scene there will be a short break where the participant will have to answer a short questionnaire regarding their experience of that actual scene.

4.3 The Experiment

(currently not finished - version 01.03.2017)

Currently, there exists no recommendations for evaluating the use of HMD in VR, or interactive multimedia content in general, suitable for this experiment. However, a combination of recommendations for sensory analysis and evaluation of multimedia content, as well as papers proposing evaluation methods in the same research field could be used. From this, a good fundamental base can be created for the methodology used in this experiment.

As a starting point, it was chosen to use the ITU BT.500 (24) and ISO 8589 (25) recommendations, as well as a paper on the ”*Evaluation of heart rate and EDA as an objective QoE evaluation method for immersive VR environments*” (20). ITU BT.500 is describing a methodology for the subjective assessment of the quality of television pictures, while ISO 8589 describes the lab setup when conducting an experiment including sensory analysis. The paper mentioned is the closest we come an experiment evaluating immersion in terms of physiological signals as a parameter for QoE of interactive content.

All the different parts of the recommendations will not be given here, but important deviations and features added specially for this experiment is listed below:

- Test of color vision and visual acuity is not needed as the experiment is not evaluating the images directly.

- How the test procedure is arranged. Single -or double stimulus does not make sense in this kind of experiment.

- Randomizing the test order of the content does not make sense, as the different content to be compared and evaluated consists of different scenes in a chronological told story. Including more stories would also make the experiment too time demanding.

- Because of the large time consume each test of the experiment takes, and to have a buffer for the data minimum of 20 participants, 25 was decided as a proper number of test participants.

The participants will before they start the experiment read an introduction form about the experiment and the tasks they will run through, as well as read and sign a consent form. They will at the same time have to answer a questionnaire regarding their demography/background. They will then put on a wristband, a Fitbit, for measuring their heart-rate. Total amount of time on this part will be approx. 10 min.

The first main task is a training session consisting of a single scene, *Trondheim 360*. After they finish this, they will have to answer a new questionnaire regarding their perceived Quality of Experience. Total amount of time spent on this part will be approx. 10 min.

The second main task is to run through the *Everest VR* experience. This consists of 5 different scenes, and after each scene is done there will be a short break where the participant will have to answer a questionnaire regarding the perceived Quality of Experience (same as for the training task). Total amount of time spent on this part will be approx. 60 min.

Since *Everest VR* is a digital story told in chronological order, we can not randomize the different scenes. This could actually do an impact on the result, since it is reasonable to think that a highly interactive scene(link to theory - immersiveness) before a less interactive and spectacular scene or opposite will have an impact on how people rate their experiences of the scenes. We can’t solve this by running the task again either, since the participants will then have some experience with the task and it is likely to think that they will perform better. Cause’ of this we will run the experiment in the same order for every participant.

When the experiment is done the participant will receive a cinema ticket as payment.

4.4 Questionnaires

Questionnaires will be used to collect subjective data of participants' perceived experience of the *Everest VR* digital story. The data will then be used to evaluate the QoE and amount of immersion in a traditional, subjective matter. The questionnaires are divided into three parts:

- *Demography background questionnaire:* Is employed before the test starts, and maps the participant's background (demography, former experience, gamer/nongamer etc).

- *Experience questionnaire*: Is performed after each experience/scene in the digital story, thus, five times for each test participant. The purpose is to cover the subjective experience. It consists of three parts:

  - **Quality**: Questions covering the perceived audio, video and overall audiovisual quality. It uses a five-point absolute category scale with the labels: excellent, 6 good, fair, poor, and bad according to ITU when evaluating audiovisual quality in general (10).

  - **SAM:** The self-assessment mannequin (SAM) questionnaire covering a participant's level of valence, arousal and dominance (23). This is used to evaluate the participant's emotional state.

  - **Immersion/Usability:** The aim of these questions is to evaluate two of the key aspects in user QoE for an immersive environment. These questions have been compiled by evaluating different prior work on immersion, including the paper by (22)(23).

- Final overall questionnaire(post-test): Is performed after the experiment is ended. It covers the usability/immersion for the whole experience of Everest VR, as well as comments and feedback from participants.

The amount of time used on each questionnaire will be approx. 5 minutes, which means that the questionnaires will take on 5 min + 30 min(5 min * 6 scenes) = 35min in total.

The questionnaires is listed in *Appendix B*.

5. Results (Analysis, Discussion)

In addition to the self-reported results from the questionnaires, physiological measurements are to be performed by tracking participant's heart rate with the use of a Fitbit Charge 2 health tracker. Heart rate data from the Fitbit can be accessed by connecting to the web-API.

The data is tracked by setting the Fitbit to initiate "Run Mode" when a participant enters a new scene. This allows the Fitbit to track data second by second. When the scene is over, "Run Mode" should be stopped. Time series of heart beats per minute (BPM) from each scene can then be seen in the raw data as time segments where the data is logged second by second.

For evaluation, the heart rate data from each participant is divided temporally into 5 segments, one for each scene. Each scene segment then is again divided into 4 heart rate zones which are derived on a participants' basis. Here, first the participant's maximum and minimum detected heart rate was calculated, and then the range in between was divided into 4 equally large zones. Using these zones, the average time spent by each participant in the different zones can be calculated.

When analyzing the results, we compare the subjective results(as described in the latter subsection) from the questionnaires with the physiological measurements of the participants. Statistical analysis should be performed on the data. This include ANOVA, correlation tests, and statistical analysis with 95% confidence intervall.

6. Priority and Timetable

Due a late Easter holiday this year, the experiment will be done in the time period between the 20th to the 28th of April. The analysis will be done in the following two-three weeks, which means that the analysis should be finished mid-May. See *Appendix A* in the following pages for timetable/project plan (milestones).

7. References

1. International Society for Haptics: Haptic technology, an animated explanation. Robles-De-La-Torre, Gabriel. 2010, Isfh. org, p. 2.
2. Digital storytelling: A powerful technology tool for the 21st century classroom. Robin, Bernard R. 2008, Theory into practice, pp. 220-228.
3. Immersive interactive virtual reality in the museum. Roussou, Maria. 2001, Proc. of TiLE (Trends in Leisure Entertainment).
4. Lukas, Scott. The immersive worlds handbook: designing theme parks and consumer spaces. s.l. : Taylor \& Francis, 2012.
5. Grau, Oliver. Virtual Art: from illusion to immersion. s.l. : MIT Press, 2003.
6. A note on presence terminology. Slater, Mel. 2003, Presence connect.
7. Immersion, engagement and presence. McMahan, Alison. The video game theory reader.
8. Le Callet, et al. Qualinet white paper on definitions of quality of experience.
9. Quality of Immersive Experience in Storytelling: A Framework. Zhang, Chenyan and Hoel, Aud Sissel and Perkis, Andrew. 2016. 2016 Eighth International Workshop on Quality of Multimedia Experience (QoMEX), Lisbon.
10. Unit, International Telecommunications. P.800 : Methods for subjective determination of transmission quality. 1996.
11. Psychophysiology-based QoE assessment: a survey. Engelke, Ulrich and Darcy, Daniel P and Mulliken, Grant H and Bosse, Sebastian and Martini, Maria G and Arndt, Sebastian and Antons, Jan-Niklas and Chan, Kit Yan and Ramzan, Naeem and Brunnström, Kjell.

2017, IEEE Journal of Selected Topics in Signal Processing. 12
12. Exploring the effects of 3D visual discomfort on viewers' emotions. BarredaAngeles, Miguel and Pepion, Romuald and Bosc, Emilie and Le Callet, Patrick and Pereda-Banos, Alexandre. 2014. Image Processing (ICIP), 2014 IEEE International Conference on.
13. Deap: A database for emotion analysis; using physiological signals. Koelstra, Sander and Muhl, Christian and Soleymani, Mohammad and Lee, Jong-Seok and Yazdani, Ashkan and Ebrahimi, Touradj and Pun, Thierry and Nijholt, Anton and Patras, Ioannis. 2012, IEEE Transactions on Affective Computing.
14. The Effects of Text-to-Speech System Quality on Emotional States, Frontal Alpha Band Power. Arndt, Sebastian and Antons, Jan-Niklas and Gupta, Rishabh and Schleicher, Robert and Möller, Sebastian and Falk, Tiago H. 2013. Neural Engineering (NER), 2013 6th International IEEE/EMBS Conference on.
15. Taylor, Laurie and Wolf, Mark and Perron, Bernard. The video game theory reader. s.l. : JSTOR, 2003.
16. Cognitive, social, and physiological determinants of emotional state. Schachter, Stanley and Singer, Jerome. s.l. : American Psychological Association, 1962.
17. Multimodal dataset for assessment of quality of experience in immersive multimedia. Perrin, Anne-Flore Nicole Marie and Xu, He and Kroupi, Eleni and Revrabek, Martin and Ebrahimi, Tourajd. s.l. : Proceedings of the 23rd ACM international conference on Multimedia, 2015.
18. Analysing the Quality of Experience of multisensory media from measurements of physiological responses. Donley, Jacob and Ritz, Christian and Shujau, Muawiyath. s.l. : Quality of Multimedia Experience (QoMEX), 2014 Sixth International Workshop on, 2014.
19. Mulsemedia: State of the art, perspectives, and challenges. Ghinea, Gheorghita and Timmerer, Christian and Lin, Weisi and Gulliver, Stephen R. s.l. : ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM), 2014.
20. An evaluation of Heart Rate and ElectroDermal Activity as an objective QoE evaluation method for immersive virtual reality environments. Egan, Darragh and Brennan, Sean and Barrett, John and Qiao, Yuansong and Timmerer, Christian and Murray, Niall. s.l. : Quality of Multimedia Experience (QoMEX), 2016 Eighth International Conference on, 2016.
21. A QoE Evaluation of Immersive Augmented and Virtual Reality Speech & Language Assessment Applications. Conor Keighrey, Ronan Flynn, Siobhan Murray, Niall Murray. s.l. : QoMEX 2017, 2017.
22. International affective picture system (IAPS): Technical manual and affective ratings. Lang, Peter J and Bradley, Margaret M and Cuthbert, Bruce N. s.l. : NIMH Center for the Study of Emotion and Attention, 1997.
23. Virtual experience test: A virtual environment evaluation questionnaire. Chertoff, Dustin B and Goldiez, Brian and LaViola, Joseph J. s.l. : Virtual Reality Conference (VR), 2010 IEEE, 2010.
24. Recommendation, I. T. U. R. B. T. “methodology for the subjective assessment of the quality of television pictures.” , Recommendation ITU-R BT. 500-11. ITU Telecom. Standardization Sector of ITU. , 2002
25. ISO, International Standard ISO 8589: Sensory Analysis: General Guidance for the Design of Test Rooms.

