Quality of Experience Evaluation for Haptic Multimedia Applications

Abdelwahab Hamam

Thesis submitted to the
Faculty of Graduate and Postdoctoral Studies
in partial fulfillment of the requirements
for the Doctorate in Philosophy degree in
Electrical and Computer Engineering

School of Electrical Engineering and Computer Science
Faculty of Engineering
University of Ottawa

Your file Votre référence
ISBN: 978-0-494-98612-7

Our file Notre référence
ISBN: 978-0-494-98612-7

ABSTRACT

Haptic-based Virtual Reality (VR) applications have many merits. What is still obscure, from the designer's perspective of these applications, is the experience the users will undergo when they use the VR system. Quality of Experience (QoE) is an evaluation metric from the user's perspective that unfortunately has received limited attention from the research community. Assessing the QoE of VR applications reflects the amount of overall satisfaction and benefits gained from the application in addition to laying the foundation for ideal user-centric design in the future. In this thesis, we address certain issues and concerns regarding QoE of virtual environments.

In essence, we propose a taxonomy for the evaluation of the QoE for multimedia applications and in particular VR applications. The taxonomy classifies QoE related parameters into groups. The groups' organization is generated from the definition we have adopted for the QoE which is the Quality of Service (QoS) plus the user experience (UX). We model this taxonomy using first mathematical modeling based on weighted averages and then a Fuzzy logic Inference System (FIS) to quantitatively measure the QoE of haptic virtual environments. We test both models conducting user study analysis to evaluate the QoE of a VR application. These models serve as engines that facilitate the calculation of QoE with minimal amount of users.

We specifically attend to the issue of the new media, haptics, within the context of increasing the QoE of virtual environments (VE). This special attention is important for comparing the effect of tactile and kinesthetic feedback on the QoE. In accordance, we investigate a particular topic that seems to have a colossal effect on QoE throughout our analysis, which is fatigue.

Our analysis involved users' studies since the main focus is on the user. The QoE for virtual environments is in its primary stages. This thesis tackles issues that are vital in dealing with and understanding the importance of QoE. The various results suggest a positive user's disposition toward haptics and virtual environments, yet there will always be obstacles and challenges such as fatigue that if minimized will enhance the QoE of haptic-based applications.

ACKNOWLEDGMENTS

I would like to pay my respect to my late supervisor Dr. Nicolas D. Georganas whom he counseled me throughout most of my graduate studies, may he rest in peace. I worked for Dr. Georganas in my undergraduate program in the MCRLab. After completing my degree, he encouraged me to enter into graduate school. Our group worked on an exciting project involving the latest haptic and 3D technologies. I obtained my Master's degree under his supervision. Immediately after, he encouraged me to enter the Ph.D. program. I was under his supervision until he passed away during the summer of 2010. Dr. Georganas was a true inspiration to me. His dedication to research, humor, and constant feedback will be missed. I am grateful for the privilege of knowing him and learning from him.

I would also like to express my everlasting gratitude to my supervisor Dr. Abdulmotaleb El Saddik for his support and guidance. I met Dr. El Saddik when I was doing my undergraduate studies. I was a cooperative program student and I was doing development work at the MCRLab, University of Ottawa. Later on he was the professor for few of my courses. When I enrolled in the graduate program, I worked in the DISCOVER Lab, University of Ottawa where we sometimes discussed my work among other various topics. Dr. El Saddik became my co-supervisor during my Ph.D. program and later on my supervisor, due to the unexpected and unfortunate incident mentioned above. Since my undergraduate program till the end of my doctorate degree Dr. El Saddik's excellent mentorship and assistance allowed me to overcome the challenges that I have encountered and, thus, complete this thesis successfully. I thank him for being a great supervisor, teacher, and friend.

I would like to thank my proposal committee for their time, Dr. Emil Petriu and Dr. Dorina Petriu. Your thorough revision of my work allowed me to enhance it and rethink angles I haven't thought of before.

I would like to thank all my DISCOVER Lab and MCRLab colleagues. Your valuable friendship and insight always provided me with the push I needed. I hope our paths will cross again one day. Special thanks to Dr. Mohamad Eid and Dr. Jongeun Cha for the exciting discussions and feedback. Special thanks also goes to Mr. Francois Malric, the DISCOVER Lab ex-manager, for providing us with a superb working environment and fixing every hardware, software, or network problem we encountered.

I want to send my greetings to very special friends that I have met during my graduate studies: Saeid Nourian, Houman Rastgar, Mohamad Eid, Oana Frunza, Ahmed Jedda, Razib Iqbal, Behnoosh Hariri, Dewan Tanvir, Naim il Far, Fayez il Far, Alistair Kennedy, Houda Rachidi, Fatima Nargesian, Maude Amyot Bourgeois, Majdi Al Rawashdeh, Omar and Bashar Assad, Ahmad Eter…. plus many others.

This work would not have been possible without the love and support of my parents Abdelbadie and Eman. I dedicate this thesis to you. I would like to also dedicate this thesis to my brother Zaid and my sister Lama and my two beautiful nieces Kayan and Ayah. Thank you all for being ever so understanding and always there for me.

I would like to thank Allah for blessing me with health, mind, food, shelter, strength, and faith to complete this thesis.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

Acronym	Description
CHAVE	Collaborative Haptic Audio Visual Environments
CVE	Collaborative Virtual Environment
DHVE	Distributed Haptic Virtual Environment
DOF/DF	Degrees Of Freedom
FCM	Fuzzy C-Mean
FIS	Fuzzy Inference System
HAVE	Haptic Audio Visual Environments
HMD	Head Mounted Display
HMI	Human Machine Interaction
ITU	International Communication Union
IPTV	Internet Protocol TeleVision
JND	Just Noticeable Difference
MF	Membership Function
MOS	Mean Opinion Score
PCA	Principle Component Analysis
QoBiz	Quality of Buisness
QoD	Quality of Design
QoE	Quality of Experience
QoP	Quality of Perception
QoPer	Quality of Performance
QoS	Quality of Service
RMSE	Root Mean Square Error
SVE	Shared Virtual Environment
TCT	Task Completion Time
UX	User eXperience
VE	Virtual Environment
VR	Virtual Reality

CHAPTER 1

INTRODUCTION

1.1 Multimedia Applications and Virtual Reality

The evolution of multimedia applications is increasing rapidly. The last decade has witnessed better graphics, bigger screens, and improved surround audio, to name a few. With this evolution a newer field of multimedia has erupted. Virtual Reality (VR) is a high end user interface that intends to shift the user's attention from the physical world into the multimedia application which could be viewed as an alternate reality. For that reason, VR applications tend to utilize realistic looking graphics and state of the art hardware such as 3D glasses. To be fair, VR applications emerged from the 1960s but it is recently that the technological advances enabled the applications to gain momentum within the industry.

(Burdea and Coiffet 2003) state three properties of a VR interface, they refer to them as the three I's of VR: Immersion, Interaction, and Imagination. Immersive properties materialize from the hardware and software technologies used. That is why bigger screens, 3D graphics, and body tracking devices are associated with VR applications. Sensorial channels are another aspect that links VR properties to immersion. VR interfaces tend to use most of the human's senses. Interactive properties tend to provide users with real time interactivity. The user is able to become part of the world, sometimes through an avatar, and provide input to modify the surrounding objects. Imagination provides the VR medium with a sense of purpose. The user is utilizing the application to solve a problem or for entertainment purposes which stimulate the human imagination.

As mentioned, VR interfaces tend to simulate the human senses. There are five sensorial modalities for humans: visual, auditory, tactile, smell, and taste. Vision and audition are utilized in VR applications and in traditional multimedia applications. Although they could be more advanced in VR applications but they are not new. Smell and taste are rarely used currently. They would certainly add more realism to VR applications and should be considered in the future. What we will be focusing on in this thesis is the sense of touch of humans, or as we might refer to as haptics.

Haptics, a term originating from the Greek language, refers to the science of manual interactions with the ambient environment through touch, including exploration for information extraction and/or manipulation for modifying the environment. Haptic technology has changed the way humans interact with computers. Incorporating the sense of touch into virtual environments (VE) has opened a new trajectory of interactive applications ranging from medical simulations and rehabilitation to more realistic video games. Gradually more and more applications will utilize haptic interfaces and they will be geared toward the three modals feedback namely: visual, auditory, and touch feedback. The promising advantages of haptics, audio, and video environments to the user are more realism, more excitement, and better manipulation of objects (El Saddik, 2007).

Haptic interfaces can be classified into different categories: point of contact (one point or multiple point of contact), mobility, and type of feedback (tactile or kinesthetic as will be seen in Chapter 6). Haptic devices have seen a surge in the 1990's. The first attempt to commercialize haptic devices was by SensAble Technologies cofounder Thomas Massie. In his master's thesis, entitled "Initial Haptic Explorations with the Phantom: Virtual Touch through Point Interaction", (Massie 1996) describes the Phantom haptic device which possess a pen-like structure and can interact with virtual objects through poking with the tip of the pen. SensAble Phantom will be utilized in most user studies in this thesis due to its popularity. The Phantom Desktop version of the device is shown in Figure 1.1.

Figure 1.1. SensAble technologies Phantom Desktop.

1.2 Motivation

Quality of Experience (QoE) is an emerging term that indicates that what ultimately matters in multimedia applications is the user's perception of the application. This is a different approach than the previously well-established Quality of Service (QoS) which focused on functional and technical requirements of the system. This shift in evaluation paradigms gave the user more abstract power from just a user to an evaluator as well. That is why users' opinions and feedback is ever more important as a design guideline and marketing strategies to commercial companies (De Marez and De Moor 2007).

De Marez and De Moor list two points that is making the shift from QoS to QoE challenging:

1. Concrete definition and clear conceptualization of QoE
2. Lack of a good QoE-measurement approach

While QoS and usability are key factors for defining QoE, QoE is more than just assessing the QoS that an application provides to users (Jain 2004). QoS would be part of the assessment, whether it is jitter and delay of the network or intermodal/intramodal synchronization of multimedia contents. Yet, there are still other parameters to consider such as ease of usage, rendering quality, and measurement of fatigue. These parameters represent the user experience and are subjective in nature. Both the QoS and the users' experience compose the overall QoE which in turn will reflect the perceptual value of multimedia applications.

The future of multimedia design metrics will eventually be based on QoE assessment rather than QoS assessment. In their 2005 ACM multimedia retreat, (Rowe and Jain 2005) indicate that users will be the focus of assessment and their experience is the basis of building future multimedia applications. There are challenges arising though which revolve around the subjectivity of the user, the diversity of the applications, and the variation of the environmental context (Jain 2004). Eventually, users are the ones who are going to be working with the application and their perceived experience is the ultimate goal.

1.3 Research Statement

Given all the promising advantages of haptics and VR technologies (from sensorial enriching and natural interaction to promised excitement), very few research efforts have objectively analyzed the way haptics improves the user's experience. Quality of Experience (QoE) is an approach that describes the evolving reality that what ultimately matters in a multimedia system is how users perceive its performance (Microsoft 2007). In contrast to the traditional methodology of Quality of Service (QoS) (which manages network configuration and performance and is ineffective to quantify the user experience), QoE examines all elements that influence user's perception of the interaction. The QoE approach is based on optimizing and monitoring the actual user experience by considering subjective quality parameters such as user satisfaction, usefulness, fatigue, distraction, among others.

Hence, simply put, our research question can be stated as "How can we quantitatively measure users' Quality of Experience (QoE) of Virtual Reality Applications?"

Two main points falls under the research question:

- Propose taxonomy to describe QoE parameters of Haptic Audio-Visual Environments (HAVE)
- Propose a model to quantitatively evaluate QoE using a subset of these parameters

Although VR elements and haptics increase excitement and enjoyment and import more perception information to the user than a regular application, there is no inclusive model that can assess the QoE of VR applications.

1.4 Thesis Contributions

The thesis main goals consist of the following:

- Design and development of a taxonomy for parameters related to the evaluation of QoE in the VR domain. This is performed through an intensive analysis of the literature for QoE related projects, ideas, definitions and thoughts
- Analysis, design, and development of a mathematical model to quantitatively evaluate QoE
- Analysis, design, and development of a fuzzy logic modeling system to evaluate QoE
- Analysis of the effect of haptics (kinesthetic and tactile) in contributing to the QoE of certain multimedia applications
- Analysis of users' fatigue while using haptic devices as the lack of undesired fatigue is a major contributor to the QoE of an application

1.5 Publications and Scholastic Output from this Thesis

Refereed Journal Papers:

[1] **A. Hamam**, A. Alghamdi, and A. El Saddik *A Quality of Experience Model for Haptic Virtual Environments*, submitted to ACM Transactions on Multimedia Computing, Communications and Applications (ACM TOMCCAP) (revised edition is submitted)

[2] **A. Hamam** and A. El Saddik, *Towards a Mathematical Model for Quality of Experience Evaluation of Haptic Applications*, IEEE Transactions on Instrumentation and Measurement, accepted to appear.

[3] **A. Hamam**, M. Eid, and A. El Saddik, *Effect of Kinesthetic and Tactile Haptic Feedback on the Quality of Experience of Edutainment Applications*, Journal of Multimedia Tools and Applications, Volume 67, Issue 2, DOI 10.1007/s11042-012-0990-7, pp. 455-472, 2013.

Refereed Conference Papers:

[1] **A. Hamam** and A. El Saddik, *Evaluating the Quality of Experience of Haptic-based Applications through Mathematical Modeling*, IEEE International Workshop on Haptic Audio Visual Environments and Games, HAVE'12, Munich, Germany, pp. 56-61, 2012.

[2] **A. Hamam**, F. Alsulaiman, A. El Saddik, and N.D. Georganas, *Deducing User's Fatigue from Haptic Data*, Proc. of the International Conference on Multimedia. Firenze, Italy, pp. 1083-1086, 2010.

[3] **A. Hamam**, N. D. Georganas and A. El Saddik, *Effect* of *Haptics on the Quality of Experience,* IEEE International Symposium on Haptic Audio-Visual Environments and Games (HAVE), pp. 1-6, 2010.

[4] **A. Hamam** and N.D. Georganas, *A Comparison of Mamdani and Sugeno Fuzzy Inference Systems for Evaluating the Quality of Experience of Hapto-Audio-Visual Applications*, Proc. IEEE Workshop on Haptic Audio Visual Environments and their Applications (HAVE'08), Ottawa, pp. 87-92, 2008.

[5] **A. Hamam**, M. Eid, A. El Saddik, and N.D. Georganas, A Fuzzy Logic System for Evaluating Quality of Experience of Haptic-based Applications, EuroHaptics, Madrid, pp. 129-138, 2008.

[6] **A. Hamam**, M. Eid, A. El Saddik, and N.D. Georganas, *A Quality of Experience Model for Haptic User Interfaces*, Haptic User Interfaces in Ambient Media System, Quebec City, Canada, pp. 1-6, 2008.

1.6 Thesis Organization

The thesis is organized as follows. In Chapter 2, we overview the related work in QoE in multimedia research field incorporating VR, including defining QoE and relating it to the QoS. Chapter 3 illustrates the taxonomy of QoE parameters. This taxonomy models the QoE and divides it into distinctive parameters which an evaluator can select according to their specific application. A case study involving a haptic application is provided in this chapter. The case study illustrates how the taxonomy can be applied in a HAVE application. Chapter 4 and Chapter 5 describe two ways for evaluating QoE of multimedia applications: mathematical modeling and fuzzy inference systems. The distinction between those two methods is explained with the evaluation study applied to each methodology to compare the results. We administrate an evaluation study in Chapter 6 in which we focus on the haptic media. We test the effect of haptics on the QoE of certain multimedia applications. Two types of haptic feedback are examined and compared; kinesthetic and tactile feedback. A very important QoE parameter is discussed in Chapter 7. Fatigue's effect on QoE is significant, therefore, detecting fatigue is the first step in rectifying any ramifications that could arise from such an effect. Novel ways to detect user's fatigue is detailed with user's subjective feedback used to validate the results. Finally, we summarize and conclude the thesis in Chapter 8 and outline prospects for future work.

CHAPTER 2

BACKGROUND AND RELATED WORK

2.1 Definition of Quality of Experience

Quality of Experience (QoE) is relatively a new topic that was introduced to the research community. Throughout the years the research community has focused on the Quality of Service (QoS) as the standard quality concept. QoS is well defined with constant metrics. The primary question to consider within the literature is a formal definition of the Quality of Experience (QoE)? What is the QoE? How does it differ from the QoS? The literature varies in the definition of the QoE. Definitions tend to be similar but are not exact. For instance, the following are some of the definitions found in the literature:

"QoE focuses on the user and is considered...as the collection of all the perception elements of the network and performance relative to expectations of the users. The QoE concept applies to any kind of network interaction such as ...Thus QoE may be seen as influenced by three factors: the user's content preferences with respect to needs and goals, the network over which the content is accessed, and the device with which the user connects to the network." (Muntean et al. 2007)

“QoE is a multi-dimensional construct of perception and behaviour of a user, which represents his/her emotional, cognitive, and behavioural responses, both subjective and objective, while using a system." (Wu et al. 2009).

"The overall acceptability of an application or service, as perceived subjectively by the end-user.

- Note 1 - Quality of Experience includes the complete end-to-end system effects (client, terminal, network, services, infrastructures, etc.).
- Note 2 - Overall acceptability may be influenced by user expectations and context." (ITU-T Rec. 2008)

"QoE is a subjective measure from the user perspective of the overall value of the provided service or application" note the authors say there is a lack of quantitative description and exact definitions of QoE, one particular difficulty consists in matching subjective quality perception to objective measures. (Hobfeld et al. 2007)

"User's perceived experience of what is being presented by the application layer where the application layer acts as a user interface front-end that presents the overall results of the individual QoS." (Siller and Woods 2003a)

"QoE can be defined as the qualitative measure of the daily experience the customer gets when he uses the services he is subscribed to including experiences such as outages, quality of picture, speed of the high-speed Internet service, latency and delay, customer service, etc. The better the consumer's experience, the higher his QoE. And that has an effect on customer loyalty." (Kumar 2005)

"QoE is not a metric but a concept comprising all the elements of a subscriber's perception of the network and its performance and how they meet expectations. QoS is the ability of the network to provide a service with an assured service level. QoE is how a user perceives the usability of a service when in use – how satisfied he or she is with a service. The term QoE refers to the perception on the user about the quality of a particular service or network." (Nokia 2006)

What is common between the definitions is threefold constructs, listed below:

- The definition centres around the user, hence the term user centric evaluation commonly found in many QoE papers
- Subjectivity is mentioned explicitly or through the term user's perception or qualitative parameters
- User's experience, and in effect the concepts that define experience

The following subsections discuss the literature foundation of the aforementioned constructs of the QoE definition.