A - Project Plan (Milestones)

Description	Deadline	Check(Y/N)
Sensors delivered \ Questionnaires for experiment ready	**17.02.2017**	
Subjective test protocol finished - All forms/applications sent to NSD/NTNU ethics council	**24.02.2017**	
Pilot test of experiment done - Writing of report started	**31.03.2017**	
Experiment done - All 30 participants finished	**28.04.2017**	
Analysis of experiment finished	**19.05.2017**	
Final report finished	**23.06.2017**	

B - Questionnaire

This is included as an individual appendix in the master thesis report.

Appendix C - The Questionnaires

Demography/Background

The following questions was used to map the demography among the participants of the experiment, as well as their background and earlier experience with VR equipment and systems.

Demography

- Age
- Gender
- Nationality
- Occupation (If student, please specify which study)

Background/Experience

Q1 - What is your gaming experience on different platforms/consoles?'

	I use it on a regular basis	I use it on an irregular basis (Once in a while)	I have tried it a couple of times	None
Xbox 360/One	◯	◯	◯	◯
Playstation 4	◯	◯	◯	◯
Nintendo Wii	◯	◯	◯	◯
Smartphone games/apps (angrybird, farmville etc)	◯	◯	◯	◯
PC/Computer games	◯	◯	◯	◯
Hand held consoles (Gameboy, PS portable, Nintendo DS etc)	◯	◯	◯	◯

Q2 - What experience do you have with Head-Mounted-Displays?

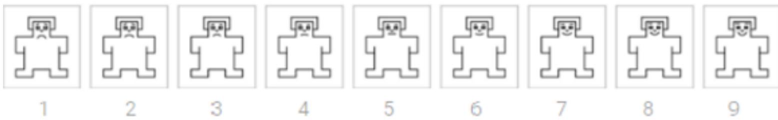
	I use it on a regular basis	I use it on an irregular basis (Once in a while)	I have tried it a couple of times	None
HTC Vive VR System	◯	◯	◯	◯
Oculus Rift	◯	◯	◯	◯
Sony Playstation VR	◯	◯	◯	◯
Samsung Gear	◯	◯	◯	◯
Google Daydream	◯	◯	◯	◯
Google Cardboard	◯	◯	◯	◯
Other	◯	◯	◯	◯

Experience

The questions asked in this question is concerning the Quality of Experience. It's separated in three parts: SAM-scale, Audio/Video, Immersion/Usability

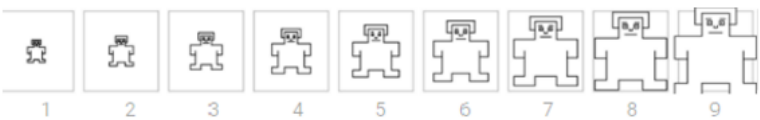
SAM/Emotions

Q1 - How happy or unhappy were you feeling during the scene? Please choose the image on the scale below that best corresponds with your feeling (the scale ranges from 1. very unhappy or sad to the left, to 9. extremely happy or joyful to the right)



Q2 - How calm or excited were you feeling during the scene? Please choose the image on the scale below that best corresponds with your feeling (the scale ranges from 1. very calm or sluggish to the left, to 9. very excited or aroused to the right)

Q3 - How controlled or in control were you feeling during the scene? Please choose the image on the scale below that best corresponds with your feeling (the scale ranges from 1. Very powerless , without control to the left, to 9. very dominant, fully in control to the right)



Audio/Video Quality

The questions in part 2 had the following rating alternatives:

- 5 – Excellent
- 4 – Good
- 3 – Fair
- 2 – Poor
- 1 - Bad

Q1 - How would you rate the overall audiovisual quality?

Q2 - How would you rate the video quality?

Q3 - How would you rate the audio quality?

Quality of Experience – Immersion and Usability

The assertions in part 3 had the following rating alternatives(Likert scale):

- Strongly Disagree

- Disagree
- Neither Agree nor Disagree
- Agree
- Strongly Agree

Note that the alternatives above are weighted from 1 – 5 where 1 corresponds to “Strongly Disagree”, and 5 corresponds to “Strongly Agree”. The exeptions are assurtion 2 and 6, where the weighting is reveresed. This means that for those two assurtions, “Strongly Disagree” corresponds to 5 and vice versa.

Q1 – I was immersed in the environment

Q2 – I did not feel a strong sense of presence whilst experiencing the system

Q3 – The virtual environment was realistic

Q4 – I enjoyed experiencing the virtual environment

Q5 – The system was easy to use

Q6 – I needed to learn a lot of things before I could get going with this system

Q7 – I did not feel any discomfort while using the application

Q8 – I found the user interface to be helpful in informing me of my current task.

Q9 – I thought that the tasks I was able to do in the virtual environment were interesting in this particular scene

Q10 – I found that the content in the story was helpful in informing me of my current task

Final - Overall Experience

The assertions in the final questionnaire(answered after all of Everest VR was finished by the participant) had the following rating alternatives(Likert scale):

- Strongly Disagree
- Disagree
- Neither Agree nor Disagree
- Agree
- Strongly Agree

Note that the alternatives above are weighted from 1 – 5 where 1 corresponds to “Strongly Disagree”, and 5 corresponds to “Strongly Agree”. The exeption are assurtion 2, where the weighting is reveresed. This means that “Strongly Disagree” corresponds to 5 and vice versa.

Q1 – I enjoyed the overall experience of Everest VR

Q2 – My experience did not meet my expectations

Q3 – I thought that the tasks I was able to do in the virtual environment were interesting

Q4 – I would have liked more time in the virtual environment

Appendix D - Fitbit Web-API Program

The following link contains the code used to connect with the Fitbit Web-API: `https://github.com/larseen/fitbit`.

***index.js*:** Contains the main JavaScript code used for the authorization and connection process with the web-API. Rest of the files are for package handling and execution of the code.

Appendix E - Python Program for Data Conversion

The following link contains the Python code used to convert the physiological results(.JSON-files) and the subjective questionnaire results(.CSV-files) to Matlab object files(.mat-files):
`https://www.dropbox.com/sh/336upihl0mcp22k/AABetERhfDPweyN7xemBmDwLa?dl=0`.

***fitbit_data_to_python_arrays.py*:** Takes the JSON-file containing the physiological data and converts it into a Python-array, which is then formated as a Matlab handable object.

***Quest_data_to_python_arrays.py*:** Does the same for the .CSV-files containing the results from the questionnaires as the former did for the physiological results.

Appendix F - Matlab Program for Data Analysis

The following link contains the Python code used to convert the physiological results(.JSON-files) and the subjective questionnaire results(.CSV-files) to Matlab object files(.mat-files): https://www.dropbox.com/sh/knfz9wi5bsxjaya/AABKCiSOAzv6X-JBXeAn4wtsa?dl=0.

***phys_data_to_*:** This folder contains the scripts used to analyze the physiological results from the experiment. The scripts provide plots of time vs hbpm results for each participant in each scene, as well as % time spent in each HR-zone.

***Quest_data_to_python_arrays.py*:** This folder contains the scripts used to analyze the subjective results from the questionnaires in the experiment. The scripts provide barplots of MOS-ratings, as well as a script providing manual feed of data that was corrupted in the conversion process.

***phys_and_subj_data_for_analysis.mat*:** This file contains a Matlab workspace with all result files on the .Mat-format converted from JSON/CSV.

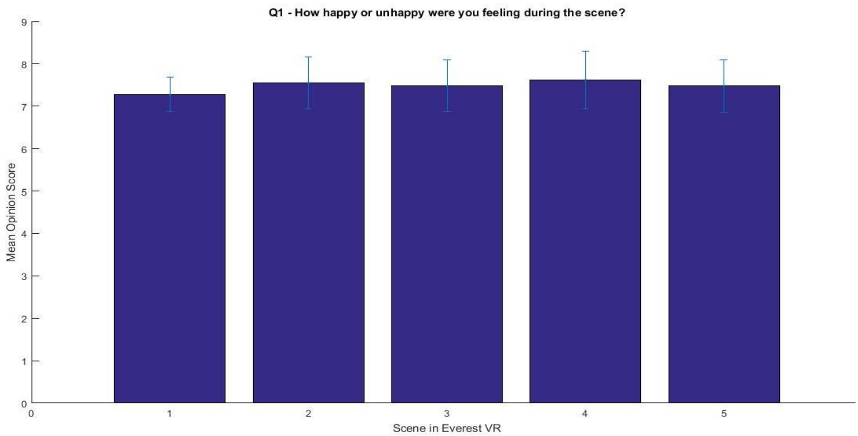
Appendix G - Questionnaire Results: MOS-ratings

Experience

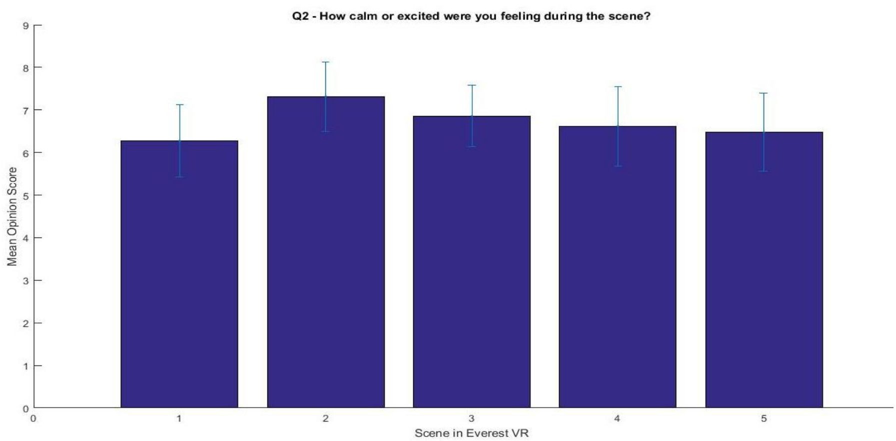
SAM/Emotions

In the following bar graphs below, the Mean Opinion Score for each of the 3 SAM questions are presented in each, separate graph. The five bars shows the average answers for each scene(scene 1 to 5).

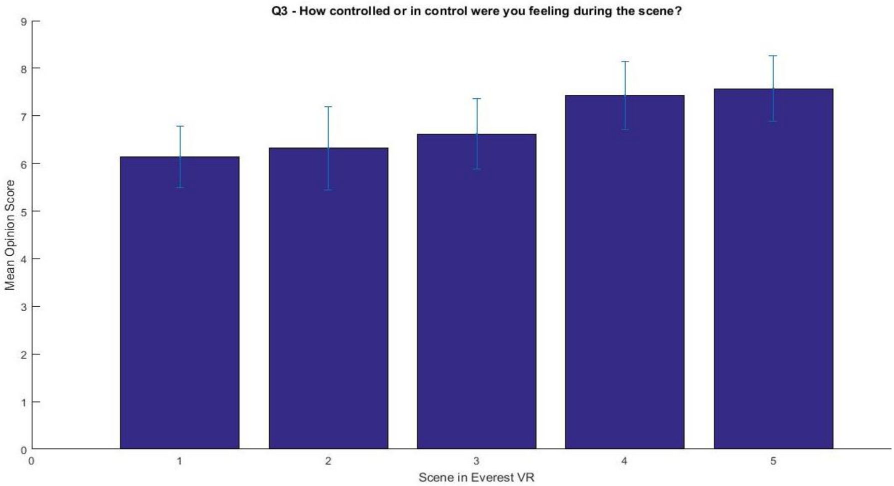
SAM - Question 1



SAM - Question 2



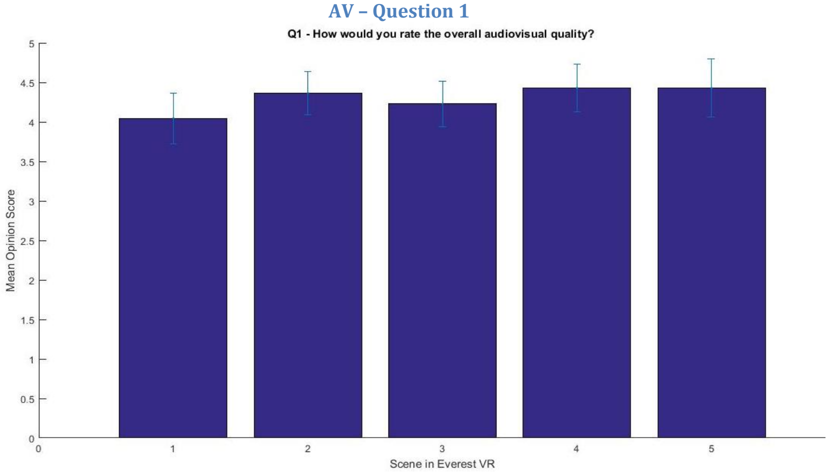
SAM - Question 3



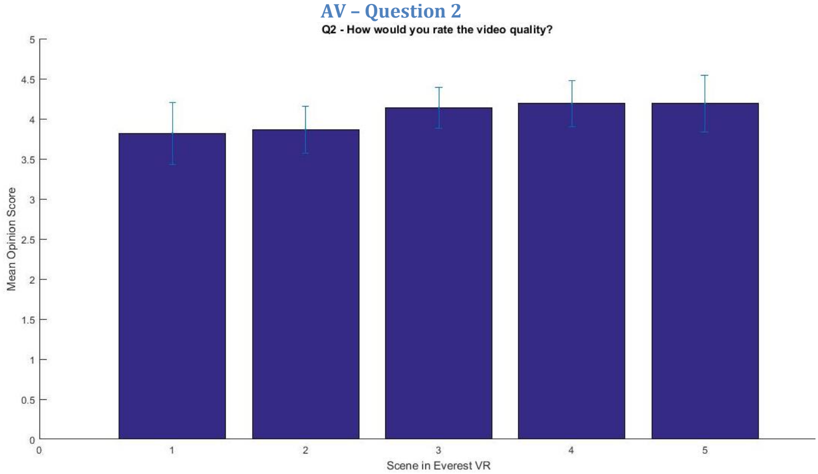
Audio/Video Quality

In the following bar graphs below, the Mean Opinion Score for each of the 3 Audio/Video related questions are presented in each, separate graph. The five bars shows the average answers for each scene(scene 1 to 5).

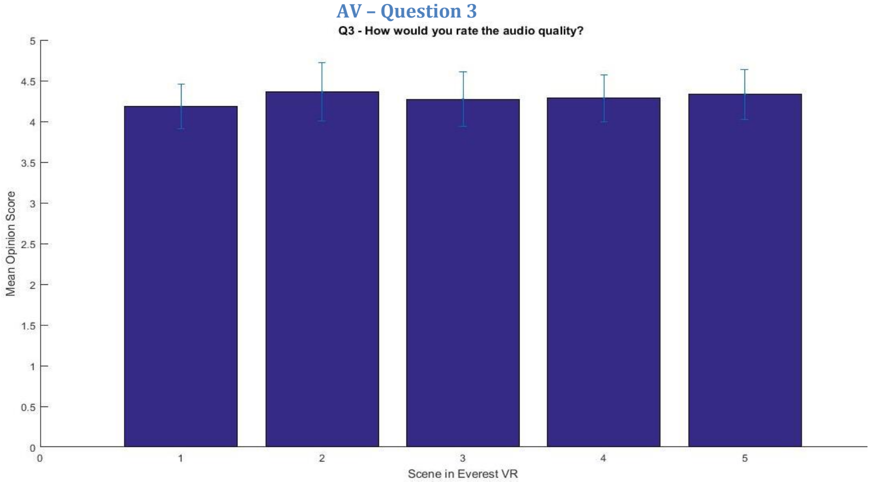
AV – Question 1



AV – Question 2



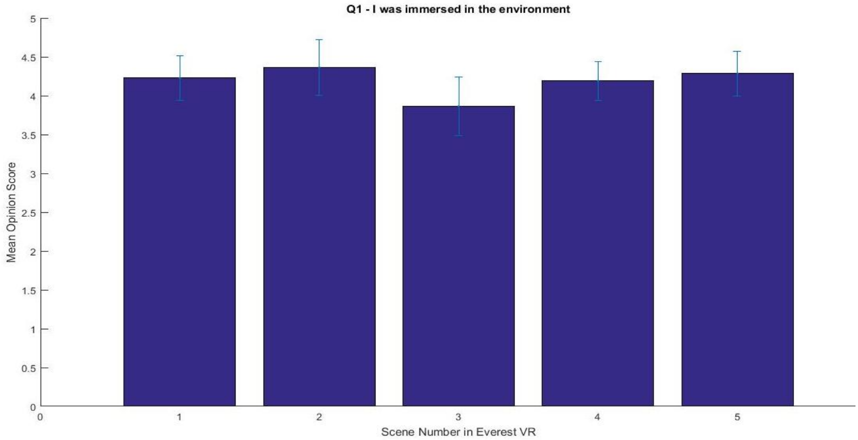
AV – Question 3



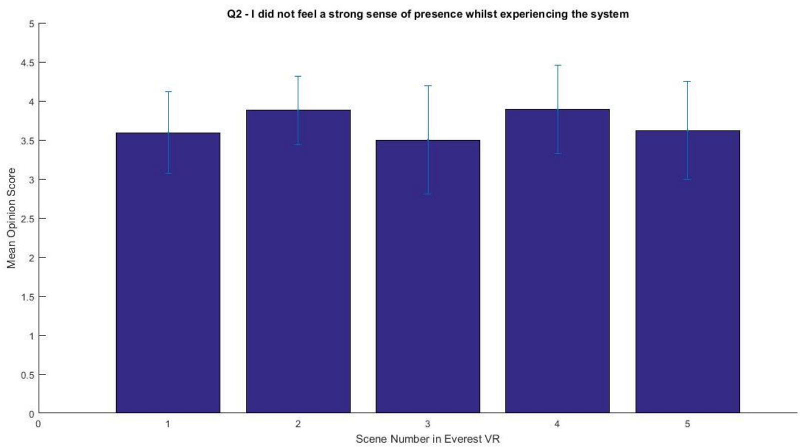
Quality of Experience – Immersion and Usability

In the following bar graphs below, the Mean Opinion Score for each of the 10 assurtions regarding QoE – Immersion and Usability are presented in each, separate graph. The five bars shows the average answers for each scene(scene 1 to 5).