2.1.1 Human (User) Factors

This subsection will focus on human factors in VR related disciplines. (Stanney et al. 1998) provided an in depth discussion about human factors issues in VE (extending Stanney 1995), which they have divided into three subtopics: human performance efficiency in virtual worlds, health/safety issues, and social implications of VE technology.

Human performance measures in VE can be maximized by examining the following factors: task and user characteristics, VE design constraints related to human sensory limitations, and effective integration of multimodal interactions. The authors are associating the effectiveness of a VE system with assessing human performance efficiency, which can be predicted by some factors including: the navigational complexity of the VE, degree of presence, and benchmark tests of users' performance.

The health and safety issue, if ignored, could result in discomfort, harm, or even injury. The authors divide these issues into direct and indirect effects. Direct effects in turn can be divided into direct microscopic effects (e.g. ergonomic design or high volume irritating the ear) or direct macroscopic effects (e.g. trauma or cyber sickness). An example of indirect effect is the eye-hand coordination after the VR experience.

Social implications of VE are often neglected. The negative social implications of VE are not fully understood. Misuses of VE could lead to negative social implications though such as violence.

Another reference that discusses human factors in VR in depth is Chapter 7 of the Burdea and Coiffet book (Burdea and Coiffet 2003) entitled Virtual Reality Technology. They give a formal definition: "VR human factors studies consist in a series of experiments, performed under very rigorous conditions, aimed at determining users' response to VR technology, VR technology usability, VR user safety, and the related societal impact of VR".

Hence human factors research is divided into four parts:

- Usability studies: look at ways to improve system or application design to make them easy to use
- User performance: conducted to measure the user's response to a given simulation and a particular hardware system

- User safety studies: conducted to better understand simulation sickness causes and effects as well as increase the user's safety (minimize physical injury)
- Sociological studies: measure factors related to effects of VR on society

Three of the areas above are shared with the previous reference (Stanney et al. 2003). The new addition here is the usability studies, which could be a subset of performance. Moreover, (Wann and Mon-Williams 1996) stress the importance of user perception in 3D VR environment and that performance might be enhanced by positive perception. To that effect, designers should consider performance requirements of the user and base their design on what is essential, desirable, and optimal.

Pertaining to the clinical applications of VR, (Lewis and Griffin 1998) discuss factors that need to be considered in the design phase to minimize any possible side effects on the human operator. The factors are divided into user characteristics (physical, past experience, and personality), system characteristics (display and delay), and task characteristics (movement, visual image, and interaction between VE and the user). As well, the paper discusses factors affecting performance, presence, and consideration in the implementation of a VR application in the clinical domain.

A human study of virtual environment (VE) that is divided into three stages is conducted by (Mania and Hawkes 2003). Stage one is a classification of current VE system and methods with interacting with the system. The conclusion is that when humans unnaturally interact with the VE system (compared to real life situation), it could affect the performance. Stage two is an experimental setup to test the effect of different environments (VE, HMD, and desktop) with the real world. Tests were administered regarding memory, presence, and simulation sickness. Stage three is similar to stage 2 but with rigorous conditions and statistical analysis. Regarding the HMD design, (Bolas 1994) focuses on the ergonomic design and comfort level for users wearing the hardware in an immersive display.

(Chen et al. 1998) summarize a special issue about human factors in VE into three categories: Realism which is discussed by (Hoffman et al. 1998), direct manipulation and learning (Sastry and Boyd 1998), and navigation which is provided by (Mukherja 1998).

Human factors for haptic interfaces and their design have been addressed in the literature (Jandura and Srinivasan 1994, Tan et al. 1994, Ellis et al. 1996, Zadeh et al. 2007). Mainly, user performance has been tested to detect and control force and torque elements to aid the design of the hardware. As part of the design as well, quantitative analysis of human threshold were performed to document the just noticeable difference (JND) of force in the human sensory system. (Tan et al. 1994) state that human haptic system has limitations that can be exploited, hence human studies are essential. (Smith 1997) goes further by not just analyzing the force resolution of the human physiological system but also including safety issues of haptic devices and a summary of the research done on the computers for the blind and handicapped, which uses haptic elements. Human factors in haptic contact of pliable surfaces are considered in (Vicentini and Bottrui 2009). Psychophysical experiments to test speed and posture position on the performance of surgeons in a task of pliable surface contact were conducted. Two force-feedback devices on a virtual surface were also employed to test the human capabilities on terms of penetration depth and responsiveness.

(Deml 2007) provides a guideline for the design of tele-presence system from a human factors point of view. The points discussed are bimanual interaction with the environment (two haptic devices), degrees of freedom (DOF) control, and presence/performance relations. (Pongrac et al. 2007) conducted experiments to see the effects of the field of view variation as well as live video streams on the human performance in a tele-presence environment.

Finally when discussing human factors, it is essential to view the ethical consideration and complexity. (Whalley 1994) warns that ethical guidelines must be accounted for in VR medical application as VR popularity as a research tool in medicine increases. (Behr et al. 2005) list the ethical problems that may be encountered in VR research: motion sickness, information overload, intensification of experience, and difficulties with re-entry to the real world. They try to answer the following question, how can the ethical guidelines applied to psychological research be extended to VR technology research?

2.1.2 Subjectivity of the User

It is clear from the previous section that user characteristics vary the performance of individuals and their perception. (Burdea and Coiffet 1998) state that there is no comprehension model of human behaviour that exists due to its multi dimensionality and large individual variability. Especially in the case of VR, where there are more parameters involved compared to other human-machine interaction (HMI) systems, it is difficult to assess such interactions. Thus determining the performance of VR simulation would be subjective and hard to quantify mathematically.

Due to the subjectivity of the user, (Burdea and Coiffet 1998) list drawbacks of the users' performance measures in VR systems. These drawbacks (e.g. noise, latency) need to be taken into account when performing HMI VR experimentation to assure the following

- Sensitivity: Data should discriminate between novice and expert users
- Reliability: Data should be repeatable and have internal consistency
- Validity: Data are truthful to the subjects' actions

Data collected is divided into objective and subjective data. Objective data refer to variables that can be measured (performance measures) such as task completion time, while subjective data refer to the preference and perception of the subjects using the hardware. (Tullis and Albert 2008) warn that even though subjective data is relative to each participant, the name implies that it lacks objectivity. On the contrary the data is objective to the evaluator point of view, in the sense that it can be captured through the experiment.

2.1.3 User Experience

We live in a world which we consider real. Our surrounding is real and anything we touch is real to us. Human experience is measured against the reality that time and space is a given (Flach and Holden 1998). Similarly in VR we have constructs of "real-time" and "real distance" built inside us on which we can compare the "virtual time" and "virtual distance" in the virtual world.

The paper (Flach and Holden 1998) is a philosophical paper that discusses how to view reality and on what aspects do users base their experiences given a certain reality. Examples of such experiences are presence and immersion in VE. Mahlke takes a different approach in explaining user experience (Mahlke 2005). He equates the user experience with the user interaction. Mahlke states that experience is a broad term that summarizes all users' aspects of interaction from their perspective. The interaction can be instrumental or non-instrumental (Mahlke 2005, Mahlke 2006). Instrumental interactions are measurable aspects such as usefulness or ease of use which often relates to usability. Interactions can also be non-instrumental with aspects such as hedonics, aesthetics, and emotions.

Mahlke analysis is similar to (Beuregard and Corriveau 2007) who warn not to confuse user experience with usability. They state that user experience definition could become a dichotomy which will confuse users due to the incompatible definition. On one end the term usability constructs (e.g. simplicity or task effectiveness) contradicts design constructs (e.g. fun, joy, aesthetic, or emotions). User experience goes beyond usability to include aesthetic, hedonic, contextual, and temporal variables. The authors provide a framework for user experience that contains psychological constructs. (McNamara and Kirakowski 2005) share their view that there is a need to change the conceptualization of usability from quality of use to quality of experience.

A different perspective framework is given in (Forlizzi and Battarbee 2004). Their framework is based on the user-product interaction, which they divide into three types: fluent, cognitive, and expressive. Those interactions in turn produce three types of experiences: experience (continuous), an experience (has beginning and end), and co-experience (creating meaning and emotions while using a product). Yet another type of experience is given by (Segerstahl and Oinas-Kukkonen 2007). They study the successful design of coherent user experience across a pervasive technological system. Through a multi device heart rate monitoring system, the authors describe the grounds of disturbed user experience. (Law et al. 2008) try to uniform the user experience definition by comparing definitions and types from different resources. (Law et al. 2009) extend this uniform definition of user experience by conducting a survey from researcher and practitioners from academia and industry.

In VR, components of the human experience have been studied by (Takatalo et al. 2008). Human experience is defined as the content of direct observation or participation in an event, and it is generated through the interaction of humans with the environment. The authors investigate patterns found in the experience of users of the VE. The experience is divided into certain components.

More specifically in haptic communication a study is conducted to see what users expect from their haptic experience (Heikkinen et al. 2009). The study is theoretical and provides a summary of the discussions held between the focus groups during the experimentation. A list of expectations and guidelines that will improve the experience of the user for haptic mobile communications is given. The user's expectations help in defining the user's actual experience and in some cases enhancing it as well. (Lin and Parker 2007) state that for investigating what enhances the user experience in VE, it is important to look at the positive aspect of VE (presence and enjoyment) and the negative aspects (simulation sickness).

2.2 Relation between QoS and QoE

This section describes the various relationships found in the literature between QoS and QoE. The paradigm shift from QoS to QoE was intended to include the "soft factors" which formerly were not regarded as important factors (Reiter 2009). These subjective or soft factors become necessary for multimodal applications that it is not sufficient to rely on the QoS metrics which worked well in the past for unimodal quality assessment.

In the seminar entitled 'From Quality of Service to Quality of Experience' (Fielder et al. 2009a, 2009b), researchers have discussed open questions regarding that community shift from QoS to QoE. The implication is that QoE augments the QoS and does not replace it. In essence, QoE is a substitute for end-to-end QoS, where the user is the end of the communication channel. QoE adds to the QoS purely technical parameters by focusing the shift towards the user and including, user satisfaction, ease of use, importance and contents of the service, pricing, among others.

Nonetheless, the relationship between QoE and QoS has been debated in research papers due to its importance. Several works are summarized below according to the type of the relationship.

2.2.1 QoS-QoE Correlation Model

In his holistic view of future interactive telecommunication services, the author stresses that QoS has lost its predominance (Reichl 2007). Instead, service design and end user perception are gaining prominence.

The quality of chain, as the author views it, is comprised of Quality of Design (QoD) plus QoS. The QoD defines the relation between the user and the device, while QoS defines the relation between the communicating devices. More specifically the relationship is given by

$$QoE(i, x) = QoD\ (i)\ o\ QoS\ (x)$$

where i = user , x= condition. This implies direct correlation between QoE and QoS.

Another correlation model is an ongoing effort by (Hyun et al. 2008, 2010a, 2010b). The authors suggest that QoE is a function of QoS parameters such that QoE = F(Delay, Jitter, Loss, Error rate, Bandwidth, Signal success rate). In any particular case, QoE would be the effect on the user perception given certain QoS parameters values. They tested their model using video on demand system and an IPTV system. They will be able to assign weights to QoS parameters and analyze hierarchical relation of QoS items and QoE.

(Gong et al. 2009) agree that QoS management scheme has little resemblance to the user's perception of the service. Instead of using regular QoS parameter mapping to QoE targets, they divide QoE into five parts based on the service integrality, retainability, availability, usability, and instantaneousness. Those five parts form a pentagon shape, and QoE would correlate to the size of the pentagon using mathematical modelling.

2.2.2 Mathematical Relationship

Some research efforts have tried to deduce particular mathematical relations between the QoS and QoE metrics. For instance, (Reichl et al. 2010) provide a logarithmic link between the physical stimuli and human perception. This can also be observed between QoS elements and the QoE on the user side, in a given communication system.

(Fiedler et al. 2010) suggest that generic QoS problems imply generic QoE problems. They derive the following mathematical formula:

$QoE = \Phi\ (I_1, I_2, \ldots, I_n)$ (meaning QoE is a function of many influences)

$QoE = f\ (QoS)$ (assume QoE is a function of one influence, the QoS)

$QoE = \alpha \cdot e^{-\beta \cdot QoS} + \gamma$ (QoE is an exponential function of QoS)

2.2.3 QoS as a building foundation for QoE

(Moller et al. 2009) describe a taxonomy of QoS and QoE composed of three layers:

1- The QoS influence factors related to the user, the system, and the context of use
2- The QoS interaction performance aspects describing the user and system behaviour and performance
3- The QoE aspects related to the quality perception and judgement process taking place inside the user

The two layers of QoS determine the system performance criteria but do not determine the user satisfaction, which is left to the QoE layer.

Authors (Bouch et al. 2001, Bouch and Sasse 1999) refer to the QoE as user-centred QoS. They argue that since user-centred QoS is gaining hold it can be assessed with a 3D approach. The 3D elements being:

- Measuring task performance

- User satisfaction

- User cost (in terms of physiological impact)

Essentially it would be possible to capture users' QoS requirements and fit it to the context of user's interaction. For example, the authors show that the quality perceived by users with respect to pricing, does not mirror the objective QoS at the network level.

Another idea is to map the many-to-one relationship between QoS metrics and QoE using a radar chart (Chang et al. 2010). The QoE function will be composed of an almost infinite number of QoS metrics $\{M_i | i=1, 2, 3, \ldots, n\}$ as in

$$QoE = f(M_1, M_2, M_3, \ldots, M_n)$$

As a proof of concept the authors measured QoE using PESQ (a standard to measure speech quality in telephony) with different QoS conditions (three QoS metrics were used).

2.2.4 User-level QoS

The authors (Ito and Tasaka 2005) describe the different levels of QoS while equating the user-level QoS, the highest level in the QoS chain, with QoE. They apply QoS mapping from application-level QoS to user-level QoS. They have used principal component analysis to determine which application-level QoS parameter to use. They have also used multiple regression analysis for the QoS mapping. To estimate the user-level QoS for conducting the analysis, the authors opted for using paired-comparison method rather than the mean opinion score (MOS).

On a relevant notation, (Siller and Woods 2003a, 2003b) have evaluated QoE using QoS metrics, network feedback, and dynamic user requirements. They state that by using QoS arbitration at the network level, QoE will be improved.

2.2.5 Other Relationship Attempts

(Du et al. 2009) relate QoE and QoS of a video application through neural network. They trained the neural network to produce QoE scores from QoS input. (Berado et al. 2008) evaluated QoE with different QoS mobility management strategies. Their results show that QoS-aware mechanism that reduces packet loss will improve QoE. Another work on mobility QoE (Verdejo et al. 2010) estimates the overall QoE of MMORPG virtual game on Android platform taking into account QoS-related parameters, player's environment context and physiological data.

Finally, (Khirman and Henriksen 2002) investigated the relationship between QoS and QoE for public Internet services. Basically, they measured the QoE based on the number of http cancellation requests (when a user reloads a page). The objective was to relate the objective network service condition with the human perception of the QoS. Their conclusion is that network delivery has less significant role on level of user satisfaction while speed (bandwidth) has a stronger effect on QoE.

Finally with QoE becoming more formal, it can be expressed in QoS terms using the following guidelines according to (Hestness et al. 2003)

`IF <communication situation>` (e.g. task, motive or user)

`USING <service prescription>` (e.g. telephony or video conference)

```
WITH <technical parameter>
```
(e.g. network delay or packet loss)

```
THEN <user behaviour>
```
(e.g. user satisfaction or communication efficiency)

2.2.6 QoS-QoE Relationship in Virtual Reality

Little has been done in this area in this extra investigation that we included since our focus is on VR. The present research usually discusses how to improve the QoS to achieve better experience in distributed virtual environments. For instance, (Marshall et al. 2008, Yap et al. 2010) discuss the network QoS effect on Distributed Haptic Virtual Environments (DHVE). They list the QoS factor and the corresponding side effect on the user of DHVE:

- Packet Delay: Degrades user's perception of collaboration
- Jitter: Instability of the system
- Loss: Inaccurate and abrupt movement of the haptic device
- Throughput: Abrupt movement

The network level QoS targets should be high enough to account for the sensitivity of DHVE compared to other applications. The authors specify preferred limits for the above factors which can be achieved under strict network conditions.

In (Matijasevic et al. 2000), the authors are equating QoE with QoS at user level in DHVE (similar to the user-level QoS section but specific to virtual environments). The highest level QoS parameters (user-level QoS) are: Interactivity, Immersion, and Plausibility. To achieve high level QoS, the authors suggest mapping the high level QoS parameters to low level QoS parameter.

Finally some work has started for QoS investigation for multiple force contact in DHVE (Abu-Tair et al. 2011), as opposed to the single point contact force most common in haptic applications. The authors suggest using CBWFQ scheduling technique which reduces delay of haptic traffic.

2.2.7 Summary of QoS-QoE Relationship

The literature presents several patterns for relating QoE parameters with QoS. Each pattern has certain advantages and disadvantages, which are summarised in the following table (only the main categories are presented in the table, the 'other' category and VR category are not included in the table. They are included in the previous discussion for inclusiveness).

TABLE 2.1- QOS-QOE RELATION COMPARISON

QoS-QoE Relationship Pattern	Advantages	Disadvantages
Correlation	QoE would be predictable based on the correlational model	Causation and Correlation are not interchangeable. This could result in erroneous QoE value, or there should be other factors involved.
Mathematical	Precise relation	Complicated, not always possible to determine an exact number
Building Foundation	QoE is built gradually according to the underneath layers.	The pattern focuses on relating the QoE construct to
User-level QoS	Simplifying the QoE mental model for the designers.	The abstraction could lead to undermine the QoE construct. In turn it could shift back the focus of the designers from QoE to QoS.

Although the previous references provide different descriptions of the relationship between QoS and QoE, they almost unanimously acknowledge that there is a relationship between QoS and QoE. This relationship is an important one since it will determine how to define QoE and improve it by also maintaining high QoS values. Our definition is provided in Chapter 3.

Based on the table, it seems that each category has its advantages and shortcomings. Calculating, estimating, or predicting the QoE from the QoS is very convenient for the developers or evaluators of the application but it deviates from the uniqueness of the QoE construct, which should not be based solely on QoS. This might defeat the purpose of having a separate construct. We acknowledge that QoS should be a part of QoE, as it affects the experience of the user. However, it should be one aspect of that construct among other aspects.

2.3 QoE in Virtual Reality and Related Multimedia

Big companies realized the importance of QoE since the beginning of the 90s. Their testing methods show that they were trying to include the user in various testing stages but nonetheless it was a non-comprehensive measurement (De Marez and De Moor 2007) such as allowing the user to test the prototype after it was built. The companies' interest in QoE is increasing, for example Nokia in its QoE white paper (Nokia 2006) stresses the importance of QoE in the mobile data services and say that the business implication of QoE is huge as it can make or break a company. The loyalty of customers will be based on high QoE with a particular operator. If the QoE is poor then the customer will simply switch to a different operator. Hence the operator should measure the QoE before it receives any complaints from the customer, otherwise the word of mouth will spread. The white paper goes on to describe a practical method for assessing the QoE of mobile data services. HP also constructed a framework to link the QoE with the Quality of Business (QoBiz) as they refer to it (Van Moorsel 2001). While discussing Internet services, they argue that QoE and QoBiz are interrelated and that QoE should be quantified and increased in order to acquire more customers and hence increasing the QoBiz.

In general, we are seeing increasing amount of work being done in multimedia applications and networks' QoE. Internet Protocol Television (IPTV) received special attention. IPTV is an application that makes use of next generation networks by sending IP packets to deliver digital television services. Slight packet loss over a network will result in a degraded video quality. For that purpose, (Asghar et al. 2009) discuss multi-layer approach in preserving the visual QoE, by monitoring the network and employing techniques to reduce packet loss and scrutinize video quality. (Wang et al. 2009) claim a new approach for measuring video quality in IPTV application by combining the assessment of both video content and the underlying network at the same time, while (Takahashi et al. 2008) explain the techniques of objective quality assessment for subjective entities such as audio and video in IPTV by utilizing the network physical properties. Accordingly, in (Greengrass et al. 2009a, 2009b) the authors discuss in detail the effect of network factors on the QoE of IPTV application. The reason researchers focus on IPTV is because it contains audio, video, and most importantly network components and their way of integration will affect the QoE of the user. Unlike VR applications, however, IPTV is not task oriented and does not contain constituents such as haptics, 3D graphics, and special stereo displays and glasses.

In virtual reality in particular, QoE was less investigated than networking and IPTV. (Ebrahimi 2009, Rowe and Jain 2005) stress the move from QoS metric to QoE for newer multimedia data types and human computer interaction, which could be the future of interaction. Measuring the QoE requires measuring the user's perception of the VR application. As (Jain 2004) puts it, we require improved performance measures over the well-established QoE measures to deal with the subjectivity of the user. (Whalen et al. 2007) discuss some of the methods and challenges in determining those performance measures in the context of VR.

There are three recognized methods for assessing the user's feedback and responses in a VE: subjective ones, performance-based, and physiological methods (Whalen et al. 2003). Each method enables the collection of a specific type of information regarding the user's responses to the application. For instance, subjective measures evaluate the user's satisfaction, fatigue, intuitiveness, preferences, etc. (collected via surveys). Performance QoE approach measures the user's behaviour when performing a task with the VR application. Finally, the physiological QoE measures non-voluntary responses of the human body during and immediately after the test session.

One research group (Iwata et al. 2010, Tatematsu et al. 2010, Kusunose et al. 2010, Watanabe et al. 2010, Ida et al. 2010) rely on subjective measures solely to assess the QoE of various applications. They use Mean Opinion Score (MOS) to determine the overall QoE

(MOS will be described in a later chapter). Their main idea is to test network jitter and delay on haptic quality reflected on the user. Using different settings and jitter parameter they test different scenarios using a networked haptic application. In one scenario they test a hockey game, while in another they test a networked writing application. One of their work assesses haptic and video and audio, another focuses on haptic, while one of them focus on the quality of a first person shooter game. From the MOS values, the research group use multiple regression analysis to link QoE parameters with application-level parameters.

Subjective and performance-based QoE evaluation research has been performed in the haptic field. Nonetheless, the evaluation methods and the aspects to be evaluated vary depending on the type of the application and the parameters to be evaluated. For instance, (Basdogan et al. 2000) combined both types in their studies to evaluate the incorporation of the haptic media in collaborative human-human and human-machine interactions in shared virtual environments (SVEs). The evaluation consisted of measurement of response variables as well as questionnaire to the users undergoing the experiment. A similar approach that measures haptic benefits in SVEs is presented in (Guerraz et al. 2006). The authors measure physical parameters generated by the haptic device directly in order to assess the quality of the application. The authors suggest that this is an assistive approach to conducting a statistical survey. Examples of parameters that are measured and included in the physical survey are gesture position and gesture velocity.

Performance-based QoE can be often interchanged with Quality of Performance (QoPer) which is another user-centric quality metric (Roid 2004). QoPer methods refer to observational procedures or testing, that are designed to evaluate the correctness of performing a particular task – or how well the user performs. Compared to QoPer approach that is task and procedure oriented, QoE approach measures the implications of the interaction at the cognitive level.

QoPer measures have been used in virtual environments research. For instance, a quality of performance model is proposed in (Alamri et al. 2008) to evaluate a stroke patient performance in a rehabilitation system. The authors proposed an evaluation taxonomy that includes the following factors: task completion time, eye-hand coordination, compactness of task, hand movement and steadiness, grasping angles and fingers grip accelerations. Another work (Kyoung and Kenyon 1999) studied the effect of network latency and jitter on performance in a collaborative virtual environment, where two subjects manipulate a ring along a complex trajectory. The performance metrics include the subjects' speed (time to complete the task) and error rate (number of collisions between the ring and the path).

Few researchers have investigated the physiological QoE approach. (Whalen et al. 2003) suggested the use of physiological measures to determine the QoE of virtual reality applications. With stress as an example, the authors argued that the sympathetic nervous system is activated and blood volume, heart rate, and respiration rate all increase. By measuring these parameters, an estimation of the stress level can be made. When dealing with cybersickeness resulting from VR environments, Ramsey (1997) claims that measuring those symptoms directly (through physiological parameters) is more effective than a questionnaire due to three limitations:

1. People are mentally aware of their internal state (emotional condition) when under the same circumstances, in the real world, they would normally not be

2. People might not understand the implication of the response in the questionnaire

3. People may not wish to report feeling any symptoms of sickness

In their research paper "Quality of Experience in Virtual Environment", (Gaggioli et al. 2003) look at experience from a holistic view. They first argue that some authors equate virtual experience with presence which is biased since many other elements are included in the human experience such as emotions and enjoyment. They define what is called optimal experience that is a state of consciousness having positive, complex, and rewarding properties. This optimal experience is combination of genetics and innate human qualities. VR has the four characteristics necessary for creating optimal experience: Opportunities for action, skills, feedback, and control.

2.4 Conclusion

Certain experience aspects can be assessed separately such as presence. However in different scenarios QoE should be considered as a whole. Up until now there is no genuine progress towards evaluating QoE of VEs, specifically when haptic devices are utilized. Here is where this thesis steps in. It puts QoE of VE into the kaleidoscope, and focuses on the users and their interaction with the environment. The next chapter describes the taxonomy that classifies the parameters involved in evaluating the QoE of a VE.

CHAPTER 3

QUALITY OF EXPERIENCE TAXONOMY FOR HAPTIC VIRTUAL ENVIRONMENTS

In this chapter we describe the QoE definition and taxonomy we used to organize the different parameters. From Sections 2.1 and 2.2, we wanted a definition that would agree with the information presented in the literature regarding the properties of the QoE metric. The definition we have adopted is that the QoE is composed of the QoS and User Experience (UX). This definition would satisfy the three points mentioned in Section 2.1 which are the QoE is user centric, subjective, and revolving around the experience of the user. Concurrently, the definition of QoE includes the QoS metric, in which the relation between the two was examined in Section 2.2. As mentioned in that section, QoE is not based only on QoS metric but rather an amalgamation of QoS and UX.

Our initial taxonomy was based on subjective vs. objective metrics (Hamam et al. 2008a). As such, we divided the parameters into two groups: ones that can be measured directly from the application such as forces and delay, and the other group that has to be deduced by other means such as user questionnaire and behavior; for instance intuition. However we felt that in such a model the taxonomy should stem from the core definition of QoE. Hence, we based our top organization into two parts: QoS and UX parameters. Surveying the literature and exploring what UX is composed of, we further subdivided this category into four parts: Perception Measures, Rendering Quality, Physiological Measures, and Psychological Measures. This higher level organization, shown in Figure 3.1, reflects an apparent taxonomy for VR applications evaluation and at the same time is more customizable depending on the parameters needed for evaluation. As an example, developers wishing to evaluate only the QoS of the application can disregard the UX parameters.

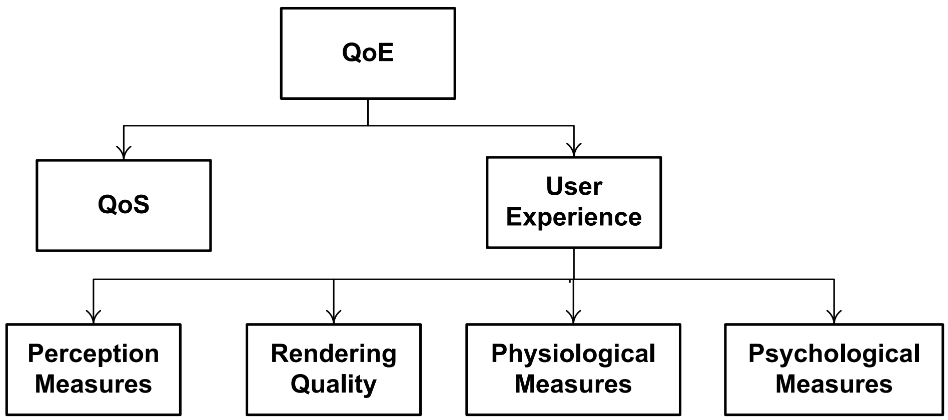

Figure 3.1. Higher level organization of QoE model.

3.1 Quality of Service Parameters

QoS parameters ensure the smooth flow of the application for the user or in certain cases the customer. Most parameters are standard for any networked application but looking at Table 3.1 we can notice that synchronization is divided into two parts: network synchronization which is common to network applications and media synchronization which is specific to the multimodal side of virtual environments, in essence the synchronization of the three media streams: graphics, audio, and haptics.

TABLE 3.1- QUALITY OF SERVICE PARAMETERS

Response Time	The time taken by a system to respond to an action, measured in millisecond or

	microsecond
Latency/Delay	Time taken for the packet to reach from source to destination, measured in millisecond or microsecond. There are different source of delay. These are 1) Propagation Delay which is the delay through a physical medium, 2) Link Speed, which determined by link bit rate, 3) Queuing delay which represent the time spent in router queues and 4) Hop Count where each traversed router or switch adds queuing delay
Price	The quantity of payment or compensation given from one party to another in return for goods or services, it can be measured by a metric related to energy, monetary, automation or other efficiency of the service.
Privacy	Deals with what personal information can be shared with whom and whether messages can be exchanged without anyone else seeing them
Security	The level of protecting the information exchanged through the use of multimedia technology
Availability	The ratio (or probability) of time a system or component is functional to the total time it is required or expected to function. Small probability values for availability indicate bad QoS, while high values indicate good QoS.
Bandwidth/Throughput	The amount of data transferred from source to destination or processed in a given amount of time. Measured typically in bits/second or bytes/second
Synchronization : Network Synchronization (CVE) Media Synchronization (intra-modal)	Network Synchronization: the temporal relations linking the various media objects within a multimedia presentation. Example: Time relations of a multimedia synchronization that starts with an audio/video sequence, followed by several pictures and an animation that is commented by an audio sequence and haptics feeling. Media Synchronization (intra-modal): Refers to the temporal relations between media units within a time-dependent media object. For a video with a rate of 25 frames per second

	each of the frames has to be displayed for 40 msec. For haptic data with 1 KHz, each of the data samples must be captured and displayed for 1 msec
Jitter	Difference in latency of network packets usually measured in microseconds or nanoseconds
Reliability	The ability of the computer system and its components, i.e., haptic audio visual environment to consistently perform according to the given specifications
Error Magnitude Frequency	Sometimes CHAVE packets are corrupted due to bit errors caused by noise and interference. The receiver has to detect this and, in case the data contained in the packet is needed, may ask for this information to be retransmitted.
Safety	The needed aspects to be considered in order to operate the haptics environment properly and use it in conjunction with other peripheral equipment without damaging the environment and the users

3.2 User Experience

Section 2.1.3 provided a holistic definition of the user experience (UX). Moreover, some authors studied how this UX can be measured. In their book "Measuring the User Experience", (Tullis and Albert 2008) discuss the proper testing environment and the proper metrics that should be selected in order to quantify the user experience and thus compare one ecommerce project to another. Although their work focuses on commercial websites and applications, it could be applicable to other domains. Their work tends to be geared more towards usability. (Zhou 2007) claims measuring and quantifying the user experience but the focus was on using fuzzy derived model to measure usability as well.

(Lin and parker 2007) take a step further than usability by the assessment of the user experience in VE. They examine positive aspects (presence and enjoyment) of a VE that employs a wide-field display and negative aspects (simulation sickness). Their results suggest that a large field of view would increase presence and enjoyment but would also increase simulation sickness.

We break down the categories of the UX into four parts in the second part that constitutes the definition of QoE which is the User Experience. This is an important evaluation category for the overall quality of the application. Even if the application possessed excellent QoS parameters still users might feel that the application is not up to their standards for some reason. The application might not be exciting enough, difficult to use, or causes dizziness which is referred to as cybersickness.

We begin by describing the first construct of the UX according to our classification.

3.2.1 Perception Measures

As depicted in Figure 3.2, perception measures mirror how the user perceives the application. This is a user-centric category, and could be unique for every user. Some users may get tired from the application, while others may feel relaxed. Some might feel the effect of collaboration in a Collaborative Haptic Audio Visual Environments (CHAVE) while others might need more stimuli. Each user may have a certain set of preferences and modality choice.

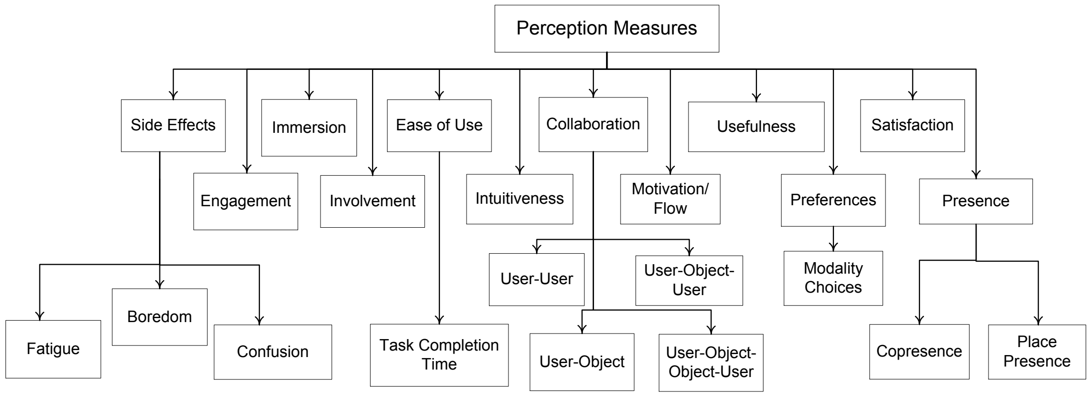

Figure 3.2. Perception measures parameters.

Furthermore, there are different levels of experience among users. While a certain group of users could be very experienced with virtual reality applications and very dexterous using haptic devices, others may be novice users and less skillful. This variation in the level of experience will cause users to have different perception regarding the application. When evaluating a HAVE application it is essential to include different categories of users and to ensure that the application suits a wide range of audience.

The taxonomy of the perception measures element is deduced mainly from the literature. There are three pillars of user perception of an interactive environment in the literature. These three pillars form the basis of Figure 3.2 along with some modifications and additions for the completion of the taxonomy. The three pillars are described next.

3.2.1.1 Presence

Presence is a term usually associated with virtual reality environments to indicate its effectiveness. (Witmer and Singer 1998) define presence as the "subjective experience of being in one place or environment even when one is physically situated in another". Involvement and immersion contributes to presence. The authors state that the strength of the presence experienced in a VE varies both as a function of individual differences and the characteristics of VE. They have established the presence questionnaire, a 7-point Likert scale, which measures the degree to which individuals experience presence. The questionnaire is constructed from presence factors appearing in various literature materials and tested for validity and reliability. The two most significant factors within the questionnaire are involvement and immersion to be discussed in the next section.

(Schubert et al. 2001) try to extend the above questionnaire by exploring components within the presence construct in oppose to exploring presence factors. They explain presence through cognitive processes: construction of a mental model and attention allocation. They constructed a presence questionnaire where most questions are not part of the Witmer and Singer's questionnaire.

Presence can also refer to the sense of togetherness in a VE, which is the feeling that there are other users in the environment. This experience is sometimes phrased as co-presence or social presence and has been addressed in the literature (Durlach and Slater 2000). (Basdogan et al. 2000) measured the user performance while playing a virtual ring on a wire game with an unknown partner. The follow up questionnaire revealed that users do experience co-presence more strongly when haptic feedback is utilized instead of just relying on vision.

(Sallnas et al. 2000) also conducted an experiment for collaborative virtual task to explore effect of haptics on social presence. The tasks consisted of rearranging virtual cubes collaboratively using a haptic device. One setting of the experiment consisted of video/only setup, while the other consisted of video/audio/haptic. The hardware apparatus remained constant throughout the two settings. The haptic feedback effect did improve the following variables: task performance, perceived task performance, perceived virtual presence, and perceived social presence.

The aforementioned strategies use subjective strategies to measure presence. (Nichols et al. 2000) state that there is no universal methodology to measure presence like any other mental model. (Insko 2003) lists three ways to measure presence: subjective, behavioral and physiological methods.

A survey of most empirical and subjective studies conducted to evaluate presence in VR is given by (Schuemie et al. 2001).

3.2.1.2 Immersion/Involvement

There has been a disagreement of what defines immersion, in the sense whether it is a subjective or an objective construct. (Witmer and Singer 1998) view immersion as a "psychological state characterized by perceiving oneself to be enveloped by, included in, and interacting with an environment". They introduce the immersive tendency questionnaire which measures the capability or the tendency of individuals to be immersed in an environment. (Slater 1999) on the other hand, views immersion as an objective characteristic of an environment. It is the extent to which a system provides a surrounding environment (that is why we have included deg. of immersion in psychological measures as it will be seen later).