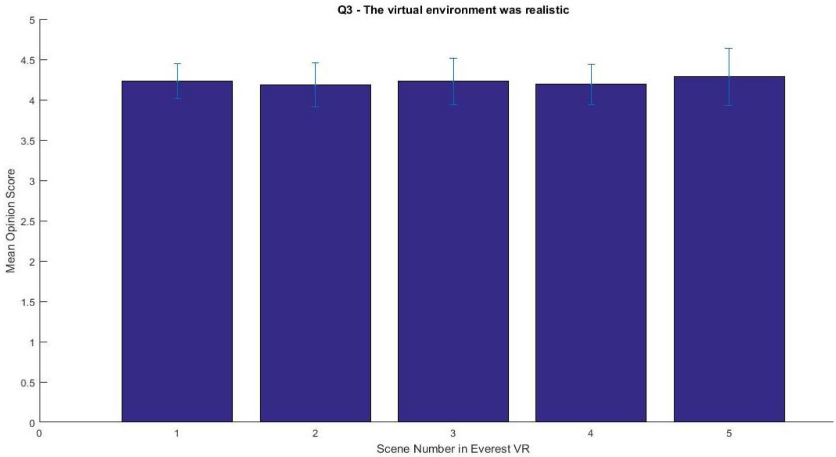
Immersion and Usability – Question 1



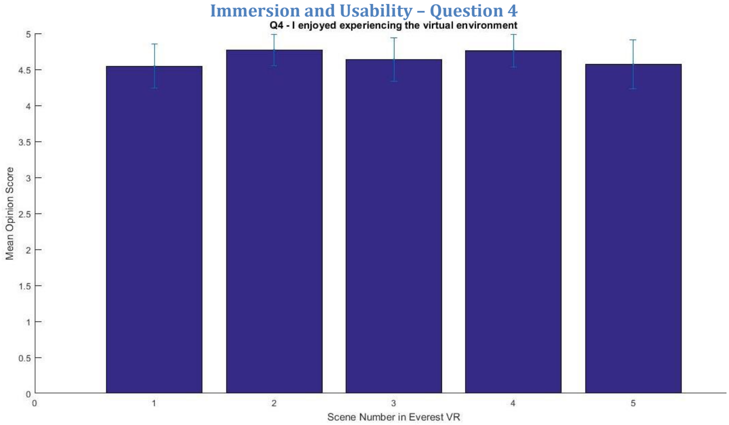
Immersion and Usability – Question 2



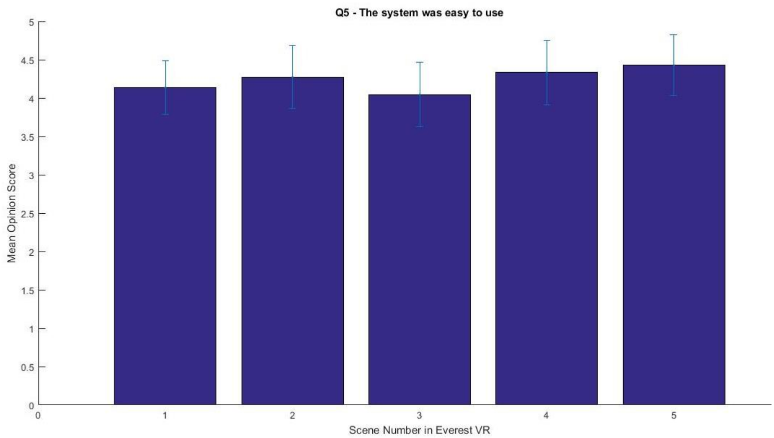
Immersion and Usability – Question 3



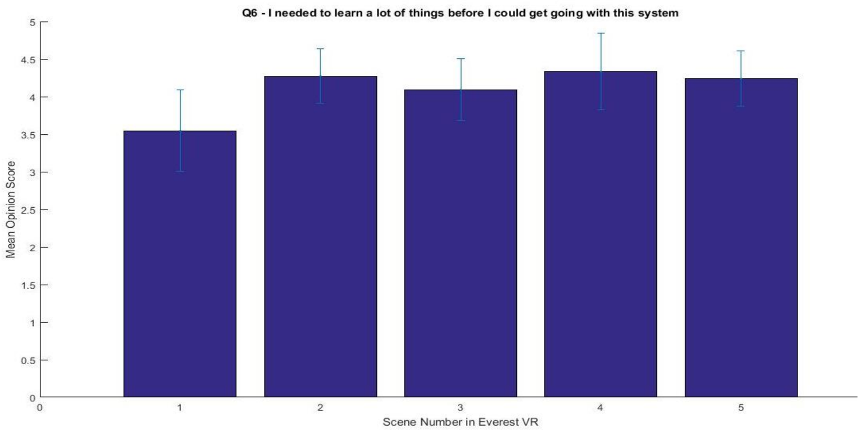
Immersion and Usability – Question 4



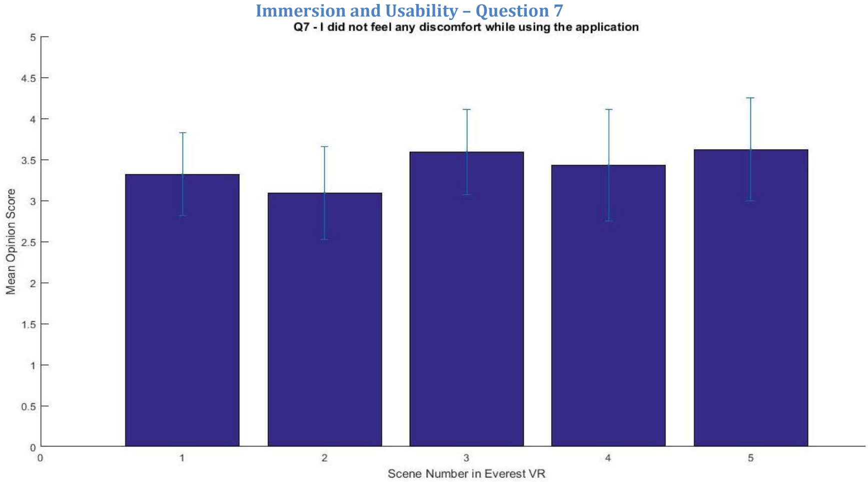
Immersion and Usability – Question 5



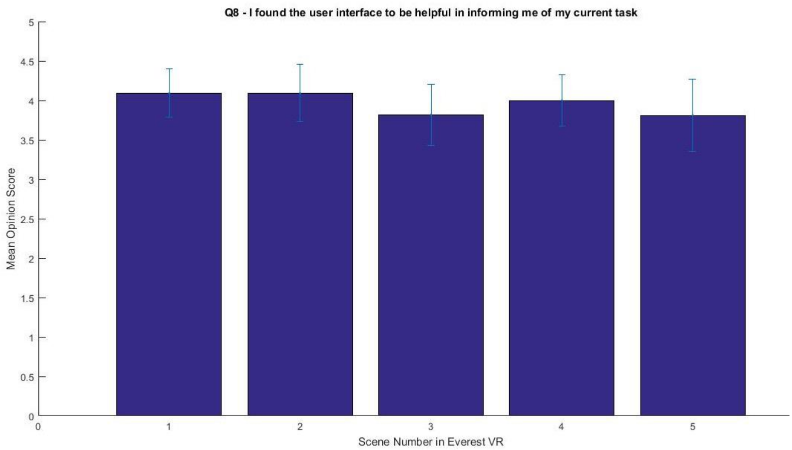
Immersion and Usability – Question 6



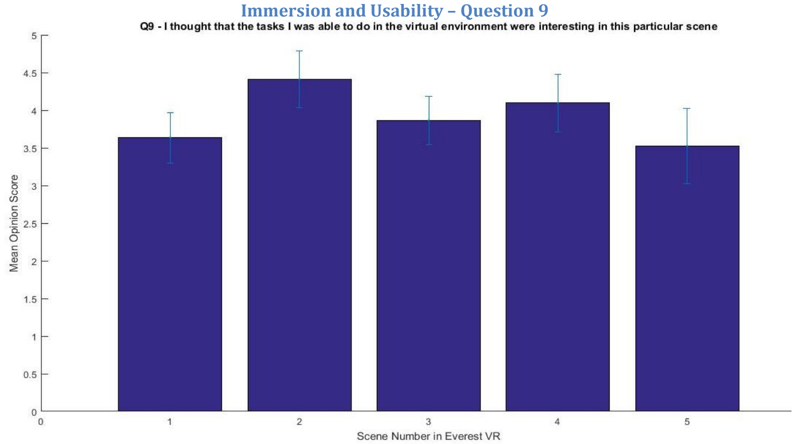
Immersion and Usability – Question 7



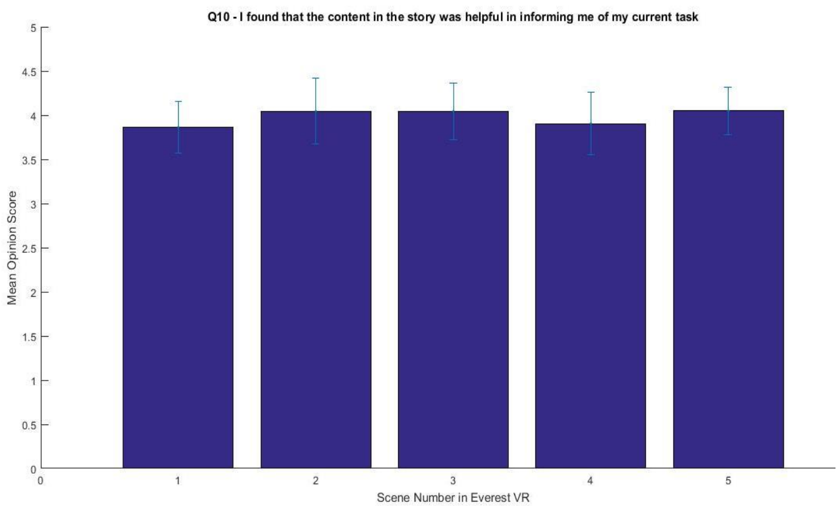
Immersion and Usability – Question 8



Immersion and Usability – Question 9



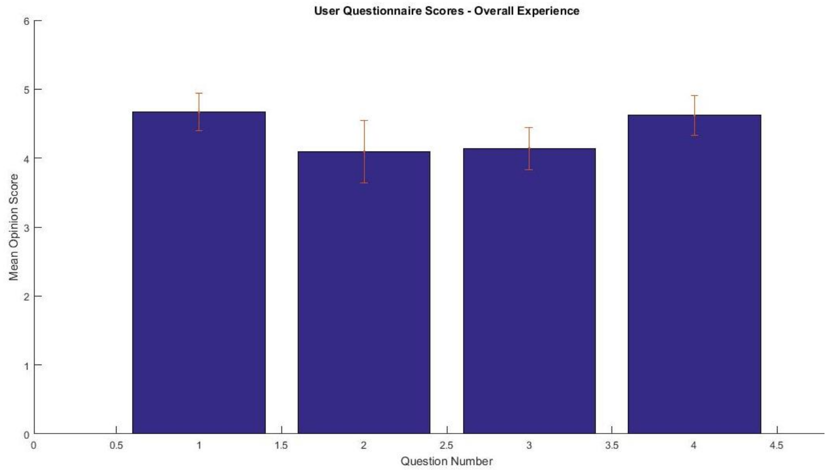
Immersion and Usability – Question 10



Final - Overall Experience

In the graph below, the Mean Opinion Score for each of the 4 – four final questions regarding the overall perceived experience is presented.

Overall Experience



Appendix H - Questionnaire Results: ANOVA tests

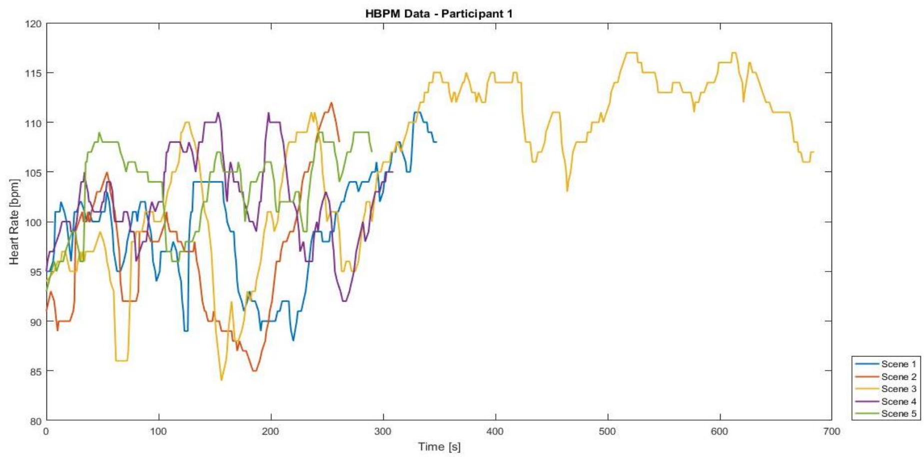
The following link contains the results and files produced by the IBM SPSS software after performing an ANOVA test: `https://www.dropbox.com/sh/732ry198fanj22f/AAB2em38psHUgyXEW8WGWyFHa?dl=0`

Appendix I - Physiological Results: Time Plots

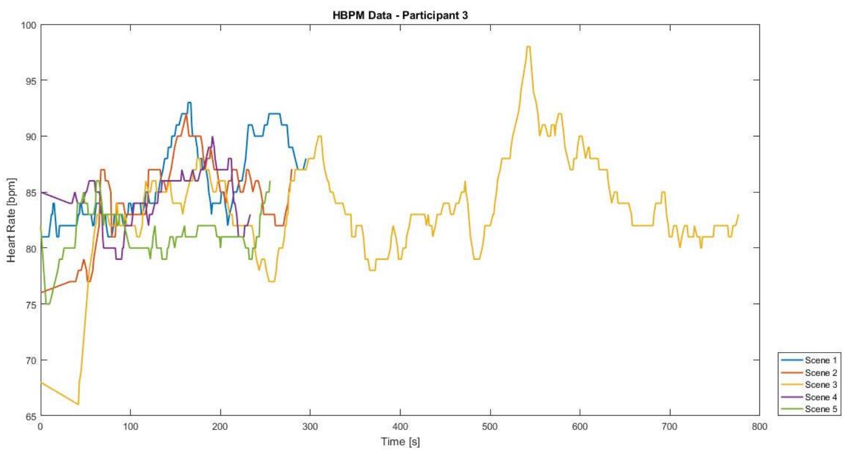
Time Series of Raw Data – Scene by Scene

This chaper consists of the result from the physiolocial measurments done by each participant(except participant 2, that was thrown). Each plot shows the heart rate(hbpm) of the participant over a time intervall, plotted in the same graph for each scene(scene 1 to 5).