(Pausch et al. 1997) compare experimental results between desktop-like configuration and a head-tracking VR configuration. The latter has immersive properties, and the task was to locate certain objects on the display. They have found that immersion reduced the search time of the users. (Slater et al. 1996) conducted similar experimentation where they used chess board reconstruction as a task for the users and compared immersive and less immersive hardware setups.

Immersion has been observed in gaming. (Jennett et al. 2008) conducted a questionnaire to measure immersion in video games. They also took into consideration objective measures to further support the questionnaire: time required for a physical task (user immersed in the game will take longer for the same task after playing the game), eye tracking, and speed (the more immersive the game is the quicker it will be played). (Brown and Cairns 2004) identify three levels of immersion in games: engagement, engrossment, and total immersion. They link the level of immersion to the hardware used and also to the level of involvement of the player. (Witmer and Singer 1998) define involvement as a "psychological state experienced as a consequence of focusing one's energy and attention on a coherent set of stimuli".

3.2.1.3 Engagement/Enjoyment/Flow

(O'Brien and Maclean 2009) define engagement as "the act of emotionally involving someone or the personal state of being in gear" which could be an outcome of experience or a process during an interaction. They state that engagement is an important component of user's experience but it is difficult to measure. In order to capture experience they suggest employing mixed methodologies to capture cognitive, affective, and behavioural components of experience.

Users' enjoyment has been studied mostly in computer games, since it is usually the ultimate goal of a game. (Ijsselsteijn et al. 2007) state that no two players experience the same level of enjoyment while playing a game. (Sweetser and Wyeth 2005) model enjoyment through flow. Flow includes eight elements: concentration, challenge, skills, control, clear goals, feedback, immersion, and social interaction. (Fu et al. 2008) followed Sweetser and Wyeth’s model to establish a valid scale to measure the level of learner's enjoyment while using e-learning games. The scale was constituted of the eight previous categories although tweaked to fit the context. The factors were then related to the level of enjoyment of four e-learning games.

3.2.2 Rendering Quality

The rendering quality relates to the quality of the three major modalities, namely: graphics, audio, and haptics. Each modality is considered separately first and eventually blended and mixed modalities are considered. As seen in Figure 3.3, there is an emphasis on haptics modality since it has very stringent requirements in terms of feedback loops which might affect the stability and transparency of the application. The rendering quality category highlights how the users experience those different modalities and their interaction as they traverse the VE. The parameters are gathered from VR and haptic books (Burdea and Coiffet 2003, El Saddik et al. 2011) as well as haptic papers such as (Srinivasan and

Basdogan 1997), some parameters were also obtained from the human factors analysis of virtual environments under the definition of QoE (Section 2.1.1).

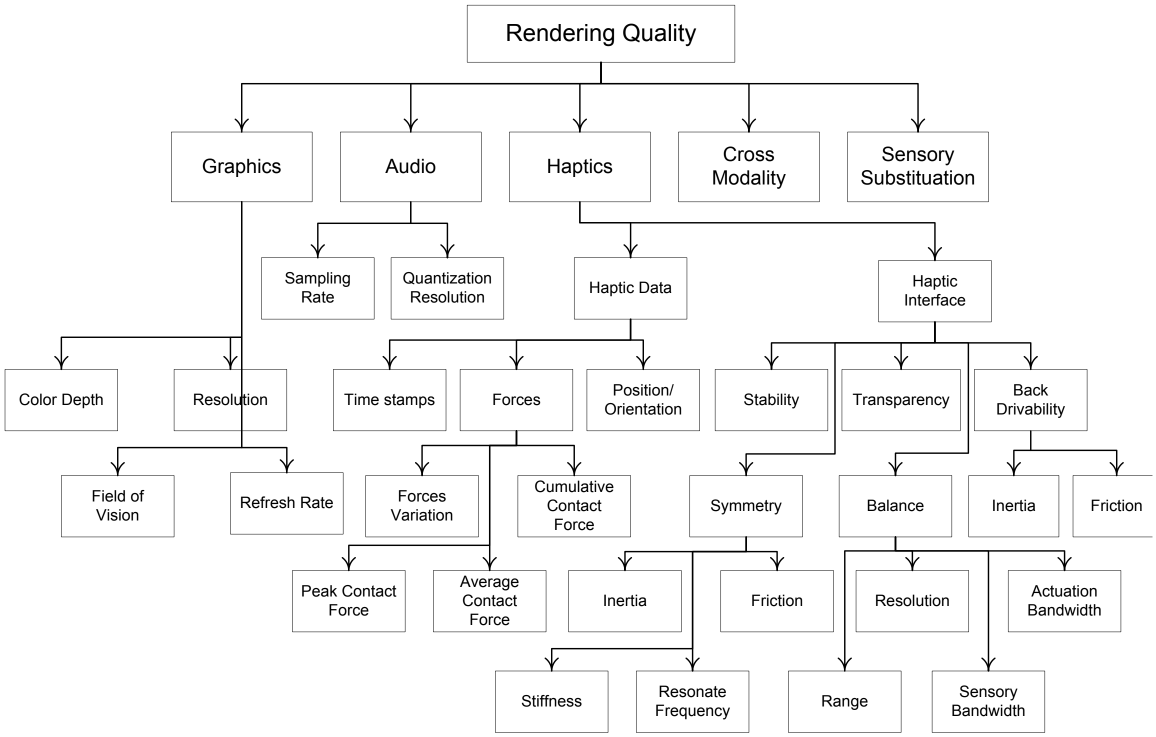

Figure 3.3. Rendering quality parameters.

3.2.3 Physiological Measure

The user state parameters can be divided into two complimentary sets. The first of these sets indicates the state of the user through biological means. Physiological measures are biological parameters measured directly from the users' bodies while they are using the application. These parameters indicate relevant factors such as anxiety, stress, and brain activity (Figure 3.4).

There is a continuing research effort in psychophysiology that tends to evaluate elements of the user experience by subjective and objective measures. The research focuses on the users' emotions during the interaction with an environment such as playing of a video game. The subjective measures would be the user emotions assessed via questionnaires while the objective measures are the corresponding physiological measurements of the user.

For example (Nacke and Lindley 2008) establishes the correlation value between the subjective and objective indicator of the experience during the gameplay of a first-person shooter game. This correlation links which physiological measurement affects which psychological measure the most.

Cybersickness falls into this category because it is generated from the interaction with the VE and it has similar symptoms as motion sickness (eye strain, disorientation, postural instability, sweating, nausea, drowsiness… etc.). Research on what causes cybersickness and ways to reduce it is being conducted in the VR community (Burdea and Coiffet 2003).

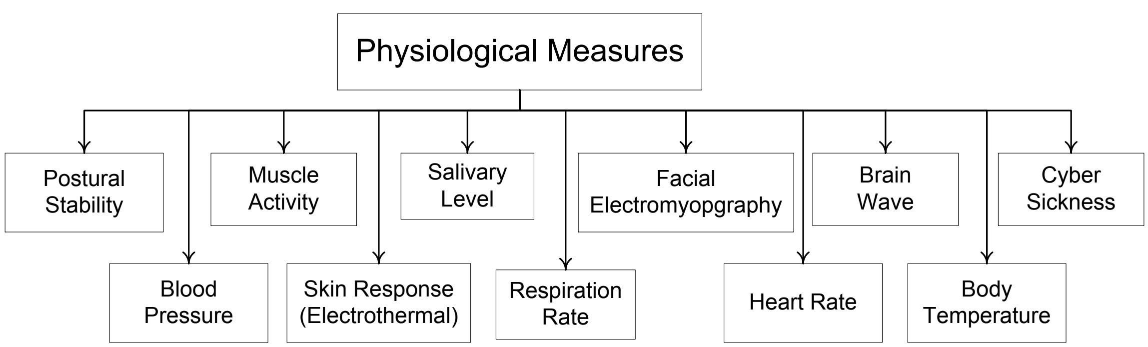

Figure 3.4. Physiological measures parameters.

3.2.4 Psychological Measures

Unlike the physiological measures, psychological measures reflect the state of the user through observation and user feedback but not direct measurements. Observation can assess the psychological behavior of users, such as stress, without hindering the user's movements by including measuring devices. Psychological measures are displayed in Figure 3.5.

In the figure, emotions are subdivided into two parts negative emotions and positive emotions. These terms are relative to the user and not the experience. Negative emotions can produce undesired responses if not desired (if the premise of the application did not intend to produce these emotions). Even users' positive emotions are capable of producing undesirable effects but to a lesser degree. For example, there is the difficulty of re-entering into the real world after a pleasant experience in the virtual world. This turbulence of emotions may decrease the user's experience if not handled gracefully (Behr et al. 2005).

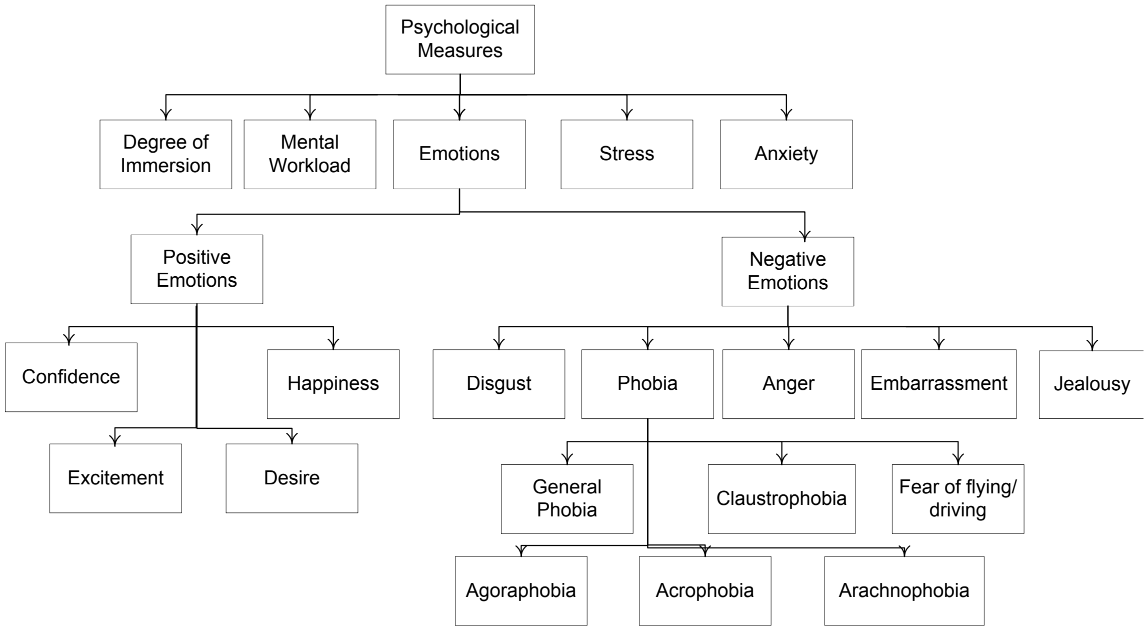

Figure 3.5. Psychological measures parameters.

For definition of the phobia terms please refer to (Strickland et al. 1997). Strickland et al. discuss VR applications that can cure individuals suffering from phobia. Nonetheless, if the application's goal is not to address phobia, any fear influence will decrease the QoE considerably.

The degree of immersion was added in this category because it is a psychological factor based on the perception of immersion. Since immersion was defined by some researchers to be a property of the system itself, then the degree of immersion would be the psychological effect of that system.

In the future it might be necessary to update the user state to include perception measures as well. The psychophysiology field is advancing and researchers such as (Nacke and Lindley

2008) are linking some perception elements with the user state through biological means (physiological parameters).

One example links biological measures and presence. As mentioned, physiological measures have been utilized to indicate the level of presence in stressful virtual environments (Meehan et al. 2002). The authors have measured users' heart rate, skin conductance, and skin temperature. They hypothesised that a VE would invoke similar physiological responses in similar real life situation (if presence was high in the VE). Heart rate indicator has best performed in terms of reliability, validity, sensitivity, and objectivity.

3.3 Applying Taxonomy to User Study

In this section, we apply the taxonomy adopted into a user study. The study involves the evaluation of five parameters from the taxonomy that are related to a VR haptic game. We start by describing the protocol that we have used for conducting the user experimentation.

3.3.1 Methodologies for User Testing

The testing methodology for human factors studies has been reported in (Burdea and Coiffet 2003) and (Tullis and Albert 2008). The main methodology for designing experimentation with human subjects can be divided into the following;

Selecting participants: Certain questions may be raised depending on the experiment at hand. Are we going to divide the users into groups? Common group categories are male/female, novice/intermediate/experts, and young/old. In all cases, the participants should represent the larger population, i.e. the targeted audience of the application. In that regard, the gender and age of the subjects are determined by the study's objectives. In our case we strived to recruit as diverse test participants as possible. We did not intend to group the recruits but in order to eliminate bias, we mixed the male/female ratio and tried to maintain balance between the new haptic users and expert haptic users. However, since our experiments were based mostly at our research lab in the university, we had difficulty varying the age range. Nonetheless, we were able to get a decent age range (18-35) since we recruited both undergraduates and graduate student (from different faculties). Hence, our participants represented the population as much as possible. Still, age limitation in our case (senior citizens) should be looked at in the future, if haptic devices are to be deployed for mass population.

Recruitment of subjects can be done through advertisement (web, email, posters, etc…). Subjects are screened to maintain the suitability of the study. (Tullis and Albert 2008) note that sample of convenience is a very common way of recruitment (that is inviting whoever is willing to participate to the experiment), however, it should be free from bias (or bias should be accounted for).

Sample Size: This depends on how much confidence it is expected from the data. This also depends on the experiment protocol details: Exact number of trials per session, number of sessions performed by the participants in a day, total number of days the study will last. The number of participants at the end of the day would depend on the scope and the budget available.

We conducted experiments for high number of users. Thirty people participated in the experiment, since we wanted high confidence rate.

Within subjects or between subjects study: Each type of these studies has its advantages and disadvantages. Within-subject refers to comparing results of different data sets of the same participant, while between-subject refers to comparing data sets of different groups typically each group is performing the experiment under different controlled conditions.

The study that we conducted is a within-subject study as we did not divide the users into different groups. Our goal is to see the variation of QoE for the whole population, not certain group of the population.

Counterbalancing: Changing the order of the tasks if possible to minimize learning effect within the experiment. This part is not always possible (as in our case) when the order of the tasks is significant.

Independent and Dependent Variable: determining what are the manipulated variables and what are the response variables to be measured. Our dependent variable was the QoE value of the application; this is the parameter we have attempted to measure. The independent variables are the different parameters of the application that affect the users' decision.

Consent: the subjects consent should be considered when performing the experiment and if they are filling out a questionnaire afterword. We took the consent from all the participants.

3.3.2 Application Description/Experimental Setup

The application we used to test the proposed model is the Balance Ball game (Al Osman et al. 2008), shown in Figure 3.6. A ball is placed on a long wooden board that is held by two players from each side. The game involves the two users collaborating in maintaining the balance of a virtual ball on a board using remote haptic devices. Each player holds one end of the board with his/her haptic device and raises it slowly over a virtual pole to predefined end mark. The challenge is to collaborate in an attempt to keep the board horizontally balanced as much as possible from the initial location to the destination. Any variation in the horizontal balance will cause the ball to roll away towards one side thus penalizing both players. The players should remedy that by using the force feedback and the 3D graphics to apply their judgment in balancing the board again. The score consists of the task completion time and the variations of the ball's position from the middle of the board.

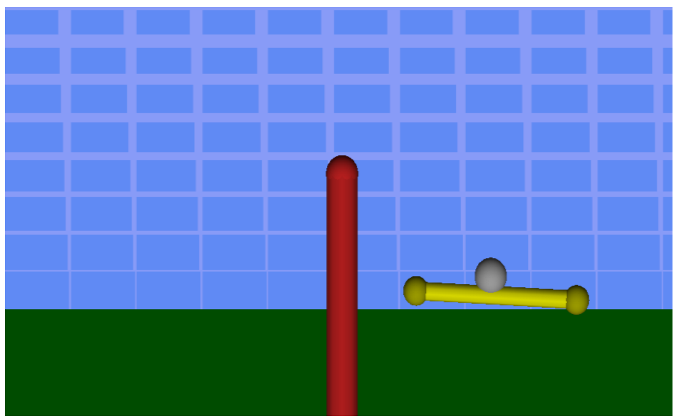

Figure 3.6. The "Balance Ball" game snapshot.

The experiment took place at the haptic laboratory of the DISCOVER Lab at the University of Ottawa. Thirty users participated in the experiment. The collaborative application ran on two computers. The computers were running WinXP SP3 on a 2x Intel Xeon 2.8GHz with 2GB of RAM and an Nvidia QuadroFX 2000XGL 128Mb DDR video card. A Phantom Desktop haptic device was attached to it. The Phantom Desktop is six degrees of freedom (DOF) positioning and sensing haptic device developed and marketed by SensAble Technologies, Inc. It has a compact design and provides three DOF output capabilities. A snapshot of the experiment setup is shown in Figure 3.7.

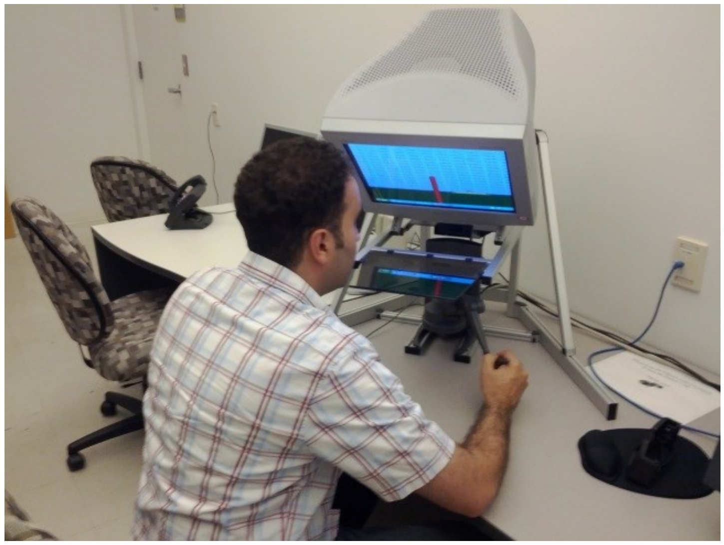

Figure 3.7. The experimental setup of the user study conducted. The figure displays the station setup of one of the users during their gameplay.

The experiment was conducted on an Ethernet Local Area Network with ALPHAN over UDP as the transport protocol. Network disturbances such as delay and jitter were controlled using a software tool we developed for this experiment. In order to make use of the jitter smoothing algorithm, the clocks of both workstations were synchronized using Network Time Protocol server. Both workstations maintained a connection with the server with clock synchronization precision falling within one millisecond.