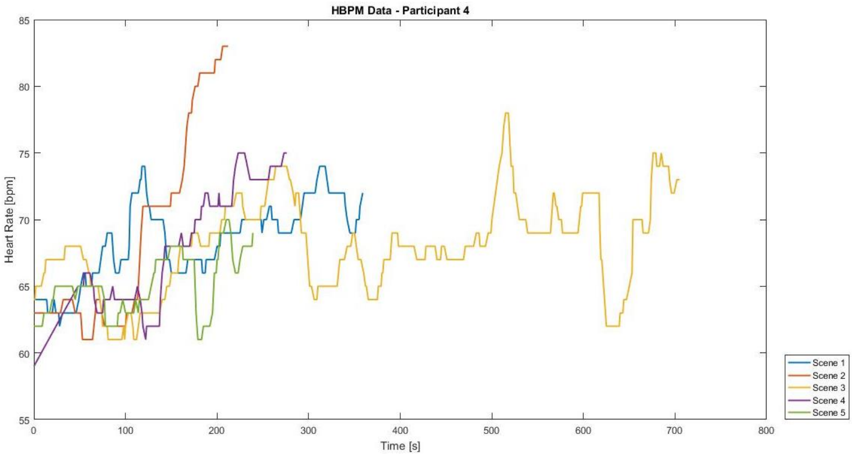
Participant 1



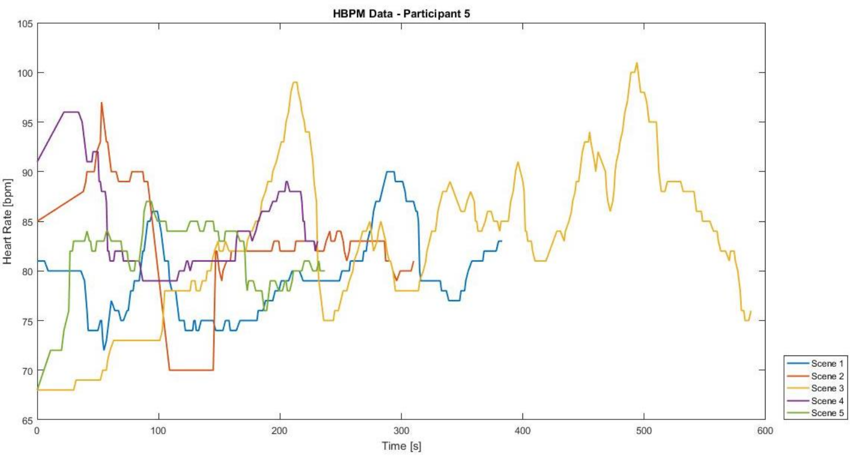
Participant 3



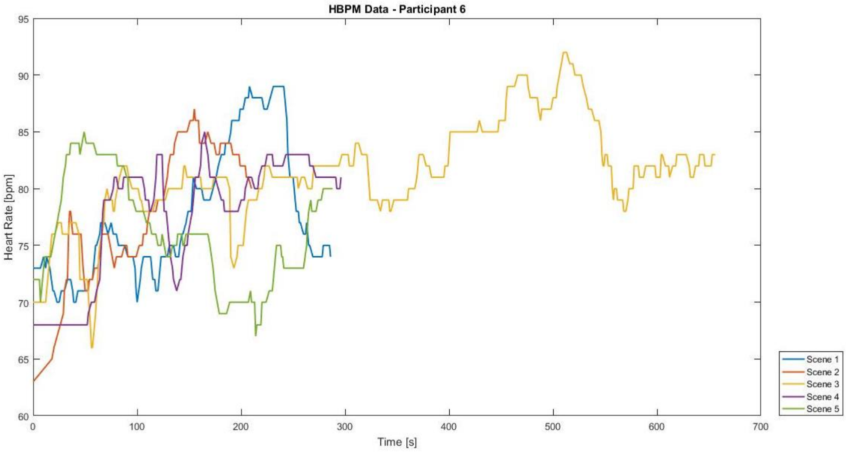
Participant 4



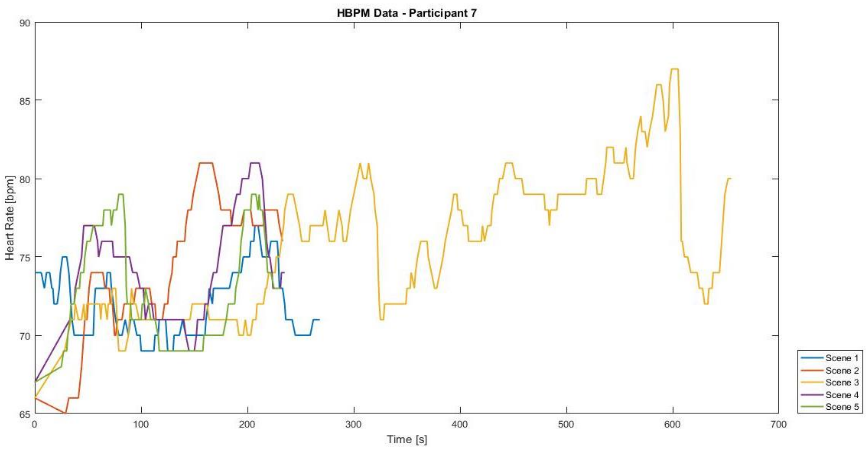
Participant 5



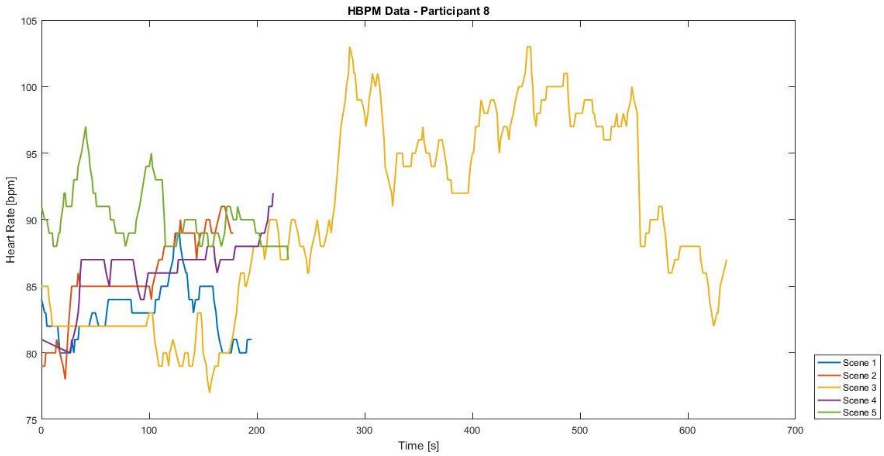
Participant 6



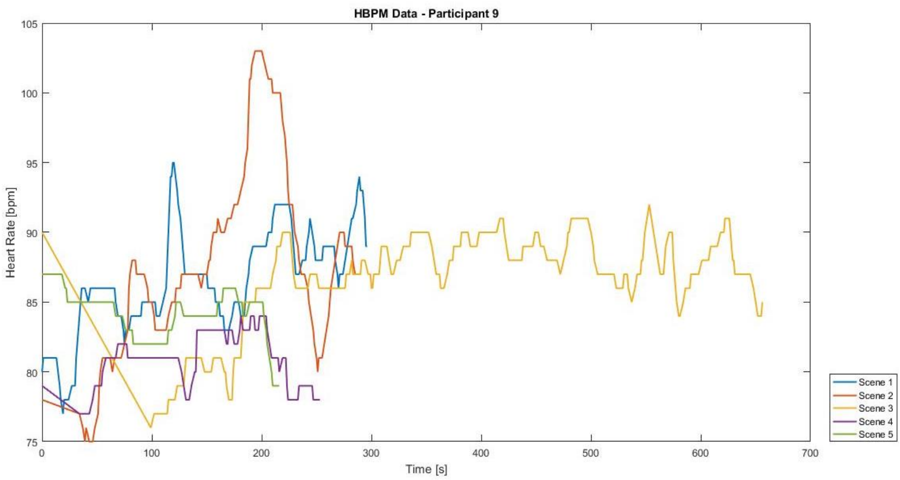
Participant 7



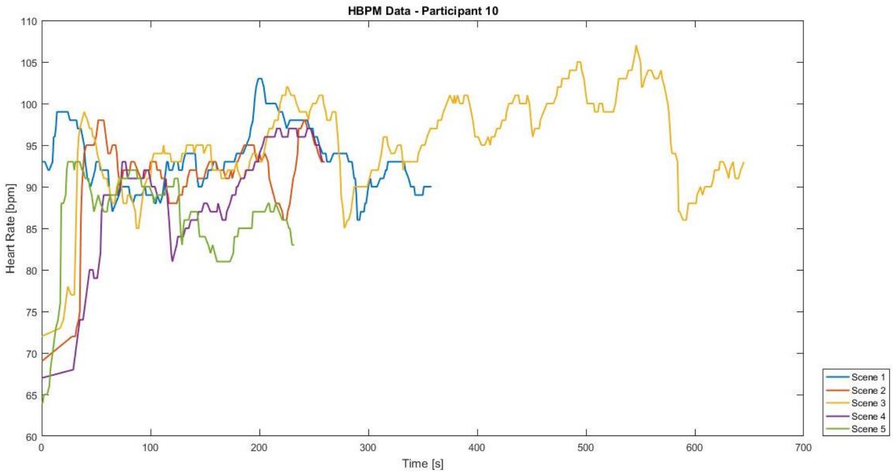
Participant 8



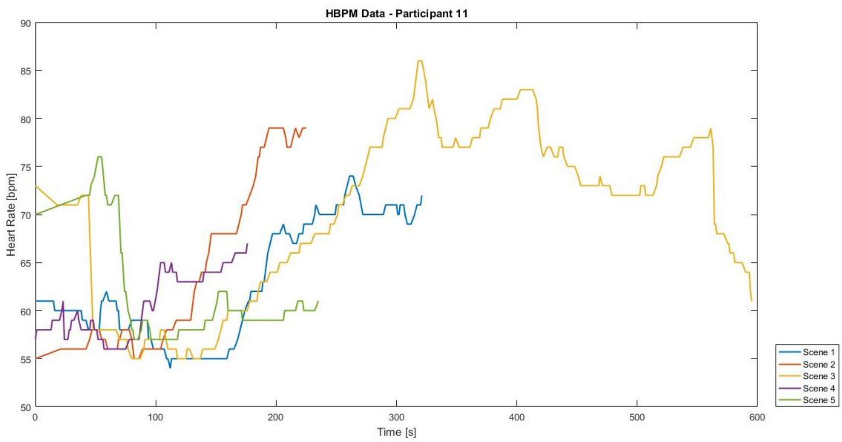
Participant 9



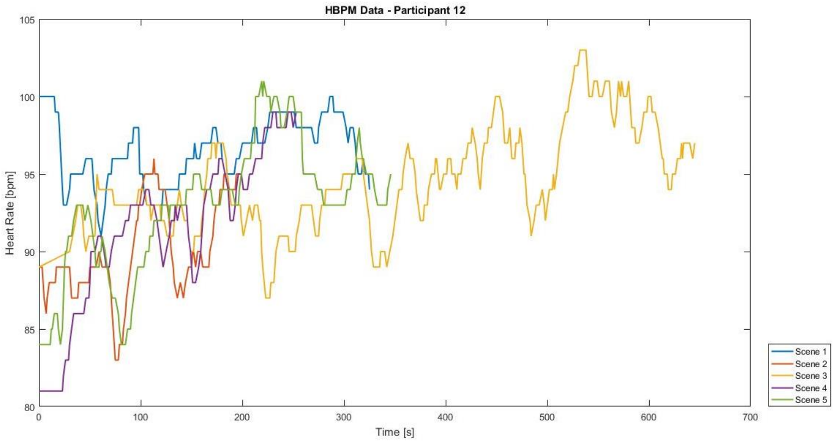
Participant 10



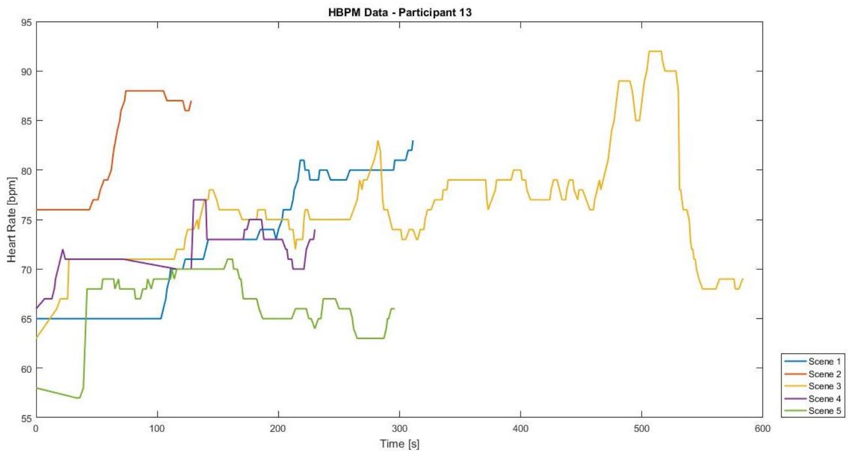
Participant 11



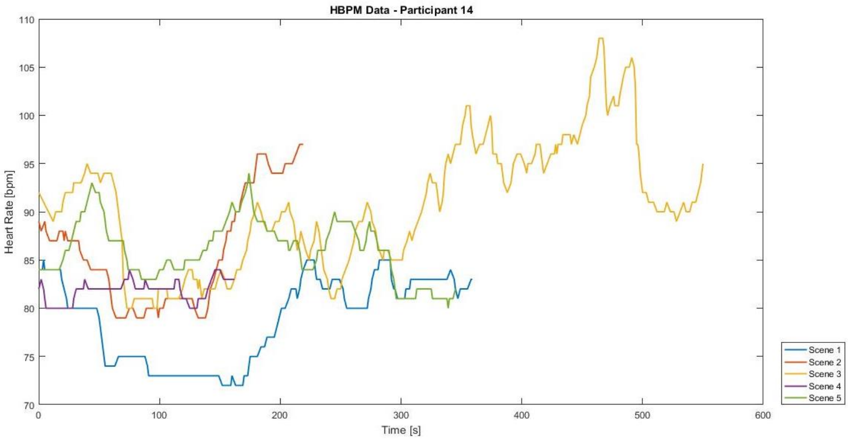
Participant 12



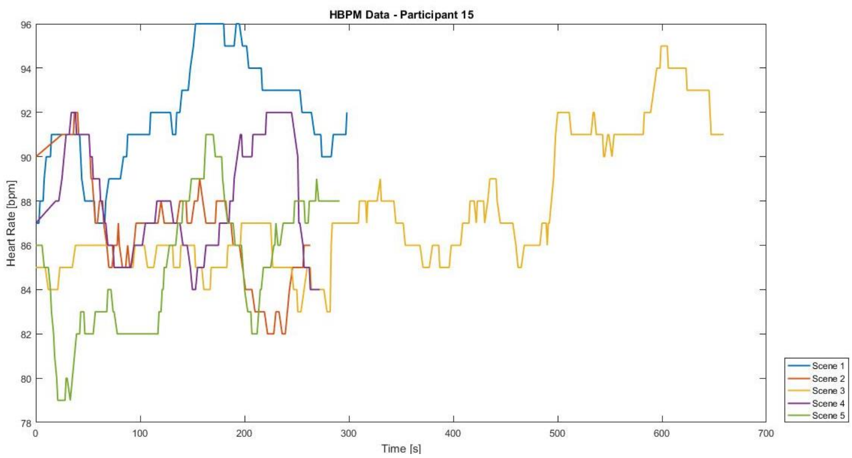
Participant 13



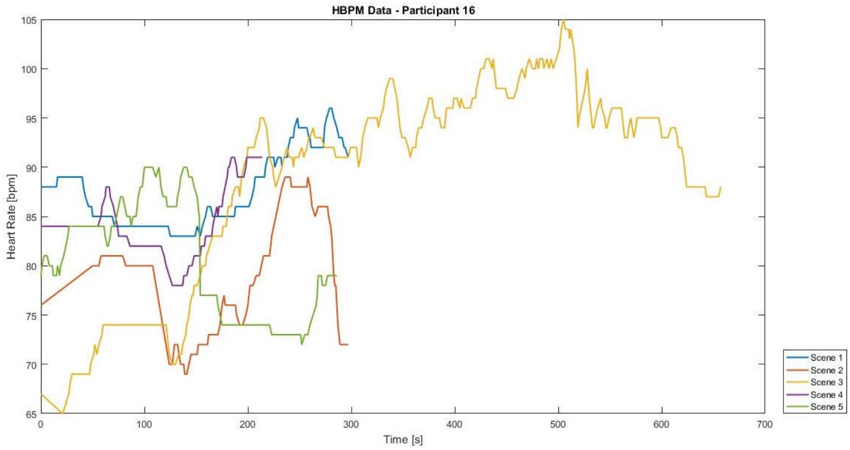
Participant 14



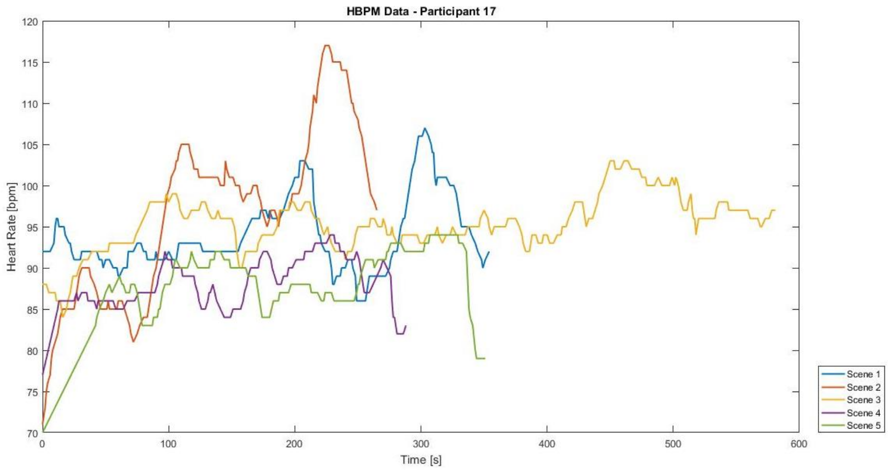
Participant 15



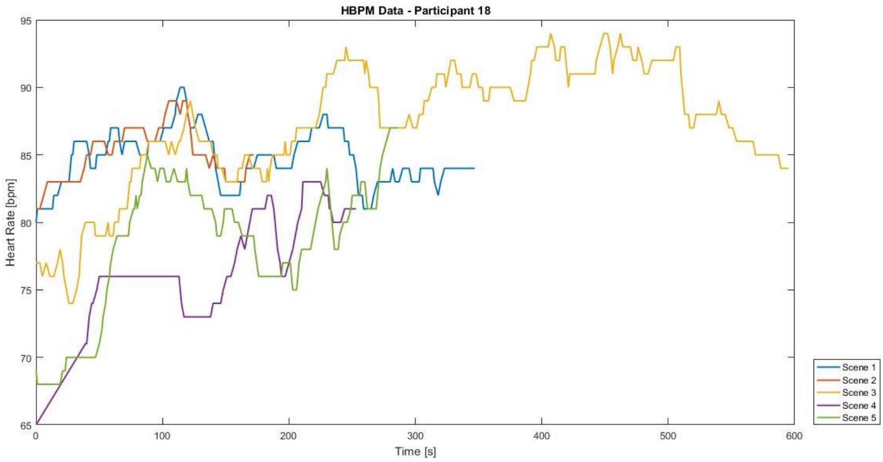
Participant 16



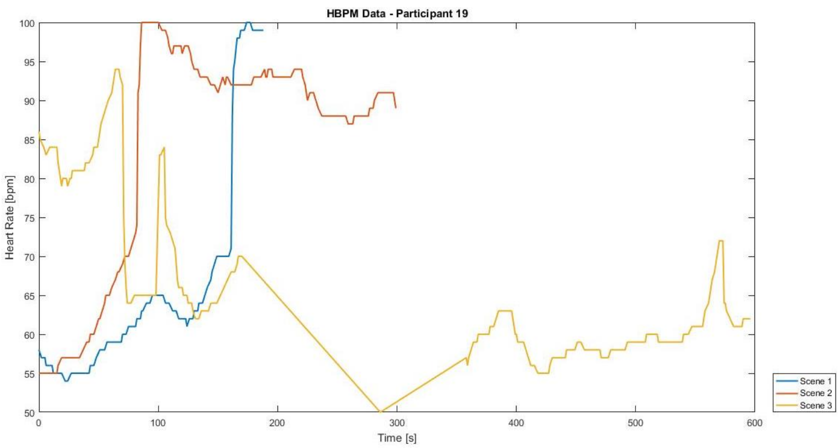
Participant 17



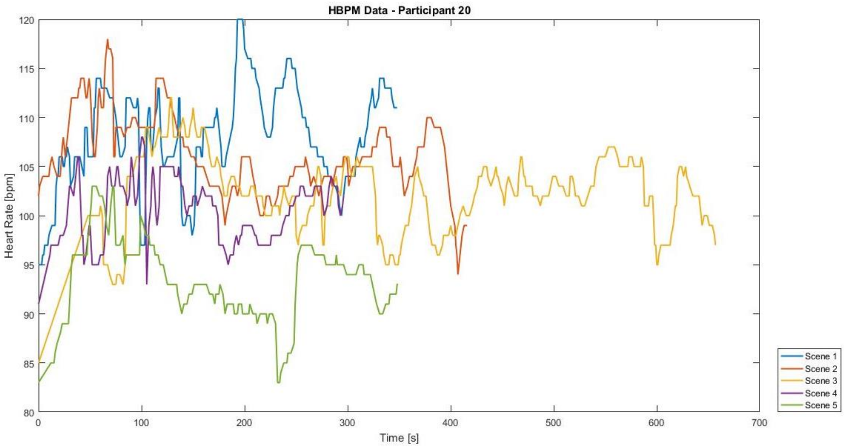
Participant 18



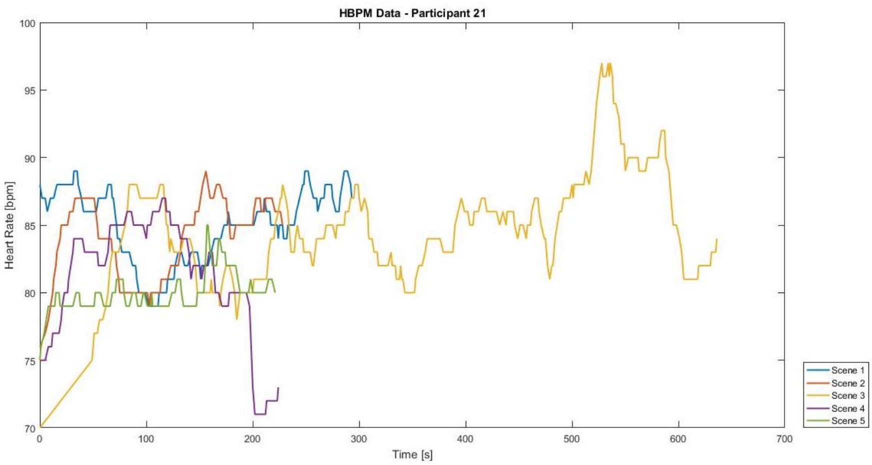
Participant 19



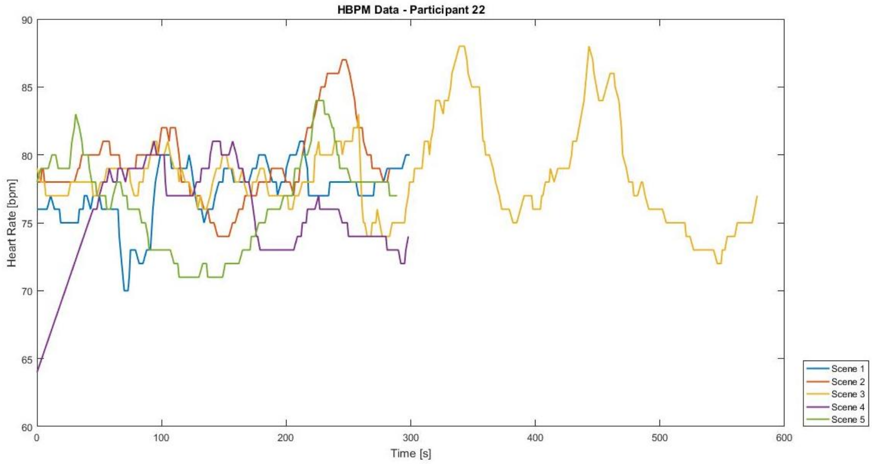
Participant 20



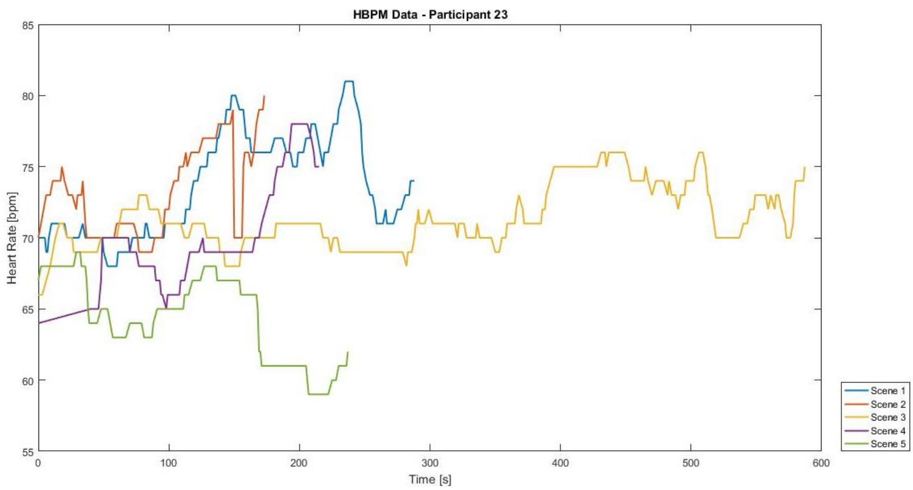
Participant 21



Participant 22



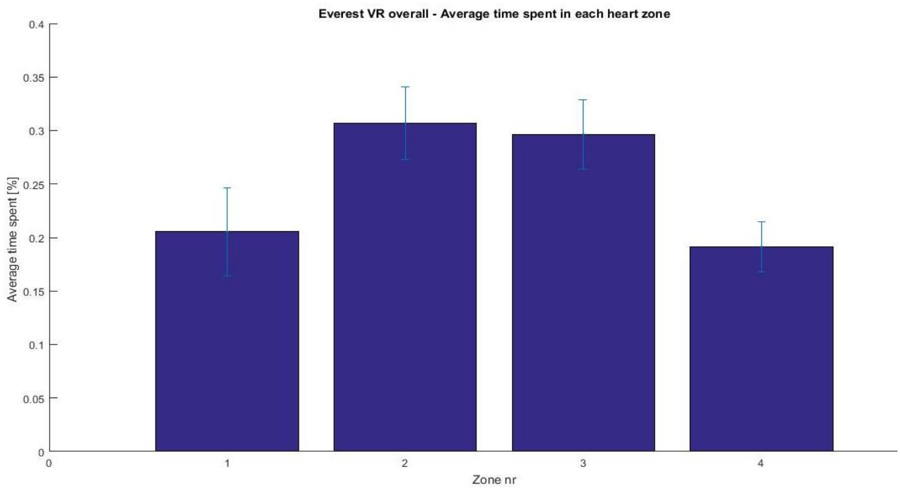
Participant 23



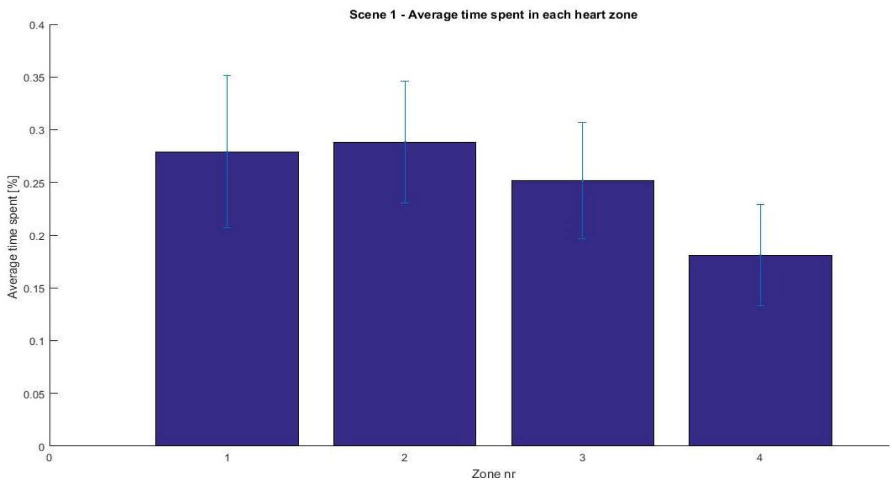
Appendix J - Physiological Results: Average Time spent in different HR-Zones