Users were selected randomly at the University of Ottawa according to the protocol described in Section 3.3.1. Some users had previous haptic experience, while other users were new to the haptic notion. In either case, the user was given a general background about the application, how to handle and hold the haptic device, and the goals of the experiment. Users were reminded that the purpose of the experiment is to evaluate the application and not the users' abilities.

Users were divided into teams. Each team consisted of two users, and the experiment lasted around fifteen minutes, which included playing the game twice (the first time was a trial game, while the second one was the actual game). After the team finished playing the virtual game collaboratively by reaching their destination in the actual game, they were asked to fill out a questionnaire with general questions about the virtual game, past haptic experience, and specific questions that reflect elements of their experience. Each user filled a separate questionnaire.

3.3.2.1 Results

The result of the questionnaire is presented in Table 3.2. The questionnaire presented to the users was a five point Likert scale in which each question is followed by two anchor labels with five points to choose in between. Essentially, most questions are followed by a five point scale in which the users are required to circle the point that is closest to their level of agreement. Each extreme poles of the scale are marked by opposing descriptive labels based on the content of the questions. The descriptive labels help the user reflect on the question as he/she is completing the questionnaire. A sample question presented to the user was the following:

To what extent, if any, did using the haptic device cause fatigue?

Not at all   Completely

1   2   3   4   5

There was mostly one-to-one mapping between the user preference and the evaluation assigned to the parameter, such that one question corresponded to the evaluation of the parameter by the user. The focus was on the preference of the user, while the performance was used in another research area (Al Osman et al. 2008). Our goal here was to quantify the subjective evaluation of the user without including the performance metrics.

The users rated the overall QoE as a percentage in order to get a more precise value for several reasons. The value is important as it is the ultimate goal of the evaluation. A precise value would ease the validation of the model and enhance error calculations. Hence given the users' percentage rating of the application, we will have two QoE values to compare: one from the user and another from an evaluation engine described in this thesis (mathematical model or fuzzy logic system).

TABLE 3.2- QUESTIONNAIRE RESULTS FOR BALANCE BALL GAME

Subject	Media Synch	Fatigue	Haptic Rendering	Deg of Immersion	User Intuitiveness	Overall Rating
1	4	3	4	5	5	85
2	4	1	4	5	5	90
3	4	2	4	5	4	90
4	5	1	4	4	5	95
5	5	1	5	3	3	90
6	5	1	3	4	4	90
7	4	2	5	4	2	92
8	2	4	2	4	3	80
9	4	3	4	4	5	90
10	5	1	5	4	5	90
11	2	2	5	4	5	90
12	4	1	5	4	5	95
13	5	1	4	4	4	90
14	4	1	4	3	4	90
15	4	4	3	3	4	75
16	5	4	4	2	4	80
17	4	2	5	4	5	90
18	4	2	4	3	4	85
19	5	1	5	5	5	100
20	2	4	2	2	2	40
21	4	1	2	5	4	70
22	4	1	3	4	3	90
23	3	1	4	4	5	97
24	3	2	3	3	3	65
2	4	3	4	4	5	80
26	4	4	4	5	5	60
27	3	1	4	4	3	70
28	3	3	4	4	3	80
29	4	1	5	5	3	90
30	5	1	5	5	5	95

The results are summarized in the following table:

TABLE 3.3- SUMMARY OF THE QUESTIONNAIRE RESULTS

Parameter[a]	Mean	Standard Deviation
Media Synchronization	3.93	0.89
Fatigue	1.97	1.14
Haptic Rendering	3.97	0.91
Degree of Immersion	3.97	0.84
User Intuitivenss	4.07	0.96
Overall Rating	84.13	12.51

a. All the parameter ratings are out of 5 except for the Overall Rating which is a percentage

3.3.3 Selecting Parameters

Five parameters from the taxonomy were selected that are relevant to any haptically rendered three-dimensional collaborative game application such as the Balance Ball game that we tested. They are listed below along with a description of each parameter as well as the reason it was selected. The category from which the parameter is selected is written in parenthesis. If the parameter is not from the QoS category, then it will be from a sub-category of the UX.

1. Media synchronization (QoS parameter): There are usually three media modals in a HAVE application. Any miss-synchronization between the audio, video, and haptics can cause a drastic loss of perception of both media that are miss-synchronized. Therefore media synchronization is necessary for players to maximize their perception and enjoy the game. In this particular case, we focus on the subjective aspect of media synchronization

from the user's point of view (even though it can be analyzed through equations our focus is the user perspective and experience).

2. Fatigue (perception measures parameter): Research has shown that fatigue, which is caused by muscle exhaustion, is linearly distributed as a function of time (Seroussi et al. 1989). Fatigue is a crucial parameter because the haptic application needs users to interact with the virtual environments by exerting force and it induces fatigue easily compared to audio-visual feedback. Depending on the specifics of an application and on the haptic device used, rapid fatigue can hinder users and limit their rapport with the application. On the other hand if the application minimized users' fatigue then their experience will be more positive.

3. Haptic rendering (rendering quality parameter): Haptic rendering quality remains the same until we reach a threshold (that is usually referred to as the JND - Just Noticeable Difference) after which the quality starts decaying (Srinivasan and Basdogan 1997). For any haptic application, we want the quality to remain above that threshold, otherwise any instability, low resolution, or low haptic fidelity will render the application virtually unrealistic from the user's point of view.

4. Degree of immersion (psychological measures parameter): Even though the degree of immersion will cause a difference in quality, this difference is still not quite understood (Gutierrez et al. 2007). However immersion in gaming application is of importance, since the more the users are immersed in the game the more they are involved and experiencing enjoyment (Ijsselsteijn et al. 2007).

5. User intuitiveness (perception measures parameter): User intuitiveness is an important phenomenon that has been considered in disciplines other than human-computer interaction, such as nursing (Miller 1993). Although the factors that contribute to intuitiveness are little known, it can be observed through swift and determined actions of the user. It can be determined through user feedback as well.

The selection of the parameters is tailored towards the new medium experience and game experience holistic theory. The new medium (haptic) is emphasized by the two parameters: media (graphic and haptic) synchronization and the haptic rendering parameter that is the top level parameter of the haptic node under the rendering quality measures in our taxonomy.

Game experience psychologists have divided the experience for users of digital games into two categories; immersion and flow (IJsselsteijn et al. 2007). Under flow, there are certain characteristics that model an acceptable level of enjoyment for the user. Most importantly, the interface should not be too cumbersome and it should be responsive to the user (intuitiveness). In addition, the lack of fatigue will increase the flow and enjoyment level among users (Sweetser and Wyeth 2005).

The parameters selected represent all the categories in the taxonomy (considering that psychological and physiological measures both represent the user state). This was appropriate since we wanted to do an overall QoE evaluation. Moreover, since the parameters selected stems from game theory, they are relevant to any haptically rendered three-dimensional collaborative game application such as the Balance Ball game that we tested.

3.4 Conclusion

This chapter presents a taxonomy for classifying QoE parameters for VE. In addition, it shows a user study in which this taxonomy was applied to a user evaluation of a haptic multimedia game. This user study will be utilized in the next two chapters to evaluate the QoE of the haptic game. In the next chapter, we will employ a mathematical model evaluation paradigm to compute the QoE of the application. In the chapter that follows, a fuzzy inference system is used to compute the QoE. In both cases, the QoE value that results from each evaluation paradigm is compared to the QoE value given by the users in the case study and analyzed.

CHAPTER 4

QOE EVALUATION THROUGH MATHEMATICAL MODELING

In this chapter we detail one of two QoE evaluation paradigms adopted: mathematical modeling via weighted averages. The other evaluation paradigm, fuzzy logic inference system (FIS), is described in the next chapter. For each chapter we used the results presented in Chapter 3, where we described the steps used for performing the user tests including the methodology, the application's description, and the way the parameters are selected.

4.1 Mathematical Model

The evaluation of the QoE metric can be done in two ways. The first type of evaluation, discussed in this chapter, is a mathematical model where the QoE is computed as the weighed linear combination of the QoS and UX for a particular haptic user interface. In turn the QoS is computed as a weighted linear combination of the parameters in the QoS category. For the UX, each sub-category is adjusted by weights of its own. Moreover, the sub-categories are treated as a weighted average of their own parameters (each sub-category is treated like the QoS category). The mathematical model equations are as follows:

$$QoE = \zeta \times QoS + (1-\zeta) \times UX \qquad (4.1)$$

Where

$$QoS = \frac{\sum_l \eta_l S_l}{\sum_l \eta_l} \qquad (4.2)$$

and

$$UX = A\frac{\sum_i \alpha_i P_i}{\sum_i \alpha_i} + B\frac{\sum_j \beta_j R_j}{\sum_j \beta_j} + C\frac{\sum_k \gamma_k U_k}{\sum_k \gamma_k} \qquad (4.3)$$

The symbols are defined as follows:

- ζ controls the relative weight given to the quality of service parameters compared to the user experience parameters.

- S_l, P_i, R_j, U_k, represent the quality values given to individual parameters of quality of service measures (S_l), perception measures (P_i), rendering quality measures (R_i), and user state measures (U_k).

- A, B, C are empirically-determined weighing constants for the respective perception measures, rendering quality measures, and user state measures.

- η_l, α_i, β_j, γ_k are weighing factors which depend on the relative quality value of individual user experience parameters underneath quality of service measures, perception measures, rendering quality measures, and user state measures, respectively.

If the quality factors are restricted between 0 and 1, then the overall quality of experience will also have a value between 0 and 1 (i.e. $0 \leq QoE \leq 1$). To achieve this condition, the constant coefficients A, B, and C in eq. ((4.3) should satisfy the constraint:

$$A + B + C = 1 \tag{4.4}$$

In this mathematical model we have combined both the psychological and physiological categories from the taxonomy presented in Chapter 3 into one category called user state. If physiological parameters are included for physical validation in the future, it can be combined with psychological measures, since both categories reflect the user state.

This three-tier organization of the mathematical model allows the evaluator to retain a greater control of the equations from the higher level to the lower one. As an example, supposedly the evaluator wants to focus all his/her study on UX, while maintaining the effects of the QoS constant. In this case, ζ would be set to zero then the full weight would go to UX in the higher-level equation. Moreover in eq. (4.3), supposedly the effect of the user state is negligible, then C could be set to zero and eq. (4.4) becomes A+B=1 and the weights then can be distributed between A and B.

4.2 Employing the Mathematical Model

We test the proposed mathematical model with the application described in Section 3.3.2 developed at the DISCOVER lab of the University of Ottawa. The user testing performed is used to assess effectiveness of the proposed model by comparing the overall QoE computed using the mathematical model with that provided by the testing subjects. This section describes the steps taken to standardize the data of the questionnaire results, determine the various weights established in the mathematical model, and apply those weights on the data to evaluate the QoE of the application.

4.2.1 Standardizing the data

To calculate a weighted average of the QoE, two modifications need to be performed in order for the data to adhere to two rules that we have specified. The first rule is that all the values selected by the users in the questionnaire should be converted to a normalized number between zero and one. This will facilitate the calculation of the QoE value in a percentage format. The second rule dictates that all the values should be in ascending order. That is, the higher end of the value of the parameter indicates better rating; while the lower end indicate worse evaluation.

We applied the two modifications necessary to the results. To normalize the numbers we applied the following formula (x-min)/range, where x is the Likert-scale value selected by the user and range = max- min (max being the maximum value that can be selected by the user and min is the minimum value).

Looking at the results in Table 3.2, all the parameters present follow rule two except 'fatigue'. Higher fatigue ratings in the questionnaire indicate that the application causes higher fatigue. In this case, fatigue is undesirable and higher subjective fatigue values will degrade the user's satisfaction. Modification of the fatigue Likert values was done by subtracting the normalized fatigue value from one. We applied 1- (norm. fatigue value) according to (Tullis and Albert 2008). We renamed the variable as 'comfort'.

4.2.2 Determining the Weights

Primarily, the weights were determined based on the number of the parameters involved, in order to maintain an equal distribution of weights to parameters ratio (Tullis and Albert 2008). For the top level equation, we have one QoS parameter and four UX parameters. Therefore $\zeta = 1/5$ and the top level equation becomes:

$$0.2\ QoS + (1\text{-}0.2)\ UX$$

$$= 0.2\ QoS + 0.8\ UX$$

The same process applies to eq. (4.3). The A, B, and C weights are determined by the number of parameters involved. In our case, we have two parameters from the perception measures category weighed by A, one parameter from rendering quality category weighed by B, and one parameter from the user state category weighed by C. Since we have a constraint in eq. (4.4) that the sum of A, B, and C should be equal to one we can calculate the weights values as follows:

$$A+B+C = 1$$

$$B = C \text{ (since each weighs one parameter only)}$$

$$B= 2A \text{ (A weighs two parameter)}$$

$$B+C=4A$$

Combining the previous equations

$$A + 4A = 1$$

A =0.2, B=C=0.4

The designation of B = 2A come from the fact that since A weighs two parameters then it should be half of the weight that controls only one parameter. This way each category contributes equally to the UX calculation (Tullis and Albert 2008).

The weights associated with each individual parameter (η_l, α_i, β_j, γ_i) can be set to one which allows the single parameters in a category to have full weight, while if a category has multiple parameters each one of those parameters can be equally weighted. Another way to treat multiple parameters in a category is to set their weight according to the correlation value between the parameter and the overall user evaluation which can be inferred from the questionnaire results.

4.2.3 Results Obtained

The formulas (4.1) - (4.4) were applied to each user. Hence for each user we have calculated a QoE rating, then averaged the values and calculated the standard deviation among other mathematical and statistical measures. The QoE results were multiplied by a 100% to get a percentage value. This was possible since we have used standardized values of the data. Table 4.1 displays the results.

TABLE 4.1- RESULTS OF APPLYING MATHEMATICAL MODEL TO DATA

Attribute	Results without parameter correlation	Results with parameter correlation [a]
Average calculated QoE value	74.33 ± 15.80	74.32 ± 15.83
Average of differences	11.73	11.76
Range of differences	0 - 37	0 – 37.3
Percent Error	14.74	14.77
Correlation	0.73	0.73

a. The difference between this column and the other set of results (previous column) is that we have used correlation as weights when there were multiple parameters in the same category

4.3 Analysis

Table 4.1 displays the average QoE value of applying the mathematical model to the data that originated from the questionnaire results along with the standard deviation. The table also displays the average difference between the users' overall ratings and the mathematical model QoE results. Furthermore it displays the range of that difference (the minimum and maximum difference values). Two statistical parameters are displayed, the percent error and the correlation.

There are two sets of calculated results observed. They differ by how the weights of multiple parameters within the same category are generated. In our case there was only one category with multiple parameters, which is the perception measures category. As mentioned before, the weights were either equally divided or they were based on correlation between the parameter and the overall user rating.

It can be noticed that both sets of results are very close to each other. They are almost identical. This is due to the fact that the parameters correlation values are very close in magnitude as well. We had one category with multiple parameters. The perception measures under the UX had two parameters. The correlation values for both parameters with the overall QoE were 0.59 and 0.44 for comfort (opposite of fatigue) and user intuitiveness respectively. No matter what the magnitudes of the weights are, if they are the same value (or close in value) then the result of the weighted average would be the same (or close to the equal weights result). For the rest of the section we will analyze only one set of results (the equal weight results) due to the similarity of both sets of results.

The average of the QoE value calculated was found to be equal to 74.33 with a standard deviation of 15.80. This is close to the QoE value rated by the users which was found to be 84.13 with a standard deviation of 12.51 (from Table 3.3). An indication of the proximity of the results would be the individual differences between the two set of values, for each user.

In Table 4.1 , the range of the differences between the two values is 0 to 37. However these extreme values occur sporadically and rarely within the data. For instance the difference of 37 is an outlier value which occurs only once. A more reasonable representation is to look at the average of the differences which is 11.73. This is a moderate value and is more indicative of the differences between the results of the mathematical model and the user rating.

To further explore the differences between the two sets of values, we have computed the average relative error as shown in eq. (4.5). The percent error (relative error in percentage form) is 14.74%.

$$re_{ave} = \frac{\sum_{i=1}^{n} |QoE_u - QoE_m| / QoE_u}{n} \quad (4.5)$$

where re_{ave} is the average relative error, QoE_u is the overall rating by the user, QoE_m is the value computed by the mathematical model and n is the number of the users.

To additionally validate the model, we can examine the correlation of the data. We have plotted the data of the thirty users in Figure 4.1. One line shows the users' QoE value and another line shows the values computed by the mathematical model. The values follow the same pattern and coincide sometimes. This indicates a high correlation pattern. The correlation value was computed to be 0.73, $p<0.001$ (degrees of freedom=28). This means that the mathematical model results significantly follow the users' ratings of the haptic application.

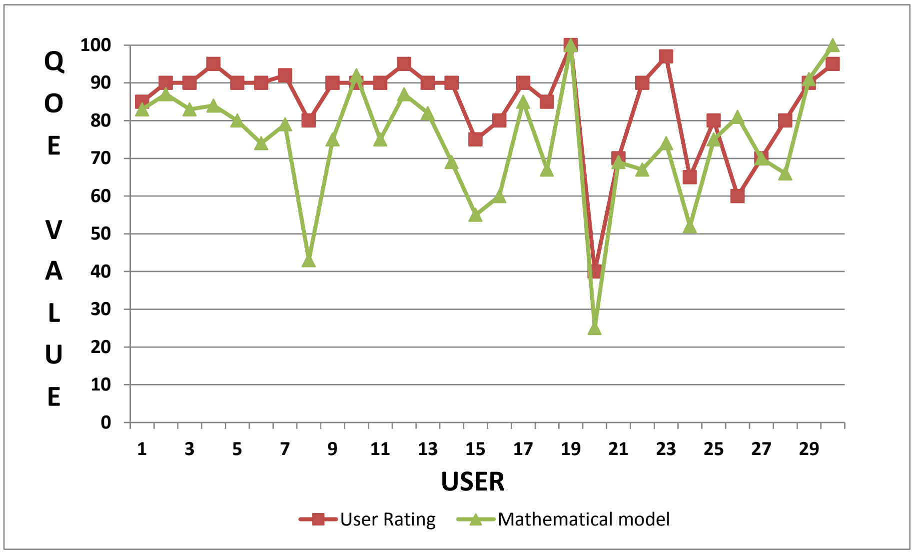

Figure 4.1. Users' ratings vs. mathematical model results.

4.3.1 Data Cleanup

To further minimize the error rates generated previously, we performed analysis on the results when the outliers of the data are removed. It is a common statistic operation to remove the outliers in a set of results (Tullis and Albert 2008). If values are in clear violation of a QoE trend (for example a user who rated all the parameters high, but rated the overall quality of the application low) then theses values can be omitted. In user experimentation jargon this is referred to as data cleanup.

After cleaning up the data, we end up with twenty-two users out of the original thirty from Table 3.1. The eight users whose values were discarded represent clear outliers. The results are summarized in Table 4.2 below along with the correlation figure.

TABLE 4.2- RESULTS OF APPLYING MATHEMATICAL MODEL TO DATA AFTER CLEANUP

Attribute	Results without parameter correlation	Results with parameter correlation
Overall User Rating [a]	84.64 ± 13.00	
Average calculated QoE value	77.59 ± 15.91	77.58 ± 15.92
Average of differences	7.77	7.76
Range of differences	0 - 18	0 – 18
Percent Error	9.95	9.95
Correlation	0.92	0.92

[a] The overall user rating average slightly changed after the data cleanup since we only averaged the data that remained.

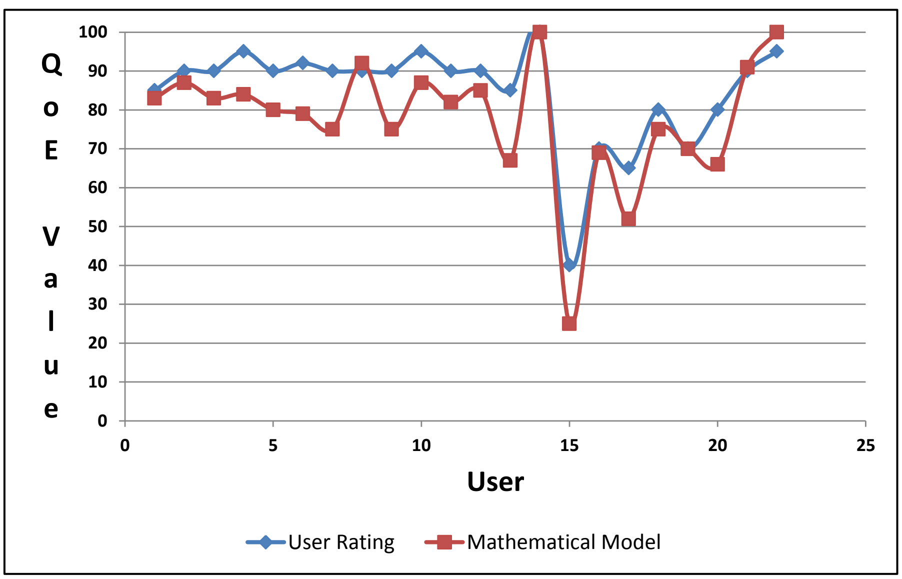

Figure 4.2. Users' ratings vs. mathematical model results after data cleanup.

Data cleanup reduces the percent error to 9.95% while also maintaining a very high correlation between results. An error below 10% is ideal since any error higher than that would result in a shift in the perception of the results. For example the difference between 80% and 90% perceptually would be different than the difference between 80% and 87% although both differences are close to each other. A percent error of 9.95% is acceptable in this case, although we would like to lower it with other weight determination techniques.

4.4 Other Weight Determination Approaches

Aside from the even weight distribution and the hybrid approach (between even weight for categories and correlation between multiple parameters) we have performed other approaches to determine the weights of the parameters in our mathematical model. The other three techniques are: pure correlation weight determination, linear regression, and principle component analysis. The discussion of these techniques assumes that data cleanup of the outliers has taken place.

4.4.1 Weight based on Correlation

This approach calculates the weight of a parameter based on the correlation of that parameter with the overall rating of the application. The higher the correlation value the more weight the parameter encompasses. The correlation formula used is:

$$r = \frac{\sum[(x - Mx)(y - My)]}{\sqrt{(SSx)(SSy)}} \qquad (4.6)$$

To determine A, B, and C of eq. (4.4), which define the weight of each parameter category of the user experience, we calculated an aggregate correlation value of that parameter category (e.g. perception measures) and the overall QoE rating, using eq. (4.6). The aggregate correlations for categories with single parameters are equivalent to the correlation of the parameter. For categories with multiple parameters, the aggregate correlation was an overall correlation value (instead of averaged correlation) according to (Monin and Oppenheimer 2005). The values were normalized in order to satisfy eq. (4.4). For weight parameters of eq. (4.1), we used aggregate correlation values normalized as well in order for them to add up to one.

4.4.2 Linear Regression Analysis

Regression analysis models the relationship between different variables. Linear regression model is a type of regression analysis where there is a regressor x and a response y related through a straight line, represented by eq. (4.7) (Montgomery et al. 2006)

$$y = \beta_o + \beta_1 x + \varepsilon \qquad (4.7)$$

where β_o, β_1, and ε are the intercept, slope and random error component respectively. The coefficient of determination R^2 describes how well the line of the linear regression fits the data.

The linear regression model for the haptic rendering parameter is provided in Figure 4.3 as an example. The QoE value of the user is the response to the regressor which is the haptic rendering parameter in this case. The weight of the parameter would be equal to the strength of the linear regression model represented by the slope of the line multiplied by the R^2.

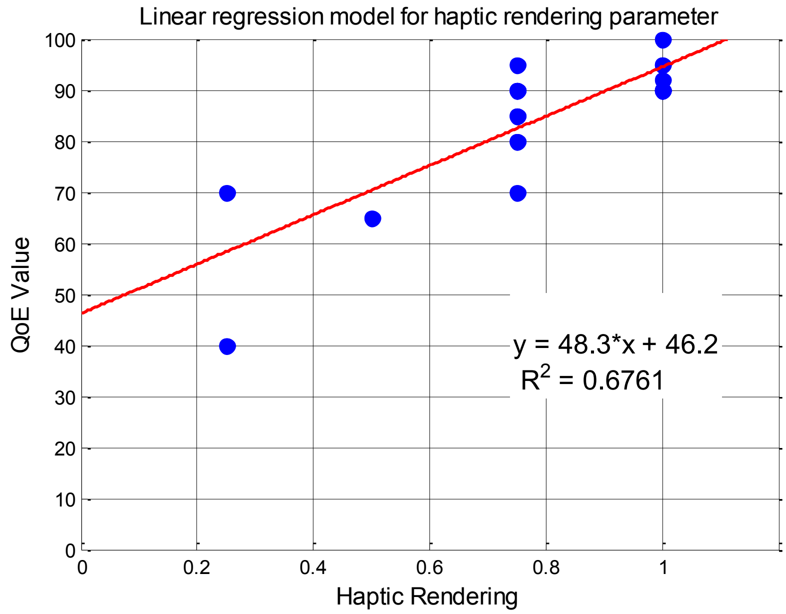

Figure 4.3. Linear regression graph for haptic rendering parameter. The equation is also displayed. The haptic rendering values are the Likert value for each user normalized between 0 and 1.

We have also attempted a linear regression model aggregated for user experience parameters and QoS parameters to compute eq. (4.1). Moreover, for each category of user experience parameters we calculated an aggregate linear regression model to satisfy eq. (4.4).

4.4.3 Principal Component Analysis

Principal component analysis (PCA) is a statistical technique that has been used in various fields such as finite data representation (Szabo and Horvath 1998) and sound recognition (Huadong et al. 1999). PCA transforms the original set of data into a simplified set of data by removing any redundancy that is present in the data (Sauro and Kindlund 2005).

Table 4.3 displays the correlation matrix of all the parameters including the overall QoE rated by the user. The table shows that most of the parameters correlate significantly with one another. This suggests that there is redundant information present in the data. By performing PCA, we remove this redundancy and maintain a set of uncorrelated variables.

TABLE 4.3- CORRELATION MATRIX OF THE PARAMETERS

Parameter	A	B	C	D	E	F
A	1	0.59	0.42	0.4	0.4	0.68
B	0.59	1	0.4	0.42	0.22	0.53
C	0.42	0.4	1	0.3	0.33	0.82
D	0.4	0.42	0.3	1	0.49	0.58
E	0.4	0.22	0.33	0.49	1	0.58
F	0.68	0.53	0.82	0.58	0.58	1

Parameters: A is Media Synchronization, B is Comfort (Fatigue inversed), C is Haptic Rendering, D is Degree of Immersion, E is User Intuitiveness, and F is overall QoE rating. Correlation is significant ($p < 0.05$) if it is greater than 0.36.

The advantages of PCA are threefold:

- Minimizing random error that may arise from one of the measures taken
- Eliminating redundant data from the variables
- Determining how much each parameter weighs in the model and which one has the highest weight

4.5 Comparison between the Weight Approaches

Similar to the previous analysis concerning the even weight distribution and weight correlation combination, the weights derived in each approach were inserted in the mathematical model and a quantified QoE was calculated. For each user there are two sets of results: one derived by the mathematical model and one provided by the user as a subjective evaluation of the overall system. The QoE value quantified by the mathematical model varied by the weight approach.

TABLE 4.4- SUMMARY OF RESULTS OBTAINED BY DIFFERENT WEIGHT APPROACHES

Attribute/Approach	Even weight distribution	Correlation weights	Weight - correlation combination	Linear regression analysis	PCA
Average calculated QoE value	77.59 ± 15.91	76.18 ± 17.22	77.58 ± 15.92	76.053 ± 17.74	78.71 ± 16.22
Average of differences	7.77	9.51	7.76	10.07	7.47
Range of differences	0 - 18	0 – 34.2	0 - 18	0 - 35.9	0-17
Percent Error (%)	9.95	12.04	9.95	12.78	9.51
Correlation ($p<0.005$)	0.92	0.86	0.92	0.86	0.89

We have applied all the weight determination techniques described in this section to the data set after the cleanup (removing the outliers). Table 4.4 summarizes the results obtained.

The similarity between some approaches can be noticed as discussed previously with the even weight and weight-correlation combination. On the other hand, just by considering correlation values as weights then the results differ from even weight distribution. Correlation weights produce higher percent error and higher average of differences. As mentioned earlier, the average of differences is defined as the average value of all the differences between the user ratings and the QoE ratings obtained from the given mathematical model under a specific approach.

The highest average of differences and percent error stems from the linear regression approach. The linear fit had low coefficient of determination values in some instances which suggest that the data may deviate from the linear regression fit and this could be the reason for the high error rate.

All approaches correlated significantly with the users' ratings. The correlation of each approach additionally validates the mathematical model. Even weight distribution had the highest correlation value in all approaches (the weight-correlation combination approach had similar correlation). This means that the mathematical model results significantly follow the users' ratings of the haptic application.

The PCA approach seems to have a slightly better percent error and average of differences than the rest of approaches. Removing the redundancy of the variables had a certain effect which could be attributed to the advantages of PCA. The correlation of PCA mathematical model results and users' ratings is not the highest but it is significant and relatively high.

4.6 Conclusion and Challenges with Mathematical Modeling

This chapter presented a mathematical model capable of quantifying the QoE of users when utilizing haptic-based applications. In this chapter, we utilized the user study conducted and showed how the data were applied in the mathematical model. The chapter focuses on approaches for weight determination of the mathematical model in order to investigate which approach produces the more accurate results when utilized.

The results suggest that there was a variation on how the approaches faired with certain degree of similarity. Linear regression analysis and correlation weights produced the highest error rates. PCA and even weight distribution produced the best results with PCA generating a slightly smaller error rate. A combination of even weights and correlation values for multiple parameters produced similar results as even weight distribution approach.

We have attempted the even weight approach first and our goal was to reduce the percent error by considering different and more advanced approaches such as the linear regression and PCA. Although the PCA approach managed to reduce the error slightly, it is still similar in range with the even weight approach. Our conclusion is that there will always be some limitations when modeling human behavior through a mathematical model and certain range of error is expected. With PCA and even weight distribution this range of error is minimal and acceptable.

Several challenges were encountered with the above mathematical model, among them are:

- The input-output relations are not linear: Given that the data are users' own subjective evaluation of the different parameters involved, we are dealing with non-linearity of the data. For example users might rate the same input values but have a different output perception. Moreover, the nonlinearity is emphasized by the outliers that we have removed in order to reduce the relative average error.

- Weight determination is not trivial: Weights can be computed by experts, or by the number of parameters, or by a mathematical procedure such as principle component analysis (PCA). These methods are error prone.
- It is difficult to infer standardized value for each input type. Since different parameters can be used in the evaluation, each parameter needs to be standardized according to its own type. The most common way to standardize the parameter was shown in the study case here. Sometimes the max and min are not defined, however, such as task completion time or number of errors. In this case we have to define an arbitrary max or min depending on the specific case (Tullis and Albert 2008).

CHAPTER 5

QOE EVALUATION THROUGH FUZZY LOGIC

It has been observed that many QoE parameters are subjective and are fuzzy in nature. For instance there is no crisp answer to whether the user is under stress or whether the application is easy to use. This idea leads to the second type of QoE metric evaluation, fuzzy logic modeling. Fuzzy logic system is needed to map the fuzzy logic inputs to a crisp output, which is in our case a QoE value. The system would vary in the number of inputs provided along with their membership functions (MFs), depending on the type of application we are trying to evaluate.

A Fuzzy logic system would tackle the challenges that were encountered with the mathematical model. For instance, with fuzzy logic systems there is no need to know precisely the weights of each parameter involved. The fuzzy logic inference system maps the input values to the QoE value based on a set of expert rules.

As a proof of concept, we evaluated the QoE of the Balance Ball game described in Chapter 3. The five parameters selected for evaluation in the user study act as the input to the fuzzy inference system (FIS) as described in the following subsections.

5.1 Fuzzy Inference System (FIS)

We first begin by describing general guidelines to build the fuzzy logic inference system. The FIS would be geared to evaluate the QoE metric, in which five steps are required.

5.1.1 Selecting the Input /Output Variables

The inputs to the FIS are parameters that will affect the QoE of the application. Designers of the application would select the most relevant from the QoE taxonomy discussed in Chapter 3.

As for the output it will be the QoE value of the application. This is the value that the owners of the application should strive for. The higher the QoE value the better it would be from the user-centric evaluation perspective.

5.1.2 Selecting a Fuzzy Inference System (FIS)

Some studies have compared different FIS systems in the evaluation of QoE from quality parameters (Hamam and Georganas 2008). There are two well established types of FIS: Mamdani (Mamdani and Assilian 1999) and Sugeno (Takagi and Sugeno 1985). The most fundamental difference between Mamdani type FIS and Sugeno type FIS is the way the crisp output is generated from the fuzzy inputs (Jassbi et al. 2007). While Mamdani FIS uses the technique of defuzzification of a fuzzy output, Sugeno FIS uses weighted average to compute the crisp output. Therefore, in Sugeno FIS the defuzzification process is bypassed.

An example of a Mamdani QoE output is given in Figure 5.1. The outer layer view of the Mamdani QoE output looks like a fuzzy output. In this particular instance the output has five linear MFs: InTolerable, UnAcceptable, Average, Excellent, and Perfect. Once the fuzzy logic system resolves the input values along the given rules, the output can be either fuzzy if the HCI system designer requires a fuzzy output or it can be crisp, as in the QoE evaluation case, which requires an extra defuzzification step.

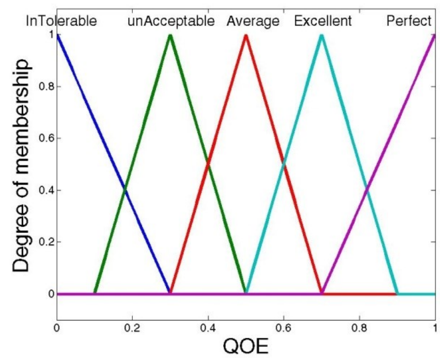

Figure 5.1. Mamdani FIS output membership function.

Since the Sugeno FIS uses weighted average for the output, there is no fuzzy, intermediate output represented by MFs. In order to conceptualize the difference from Mamdani, it is possible to divide the Sugeno output into five levels and label them to correspond to Mamdani's five output MFs in Figure 5.1. The five constant MFs along with their values are given in Table 5.1.

TABLE 5.1- SUGENO FIS CONSTANT OUTPUT

Perfect	1.0
Excellent	0.75
Average	0.5
UnAcceptable	0.25
UnTolerable	0

It can be noted that these are output values, and the labels are just there to assist the design in software suits such as MATLAB. Hence, critics of the Sugeno FIS argue that the expressive power and interpretability of the Mamdani output is lost in the Sugeno FIS since the consequents of the rules are not fuzzy (Jassbi et al. 2006, Jassbi et al. 2007). Table 5.2 summarizes the differences between the Mamdani FIS and the Sugeno FIS (Meitzler and Sohn 2005, Jassbi et al. 2006, Jassbi et al. 2007).

TABLE 5.2- COMPARISON BETWEEN MAMDANI FIS AND SUGENO FIS

Mamdani	**Sugeno**
Output membership function	No output membership function
Output distribution	No output distribution only 'resulting action': Mathematical combination of the rule strength and the output
Crisp result obtained through defuzzification of rules' consequent	No defuzzification: crisp result is obtained using weighted average of the rules' consequent
Non-continuous output surface	Continuous output surface
MISO and MIMO systems	Only MISO systems [1]
Expressive power and interpretable rule consequents	Loss of interpretability
Less flexibility in system design	More flexibility in system design; more parameters in the output

Based on the above, there are some advantages of using either Mamdani FIS or Sugeno FIS. The advantages of using Mamdani FIS are:

- Expressive power
- Easy formalization and interpretability

1 MISO : Multiple Input Single Output
MIMO: Multiple Input Multiple Output

- Reasonable results with relatively simple structure
- Intuitive and interpretable nature of the rule base. For this reason Mamdani FIS is widely used in particular for decision support application
- Can be used for both MISO and MIMO systems
- Output can either be fuzzy or a crisp output

The advantages of using Sugeno FIS are:

- There are algorithms which can be used to automatically optimize the Sugeno FIS. One of the tools that can calibrate the weights of the Sugeno FIS output is *Fuzzy Logic Toolbox for Use with MATLAB* by Math Works Inc.
- Better processing time since the weighted average replace the time consuming defuzzification process
- Computational efficiency and accuracy
- More robust when in presence of noisy input data such as sensor data
- Rules' consequents can have as many parameters per rule as input values allowing more degrees of freedom and more flexibility in the design
- Adequate for functional analysis because of the continuous structure of output function (same inputs do not originate substantially different outputs)

5.1.3 Designing the Membership Functions and Defining Fuzzy Sets

One of the techniques to define the fuzzy sets and generate membership functions (MFs) for the input parameters is to produce data clusters from users' tests. Using MATLAB Fuzzy c-means (FCM) clustering technique, it would be possible to generate clusters of data for each parameter, and based on the result of the clustering generate a standardized fuzzy set.