Average Time Spent in Each Heart Zone

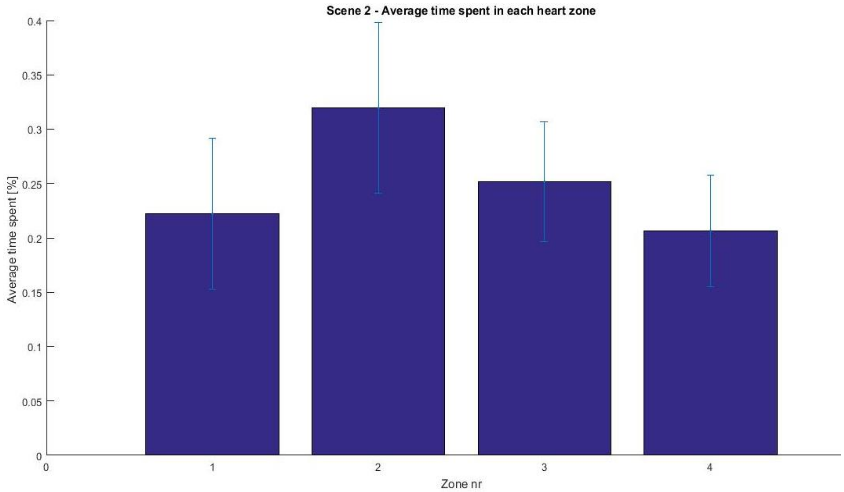
Everest – Overall Average



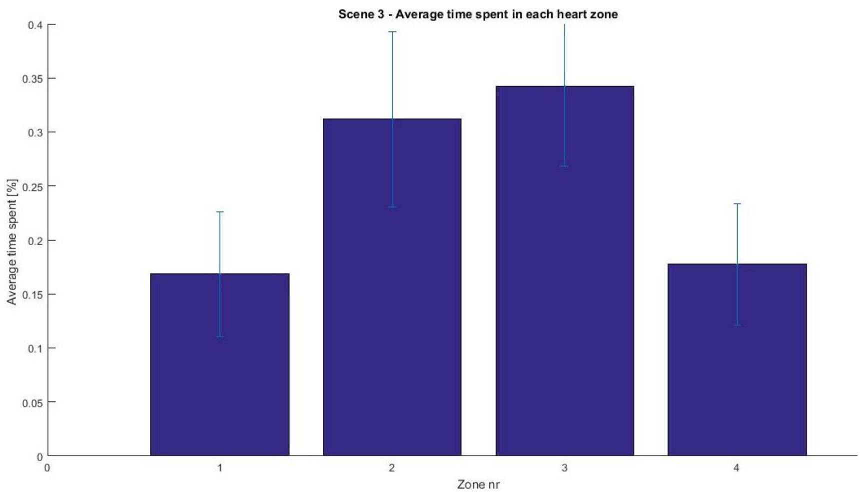
Scene 1 – Average time spent in each heart zone



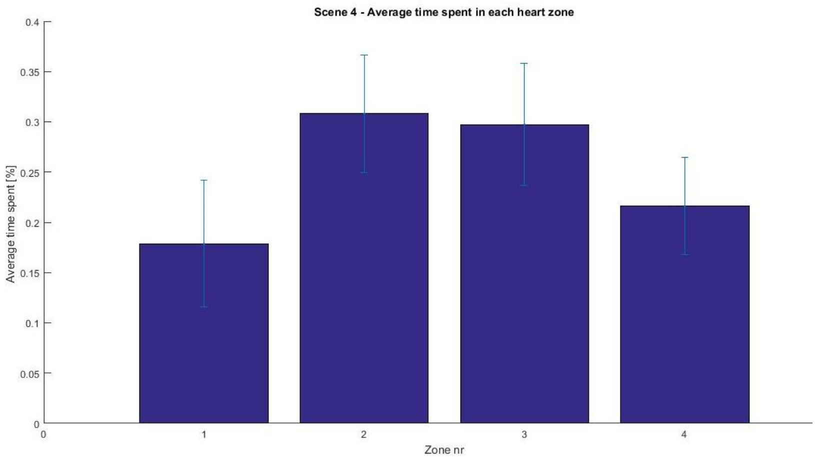
Scene 2 - Average time spent in each heart zone



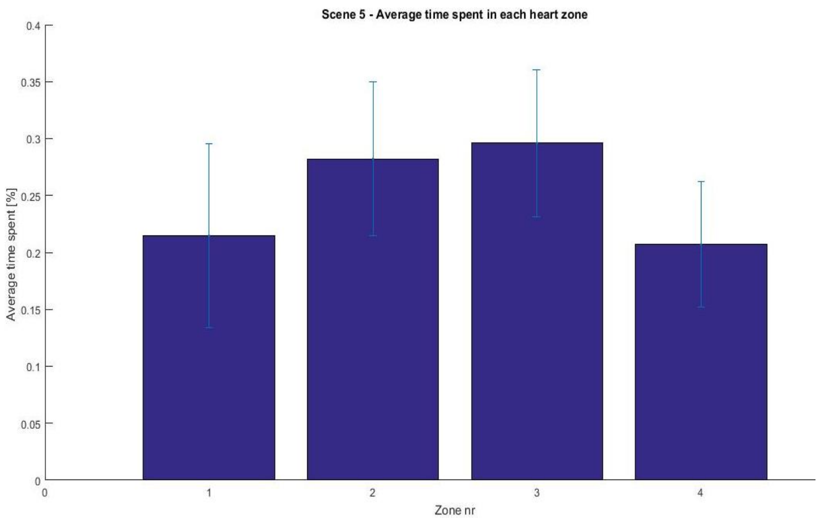
Scene 3 - Average time spent in each heart zone



Scene 4 - Average time spent in each heart zone

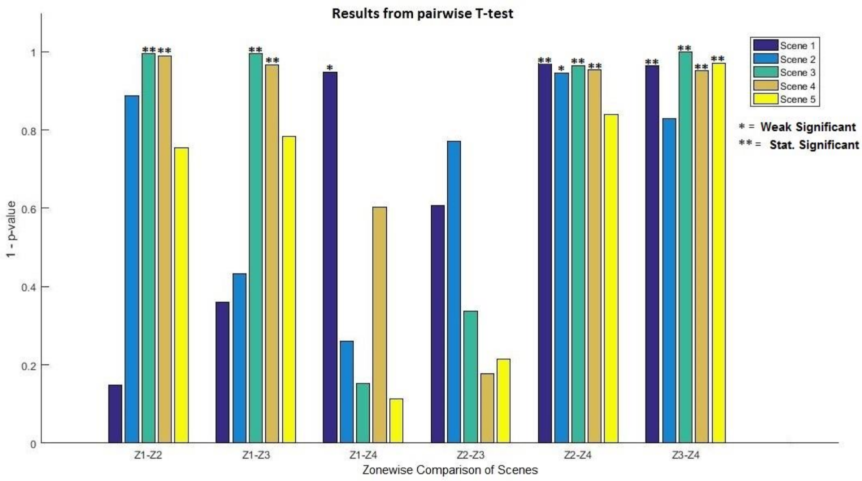


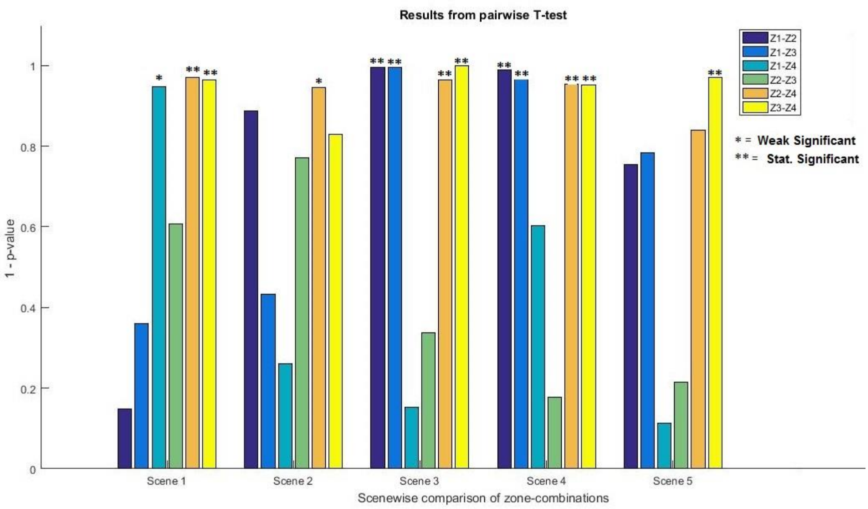
Scene 5 - Average time spent in each heart zone



Appendix K - Physiological Results: Pairwise T-test Between HR-zones

Results from pairwise T-test between HR-zones





Z1-Z2	p	h
S1	0.8519	0
S2	0.1130	0
S3	0.0047	1
S4	0.0106	1
S5	0.2464	0

Z1-Z3	p	h
S1	0.6395	0
S2	0.5676	0
S3	0.0058	1
S4	0.0344	1
S5	0.2175	0

Z1-Z4	**p**	**h**
S1	0.0521	0
S2	0.7403	0
S3	0.8470	0
S4	0.3981	0
S5	0.8876	0

Z2-Z3	**p**	**h**
S1	0.3919	0
S2	0.2290	0
S3	0.6622	0
S4	0.8235	0
S5	0.7845	0

Z2-Z4	**p**	**h**
S1	0.0294	1
S2	0.0541	0
S3	0.0360	1
S4	0.0453	1
S5	0.1600	0

Z3-Z4	**p**	**h**
S1	0.0353	1
S2	0.1717	0
S3	0.00347	1
S4	0.0477	1
S5	0.0303	1