FCM is a clustering tool provided by MATLAB to find natural grouping in data to represent system's behavior. Unlike other clustering techniques, FCM exploits membership grade to group data points in clusters. Each data point belongs to a cluster to some degree that is specified by a membership grade, thus enabling the clustering of multidimensional data.

As an example, assume a hypothetical case in which "fatigue" was included in the QoE evaluation process. The designers of the application conducted user tests, and based on the users' questionnaire results, a two dimensional data vector was created: The index and the value corresponding to the users' rating of fatigue.

TABLE 5.3- A DATA VECTOR GENERATED BASED ON USERS' QUESTIONNAIRE RESULT FOR FATIGUE

User	Fatigue Rating
1	20
2	20
3	20
4	40
5	60
6	80
7	80

This data vector acts as an input to generate three clusters using the FCM command in MATLAB. The following figure displays the result after running the clustering script

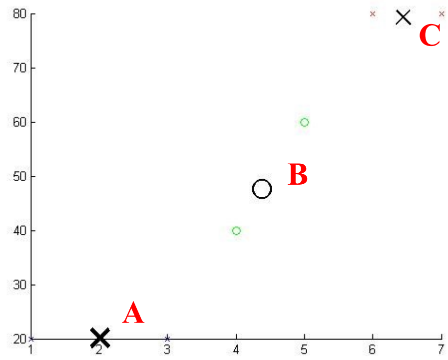

Figure 5.2. Clustering results of the fatigue parameter.

Many papers that standardized fuzzy sets from training data used trapezoidal and triangular membership functions for the input parameters. This common membership functions' shapes were also used in the design of the parameter fuzzy set in this example:

- The first membership function is a trapezoidal membership function. It starts with the maximum truth value of one until the value of the first cluster center is encountered. The membership function starts decreasing until it reaches zero truth value at the second cluster center value.

- The second membership function which is triangular in shape starts at the value of the first cluster center and ends at the value of the third cluster center. It reaches its maximum truth value of one at the value of the second cluster center.

- The third membership function is also a trapezoidal function. From the value of the second cluster center the function starts rising gradually to the maximum truth value of one at the third cluster center. The membership function stays at one till the end of the range of values.

- At each x-axis value, the aggregate truth value of all membership functions is one. To that effect the intersection of membership functions occurs at 0.5 truth value. At this point each membership function has equal truth value of 0.5 and the aggregate is one.

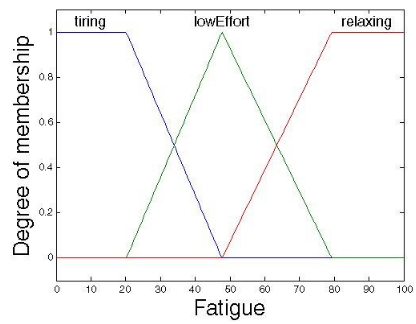

Figure 5.3. Membership functions of the fatigue parameter.

It can be noted that when building the fuzzy sets, it is advantageous to use the linguistic labels to name each membership of every parameter. By harnessing the linguistic advantage of fuzzy logic, you can give meaning to the membership functions instead of just labeling them low, medium and high, such as the example above.

5.1.4 Deriving Fuzzy Rules

The rules set can be introduced and refined by experts of the system. Depending on the application, the experts might decide if the parameters reach certain threshold then the result should be in a certain range. An example of a fuzzy rule is:

If parameter X is high AND parameter Y is medium THEN output is medium

Another way to derive rules sets is from user studies conducted previously that utilizes the application being studied. Based on users' results and comments, the rules can be setup to be parallel with the users' opinion, which is what the user-centric studies should focus on.

5.1.5 Defuzzification and Generating the Output

This step is particular to the users of the Mamdani FIS. A defuzzification step is essential to get that crisp QoE output from the fuzzy aggregate that is generated after applying the rules to the input. The most common defuzzification method is the centroid calculation which happens to be the default selection in the MATLAB Fuzzy Logic Toolbox. This method calculates the center value of the aggregate fuzzy curve generated before the defuzzification step. The centroid method returns this value as the crisp output of the FIS.

5.2 Results Adaptation

The results in Table 5.4 are presented in their normalized percentage form to aid in building and testing the FIS. Since the questionnaire presented to the users was a five point likert scale, each user selects a value from one to five. To convert this value into a percentage, a formula is given by (Preston and Colman 2000) to accomplish the conversion. The formula is

$$(\text{rating} - 1) / (\text{number of response categories} - 1) \times 100 \qquad (5.1)$$

Hence after the conversion, the raw data become: 1 → 0, 2→ 25, 3→ 50, 4 →75 and 5 → 100. The overall quality of experience is given by the user as a percentage form and no conversion was necessary in that case.

We will use the non-shaded results to build the FIS elements, while the shaded results are used for testing.

TABLE 5.4- QUESTIONNAIRE RESULTS OF BALANCE BALL GAME ADAPTED FOR FIS

Subject	Media Synch	Fatigue	Haptic Rendering	Deg of Immersion	User Intuitiveness	Overall Rating
1	75	50	75	100	100	85
2	75	0	75	100	100	90
3	75	25	75	100	75	90
4	100	0	75	75	100	95
5	100	0	100	50	50	90
6	100	0	50	75	75	90
7	75	25	100	75	25	92
8	25	75	25	75	50	80
9	75	50	75	75	100	90
10	100	0	100	75	100	90
11	25	25	100	75	100	90
12	75	0	100	75	100	95
13	100	0	75	75	75	90
14	75	0	75	50	75	90
15	75	75	50	50	75	75
16	100	75	75	25	75	80
17	75	25	100	75	100	90
18	75	25	75	50	75	85
19	100	0	100	100	100	100
20	25	75	25	25	25	40
21	75	0	25	100	75	70
22	75	0	50	75	50	90
23	50	0	75	75	100	97
24	50	25	50	50	50	65
25	75	50	75	75	100	80
26	75	75	75	100	100	60
27	50	0	75	75	50	70

28	50	50	75	75	50	80
29	75	0	100	100	50	90
30	100	0	100	100	100	95

5.2.1 Systematic Sampling

We have planned to divide the results into two parts, one part for building the FIS and the other part for testing. We wanted 25 users' data to be utilized for the building process and the other five users' data for testing. To select which five to use for testing we used a systematic random selection process in which every sixth user of the table be selected. Systematic sampling is used in statistical psychology and is valid as long as the data are randomly organized (Black 2004). In our case there was a random organization of the data where each user had a random chance of being the nth user in the list. The shaded part of Table 5.4 indicates users selected for the testing, which is the 6(*n)th users, $1 \leq n \leq 5$. The rest of the users (non-shaded) are used for constructing the FIS.

5.3 Customizing the fuzzy Inference System

Using MATLAB, we have built a fuzzy logic system to test our QoE model and apply it in evaluating a multi-modal environment. We have followed the steps outlined in Section 5.1, tailored to our user study as it is described next.

5.3.1 Input/output Variables

Five parameters indicated in Section 3.3.3 were selected to act as input to our FIS. The number of parameters is appropriate to be comprehensive yet not too cumbersome in number such that it is difficult to create the rules and logic of the system. More than five parameters will increase the complexity of the fuzzy logic system and it will become harder to construct (Hajshirmohammadi and Payandeh 2007).

5.3.2 Selecting FIS

We have implemented the model using the well-known and established Mamdani inference system based on our previous work (Hamam and Georganas 2008). Moreover the time performance of the Mamdani FIS was very efficient and the results were generated within milliseconds. As seen in Figure 5.4, the five parameters described above act as an input to the system. The Mamdani system applies defuzzification to the output, which is as well modeled with MFs, to generate a crisp output value. Naturally in our case, this output value is the QoE of the user based on the input parameters.

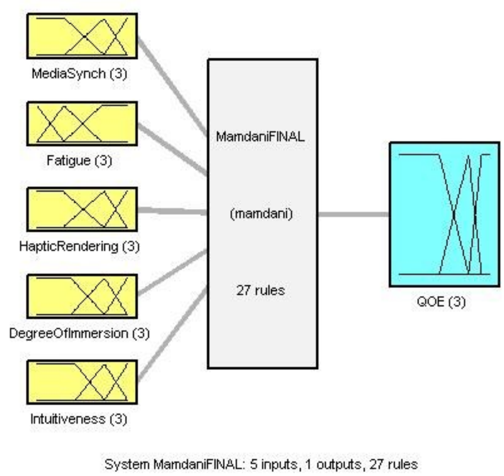

Figure 5.4. Mamdani fuzzy inference system.

5.3.3 Defining Fuzzy Sets (Clustering technique)

In order to define our fuzzy sets and generate membership functions for our input parameters, we have clustered the inputs from the questionnaire results of Table 5.4 (the non-shaded users' results which are used for building the FIS). We have followed the method described in Section 5.1.3, using MATLAB Fuzzy c-means (FCM) clustering technique. All the input parameters and their membership functions can be viewed in Figure 5.5.

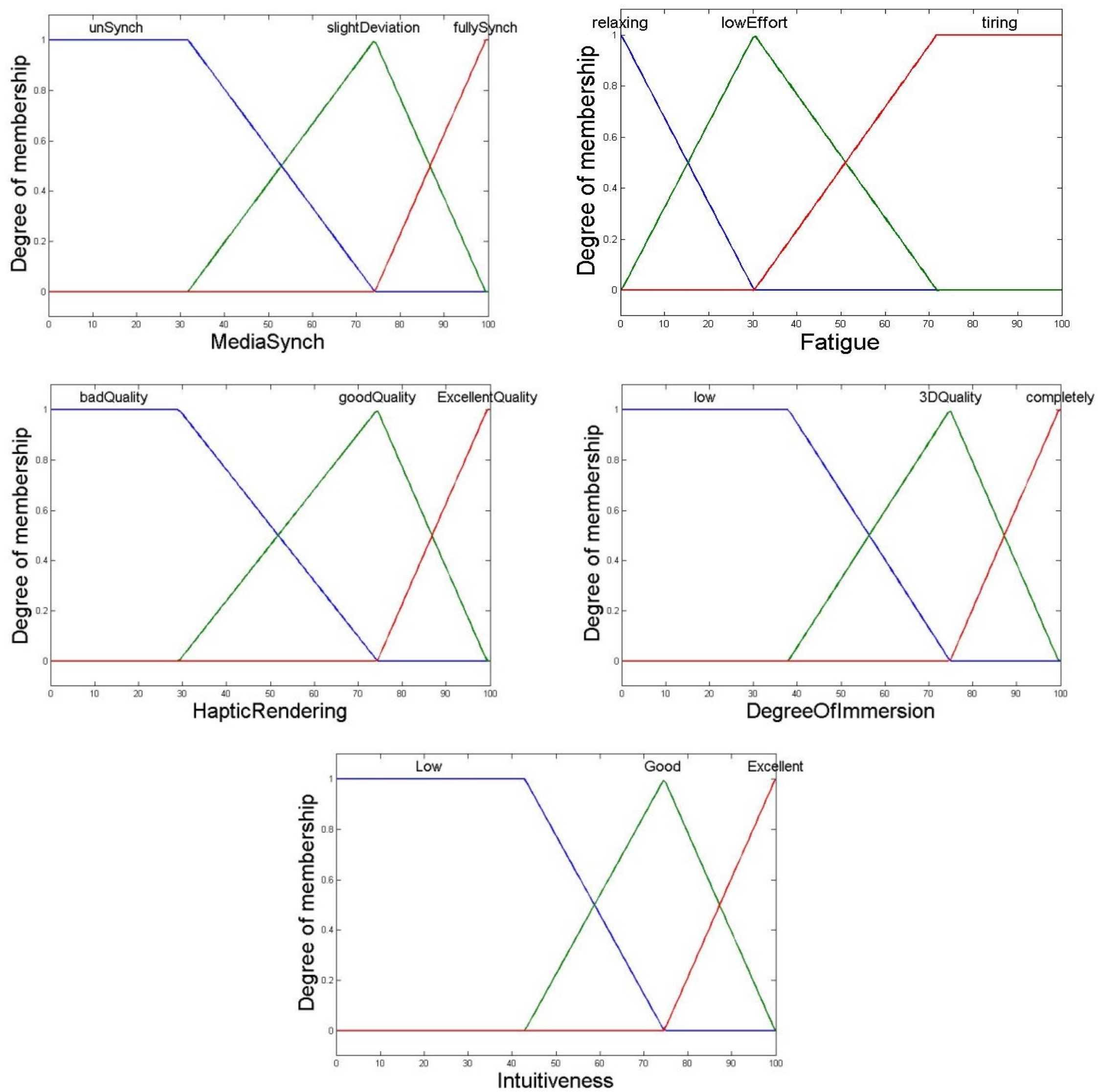

Figure 5.5. Inputs fuzzy sets and membership functions.

The output was divided into three membership functions. Unlike other FIS we looked at in the literature, we wanted to standardize the output MFs too instead of just dividing them into three equally spaced MFs. As a result, we applied clustering technique and the same set of guidelines to the output data set (Figure 5.6), which is the 'overall rating' column of Table 5.4. The output membership function is given in Figure 5.7.

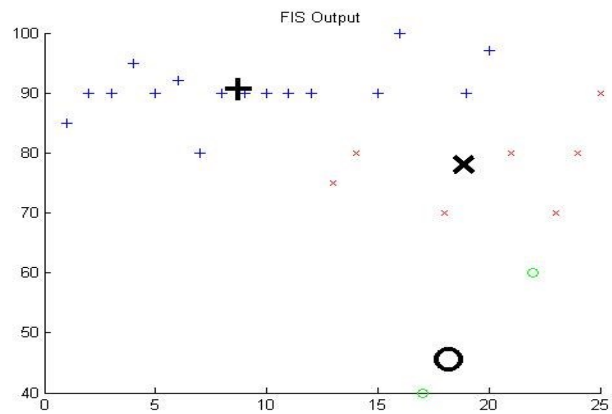

Figure 5.6. Clustering result for the output function.

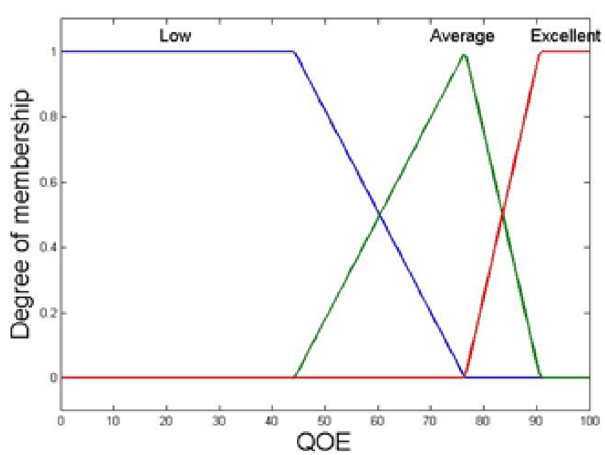

Figure 5.7. Output membership functions for Mamdani FIS.

5.3.4 Deriving Fuzzy Rules

Fuzzy rules were derived from Table 5.4. We generated the rules based on the input and output of the twenty five users that were used to build the data clusters. Table 5.5 illustrates this with an example of rule generation from user three's data. The notion when referring to the rules is L for the lower membership function, M for the middle membership function, and H for the higher membership function.

TABLE 5.5- RULE CONVERSION EXAMPLE

	Media Synch	Fatigue	Rendering	Deg. of Immersion	User Intuitiveness	Overall Rating
U3	75	25	75	100	75	90
Rule	M	M	M	H	M	H

The following rule is translated as: if MediaSynch is *medium* (slight deviation) AND Fatigue is *medium* (low effort) AND Haptic Rendering is *medium* (good quality) AND Degree of Immersion is *high* (completely) AND User Intuitiveness is *medium* (good) THEN QoE is High (excellent).

Often, to generate the rule we have to resolve the value to its most probable linguistic membership function. For example, if we took the input of fatigue in the above table and looked it up in the membership plot of fatigue (Figure 5.5), we will find that 25 on the x-axis has 0.18 truth value in the low MF and 0.82 truth value in the medium MF. The higher the truth value, the more will the value belong to a membership function. Hence we considered 25 to belong to the medium membership function. The rest of the rules were decided similarly, Table 5.6 list all the rules we used.

TABLE 5.6- LIST OF RULES GENERATED FOR THE FIS

Rule	Media Synch	Fatigue	Rendering	Deg. of Immersion	User Intuitiveness	Overall User Rating
R1	M	M	M	H	H	H
R2	M	L	M	H	H	H
R3	M	M	M	H	M	H
R4	H	L	M	M	H	H
R5	H	L	H	L	L	H
R6	M	M	H	M	L	H
R7	L	H	L	M	L	M
R8	M	M	M	M	H	H
R9	H	L	H	M	H	H
R10	L	M	H	M	H	H
R11	H	L	M	M	M	H
R12	M	L	M	L	M	H
R13	M	H	L	L	M	M
R14	H	H	M	L	M	M
R15	M	M	H	M	H	H
R16	H	L	H	H	H	H
R17	L	H	L	L	L	L
R18	M	L	L	H	M	M
R19	M	L	L	M	L	H
R20	L	L	M	M	H	H
R21	M	M	M	M	H	M
R22	M	H	M	H	H	L

R23	L	L	M	M	L	M
R24	L	M	M	M	L	M
R25	M	L	H	H	L	H
R26	H	L	H	H	H	H
R27	L	H	L	L	L	L

The first twenty five rules correspond to the twenty five users' data we used for clustering. Rules twenty six and twenty seven are intuitive. They are basically stating if all input parameters are high then QoE is high. Similarly if all inputs are low then QoE is low (except for the fatigue parameter which is reversed; meaning that low fatigue increase QoE).

5.3.5 Generating Output (centriod defuzzification method)

Since our system is a Mamdani FIS, a defuzzification step is essential to get that crisp QoE output from the fuzzy aggregate that is generated after applying the rules to the input MFs. We used the centroid method which calculates the center value of the aggregate fuzzy curve generated before the defuzzification step. The centroid method returns this value as the crisp output of the FIS.

5.4 Testing and Analysis

5.4.1 Testing the Fuzzy Inference System

To test the fuzzy inference system, we ran the data from the rest of the users that were not used in the clustering procedure (highlighted users in Table 5.4). The data are presented in Table 5.7.

TABLE 5.7- USERS' DATA USED FOR TESTING THE FIS

	Media Synch	Fatigue	Rendering	Deg. of Immersion	User Intuitiveness	Overall User Rating	FIS Output
U6	100	0	50	75	75	90	90.7
U12	75	0	100	75	100	95	89.5
U18	75	25	75	50	75	85	85.9
U24	50	25	50	50	50	65	72.9
U30	100	0	100	100	100	95	92

The last column of the table displays the FIS output value when the five input values of the FIS are taken from the user. For example when the data vector of user six [100 0 50 75 75] is given as input to the FIS, the output of the FIS is 90.7. The actual rating of QoE by user six is 90%, provided in the 'Overall User Rating' column.

5.4.2 Error Calculation

5.4.2.1 Relative and Percent Error

Observing the last two columns of Table 5.4 we can compare the actual value of the user rating and the FIS output. The relative error between the two can be calculated by the following equation:

$$re = {|QoE_u - QoE_f|}/{QoE_u} \qquad (5.2)$$

where QoE_u is the user rating of the application and QoE_f is the output of the FIS. Table 5.8 displays the results.

TABLE 5.8- RELATIVE ERROR CALCULATION TABLE

Overall User Rating	FIS Output	Relative Error (re)	Percent Error (re*100%)
90	90.7	0.0078	0.78%
95	89.5	0.0579	5.79%
85	85.9	0.0106	1.06%
65	72.9	0.1215	12.15%
95	92	0.0316	3.16%

The average relative error can be computed by summing all the relative error and dividing by the number of users utilized for testing (n), according to the following equation:

$$re_{ave} = \frac{\sum_{i=1}^{n} {|QoE_u - QoE_f|}/{QoE_u}}{n} \qquad (5.3)$$

$re_{ave} = (0.0078+.0579+0.0106+0.1215+0.0316)/5 = 0.2294/5 = 0.04588$

Hence, the average relative error in the FIS output is 0.04588 and the percent error is approximately 4.6%.

5.4.2.2 Root Mean-Square Error

Another way to evaluate our FIS output is to consider the root mean-square error (RMSE). RMSE could be calculated according to the following equation:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(QoE_u - QoE_f)^2}{n}} \quad (5.4)$$

RMSE measures the difference between the predicted value of an estimator and the actual outcome. The idea is to calculate how far is the FIS output from the general category of the results.

In our case RMSE = 4.538. This is an absolute and not a relative value. This indicates that on average the estimator (FIS) value deviate from the original value (user's value) by approximately 4.5 points on a scale of one hundred.

5.4.3 Correlation and Statistical Testing

In this section we provide statistical analysis to the acquired results. First we calculated the correlation between the users' responses and the FIS output. The correlation value is 0.98 $p<0.01$. The two sets of data correlates with each other significantly, indicating that the two groups are close in value to each other.

Yet, in this case we want to test whether the two sets of data are actually the same and there is no significant difference between the FIS output and the user evaluation. That is, we want to test that both the FIS output and user evaluation represent the same group of users.

In statistics layman terms we have the following

$$H_0: \mu_1 = \mu_2; H_A: \mu_1 \neq \mu_2 \qquad \text{(H5.1)}$$

Where H_0 is the null hypothesis stating that both sets of data come from the same population, while H_A is the alternate hypothesis which states that both sets of data represent different groups, i.e. both sets of data are not a representative of the same results and differs significantly.

Unlike usual statistics testing (Student's T-test) where the researcher tries to prove the alternate hypothesis, our goal is to reject the alternate hypothesis and accept the null hypothesis. In this case we can be significantly confident that the FIS output is a valid representation of the user's QoE.

To prove the equivalence of the two groups (the Student's t-test null hypothesis previously) we followed the methodology called test of equivalence found in the literature (Streiner 2003, Cribbie et al. 2004). The method consists of finding an acceptable difference (D) between both means of the results in which we still consider them equivalent. Hence the new null and alternate hypothesis become

$$H_{01}: \mu_1 - \mu_2 > D; H_{02}: \mu_1 - \mu_2 > -D; H_A: \mu_1 - \mu_2 \leq D \qquad \text{(H5.2)}$$

Writing the null and alternate hypothesis in this way indicates that if the difference between the two means is greater than D then the two groups are different (null hypothesis) otherwise if the difference is less than or equal to D then the two groups are equivalent (alternate hypothesis) which is what we want to prove. Note that the null hypothesis is divided into two parts to ensure any sign differences that may arise. In other words, both null hypotheses state that the difference between the two means reside outside the range of [-D, D].

The t-value equations for rejecting each of the test of equivalence's null hypotheses and thus accepting the alternate hypothesis are

$$t_1 = \frac{(M_1 - M_2) - D}{S_{M_1 - M_2}} \qquad (5.5)$$

$$t_2 = \frac{(M_1 - M_2) - (-D)}{S_{M_1 - M_2}} \qquad (5.6)$$

where M is the mean of the series, D is the acceptable difference, and $S_{M1\text{-}M2}$ is the standard error of the difference defined by:

$$S_{M_1 - M_2} = \sqrt{\left[\frac{(n_1 - 1){s_1}^2 + (n_2 - 1){s_2}^2}{n_1 + n_2 - 2}\right] x \left[\frac{1}{n_1} + \frac{1}{n_2}\right]} \qquad (5.7)$$

where s represents the standard deviation, n is the number of elements, and the degrees of freedom (DF) = $n_1 + n_2 - 2$.

In our case, $M_1 = 86.0$, $S_1 = 11.13$ for the column 'User Overall Rating', and $M_2 = 86.2$, $S_2 = 6.95$ for the column FIS output. Therefore our standard error is

$$S_{M_1 - M_2} = \sqrt{\left[\frac{4 * (11.13)^2 + 4 * (6.95)^2}{8}\right] x \left[\frac{1}{5} + \frac{1}{5}\right]}$$

$$S_{M1\text{-}M2} = 5.87$$

Next we want to calculate the t-values. The challenge here is to find an acceptable D value. Observing the output function in Figure 5.8, there are four points of interest to consider: the two pole points of the graph (A and F) at x = 0 and x =100 and the intersection of the MFs (B and E) that happens at x = 60.5 and x = 83.5.

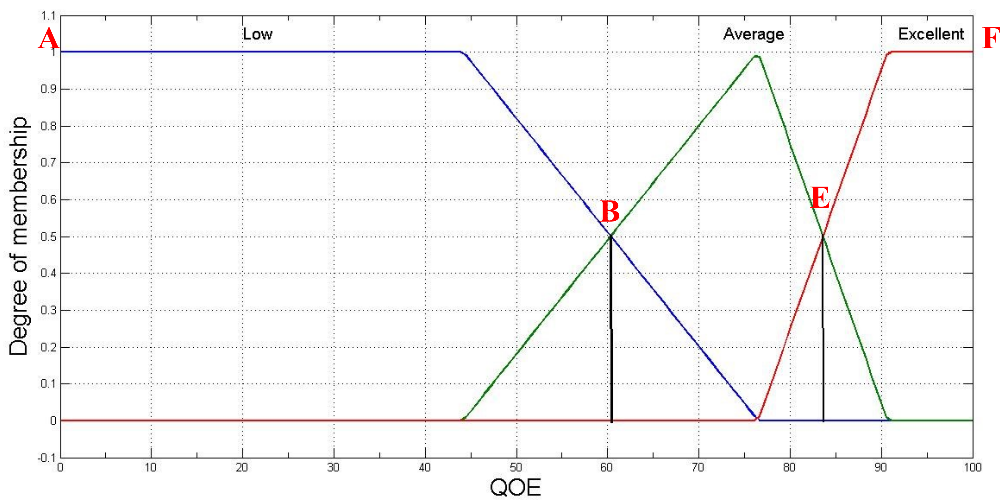

Figure 5.8. Output membership functions for Mamdani FIS with points of interest.

If the FIS output is between the two intersection values (B and E) then the output would be resolved to the Average MF, since it will have the highest truth value (see deriving the rules of Section 5.3.4). If the output is above point E and below point F then it would be resolved to the Excellent MF. If the output is below point B and above point A then it would be resolved to the Low MF. Hence, to find a suitable D, we require that the output does not change its MF status (still be resolved to the previous MF even if we added or subtracted difference D from the output). To find that D value, we subtracted the FIS output from the points of interest (intersection point or the pole point) of the MFs that would apply to it:

|Output – applicable point of interest| = difference in question

|90.7 – 100| = 9.3

|90.7 – 83.5| = 7.2

|89.5 – 100| = 10.5

|89.5 – 83.5| = 6.0

|85.9 – 100| = 14.1

|85.9 – 83.5| = 2.4

|72.9 – 83.5| = 10.6

|72.9 – 60.5| = 12.4

|92.0 – 100| = 8.0

|92.0 – 83.5| = 8.5

The D value in this case would be the average of the differences. D = 89/10 = 8.9. Applying eq. (5.5) and (5.6),

$$t_1 = \frac{(86 - 86.2) - 8.9}{5.87} = -1.55$$

$$t_2 = \frac{(86 - 86.2) - (-8.9)}{5.87} = 1.48$$

The critical value (t_v) when DF = 8 (5 + 5 - 2) is 1.397, p<0.1. According to (Cribbie et al. 2004) if $t_1 < -t_v$ we can reject the null hypothesis H_{01}. Moreover, if $t_2 > t_v$ we can reject the null hypothesis H_{02}. We have t_1 (-1.55) < -1.397 and t_2 (1.48) > 1.397. Therefore we can reject the hull hypotheses and accept the alternate hypothesis with $p<0.1$. This suggests that the two groups are equivalent significantly with $p<0.1$. This result is in agreement with the Student's t-test ($t=-0.0305$, $p=0.98$) that indicates that we cannot reject the null hypothesis of H5.1.

5.5 Evaluation Model Comparison

In the previous and current chapters, we have presented a mathematical model and a fuzzy logic model to evaluate the user's QoE of an application. A user study to evaluate a haptic game was conducted to generate the data to feed both models. Users' actual evaluation was compared to the results obtained by the mathematical model or the fuzzy logic model.

For the mathematical model, initially there were two sets of results acquired. The difference was the weight considered for multiple parameters within the same category. In one set of the results, we used equal weights from multiple parameters while for the other set we used correlation value for multiple parameters (we referred to it as weight-correlation hybrid approach). However, due to the small variation of the parameter correlation, there was no real difference in the results between equal weights and correlation-based weights.

We have also performed other weigh determination approaches, such as PCA and linear regression. However in the analysis below we will use the former approach (even weight distribution) due to the similarity of the approaches' results and to simplify the comparison that will follow between the two QoE evaluation methodologies.

After the data cleanup of the results, the correlation between the mathematical model QoE values and the user QoE overall rating was 0.92 which is high and significant ($p<0.001$).

The average difference between the two sets of results was 7.77. The percent error was 9.95%. The error value is acceptable but we would like to decrease it in the future.

The other QoE evaluation paradigm is the FIS. For testing the FIS, we applied the following evaluation assessments: relative error, root mean error square, correlation, and statistical testing. Although the values are not identical between the user and the FIS output, the percentage error indicates that the FIS values are within reasonable error value compared to the users' values (4.6%). The RMSE calculation confirms that the deviation between the FIS output and the users value is less than five, which on a scale of a hundred is moderately low.

Correlation value between the two results was high as expected given that the results generated low relative and percent error values. The high correlation value is significant, $p<0.01$, which indicates that the two groups of values follow closely the same pattern. Correlation values are important in detecting the direction of the results, but in this case we wanted to prove that the FIS values and the users' ratings represented the same group and that there is no significant difference between the two.

We resorted to two types of statistical analysis which complemented each other. While the Student's t-test produced very high p value indicating that we should reject the alternate hypothesis of (H5.1), that didn't automatically indicate that we can accept the null hypothesis (although it indicates that accepting the null hypothesis would still be an option), as statisticians suggest. This is the reason we performed the second type of statistical analysis, the test of equivalence. With a calculated difference D, we were able to accept the alternate hypothesis of (H5.2), p<0.1. The alternate hypothesis states that both results represent the same group within a small range of [-D, D].

The table below summarizes the difference between the two models.

TABLE 5.9- SUMMARY OF DIFFERENCES BETWEEN EVALUATION MODELS

Criteria	Mathematical Model		Fuzzy Logic Model
	Before Cleanup	After Data Cleanup	
QoE Average	74.33 ± 15.8	77.59 ± 15.91	
Degrees of Freedom (n-1)	29	21	4
Percent Error	14.74	9.95	4.6
Average of Differences (between user rating and calc. rating)	11.73	7.77	4.54 (RMSE)
Correlation	0.73 (p<0.001)	0.92 (p<0.001)	0.98 (p<0.01)

From the table, it can be noted how the results favor the FIS paradigm more than the weighted average, especially if data cleanup is not performed. This is clearly reflected in the percent error value.

5.6 Conclusion

Mathematical modeling is simple to perform once the weights in the equations are established. However, determining the value of these constants with fixed accuracy is a challenging task. Moreover, it is not always feasible to do so, depending on the application and the parameters being evaluated. In this case, a different type of evaluation should take place.

The fuzzy logic inference system introduced is capable of evaluating the user's QoE based on certain input parameters. FIS was chosen due to the vagueness of the inputs and also to the nonlinearity of the input/output relation. The set of rules derived maps the input parameters to a crisp QoE value. The QoE value obtained represents an alternate to the user's rating. The FIS system addresses the shortcomings of the mathematical model and thus can be used as an alternative to a mathematical system without the need for meticulous calculations.

Both the evaluation paradigms attempt to objectify the subjective nature of the QoE. The results verify that the evaluation engines (especially the FIS) can mimic the behavior of the users after user results are used to build the given engines. However, the user is still required to provide the input parameters to the engines. Our way of thinking is that the user at the current stage cannot be replaced completely when evaluating such a subjective metric as the QoE. However the aim of the engines is to build a framework for QoE evaluation such that the number of users can be minimized when testing a given environment, and hence reducing the time-expensive users tests to a small amount. For instance, after building the FIS we can test and evaluate the system with just five users. Even if the input parameters' values were automated (such as Fatigue in Chapter 7), we still require some users to go through the application steps to produce the automated values of the parameters.

We view our engines as a mixture of subjective and objective system. It requires subjective input from the user, but produces an objective QoE value, which designers can use to enhance the quality of their applications by focusing on the most important parameters.

CHAPTER 6

EFFECT OF KINESTHETIC AND TACTILE FEEDBACK ON THE QOE

The previous two chapters focused on multimodal applications incorporating both graphics and haptics. The haptic device utilized in the user study of those two chapters is a kinesthetic device providing mostly force feedback associated with limb movement. The two methodologies described provided a structured solution for evaluating the QoE.

This chapter provides a different perspective of QoE investigation by first focusing solely on the haptic media, including the differentiation and comparison of kinesthetic and tactile feedback. Moreover, the usage of unstructured tactics such as correlation and statistical analysis are used to link QoE to the haptic media.

The integration of haptic modality in multimedia systems depends on whether it enhances the overall quality of the user experience. Furthermore, questions can be raised as how do the advantages of heightened senses overweigh the costs of adapting to a new technology. Moreover, how will users experience such advantages and whether they will be overwhelmed or exhausted. In this chapter, we attempt to answer such questions by exploring two haptic applications: the haptic learning tool (Mansour et al. 2007) (to measure the QoE associated with kinesthetic haptic feedback) and the haptic YouTube application (Abdur Rahman et al. 2010) (to measure the QoE associated with tactile haptic feedback). We base our analysis on the fact that most computer users are adjusted to using the mouse and keyboard and are acquainted with online video streaming websites such as

YouTube. We administer two user studies in this chapter. In one study, the participants evaluate a writing application that integrates force feedback through a stylus-based haptic device. The other study introduces the user to a new way of watching online streaming videos; videos with tactile haptic feedback. The studies also focus on important haptic properties, including positive and negative attributes. The studies along with their accompanied questionnaire are geared to explore the effect on the QoE when haptics are involved.

In this chapter we attempt to find the effect of just the haptics modality on the QoE of the application. The parameters investigated relates solely to the haptic media. For the evaluation part, we relied on the Mean Opinion Score (MOS) (ITU-T Rec. 2006) values to draw correlation and statistical analysis from the results as it will be described later.

6.1 Kinesthetic and Tactile Haptics

Haptic devices are the instruments that enable us to feel the objects in a virtual environment. Each application is usually designed to adapt to a certain type of haptic devices. Human haptic senses can be divided into two distinct categories: tactile sensation and force feedback (kinesthetic) sensation (El Saddik 2007).

Tactile sensation deals with information about the physical surface such as contact force, geometry of the object and temperature. Kinesthetic sensation, on the other hand, deals with forces resulting from position and velocity of the hand motion. Net forces experienced by humans are usually a combination of both types of haptic sensation. However, haptic devices can be classified into tactile haptic devices or force feedback (kinesthetic) haptic devices. The two categories of haptic devices can be distinguished by the users' interaction with the virtual environment. Tactile devices or tactile displays distributes the tactile forces over the region of contact, and simulate that contact to the skin. Currently tactile devices are composed of shape-memory alloys, pneumatic actuators, or vibrotactile elements, among others. Kinesthetic devices will simulate the force and torque to the user through tool such as a rigid stick (Srinivasan and Basdogan 1997).

Since haptic devices primary target the human sense of touch, evaluating these devices and their advantages became an essential research topic. (Robles-De-La-Torre 2006) discusses the importance of haptics technology by discussing the biological sense of touch in humans. He stresses that it is hard to cope without the sense of touch in the real world, so in the virtual world vital information would be lost without the inclusion of haptic devices.

Force feedback effect has been considered in the literature. (Pawar and Steed 2009) discuss the effect of haptic cues on three dimensional selection tasks. They argue that the conception of haptic cues always improving performance is not always true. On the positive side, the authors of (Wagner et al. 2007) discuss the benefit of force feedback in laparoscopic surgery. Their analysis is based on the number of errors and precision of dissection. Reinforcing this idea, (Gwilliam et al. 2009) discuss force feedback effect on surgical performance in minimally invasive surgery.

On the other hand, the effects of tactile feedback on movement accuracy and speed were studied by (Viau et al. 2005). The researchers concluded that tactile feedback may enhance performance when feedback is event related. Tactile feedback has also been proved constructive for several other applications including military and robotics tasks in multitasking environments (Chen and Terrance 2008, Greenway and Faddis 1993), visual speech-reading (Andersson et al. 2001), task performance (Naud et al. 2009), as well as grasping in laparoscopic surgery (Van der Putten et al. 2010, King et al. 2009).

It can be noticed that medical procedures and simulations have received special attention in the literature when investigating the effect of haptic feedback on medical applications. A survey of the effect of both types of haptic feedback on medical simulators has been addressed in (Coles et al. 2011). At the same time, the literature seems to lack studies about the effects of haptic modality onto the quality of user experience.

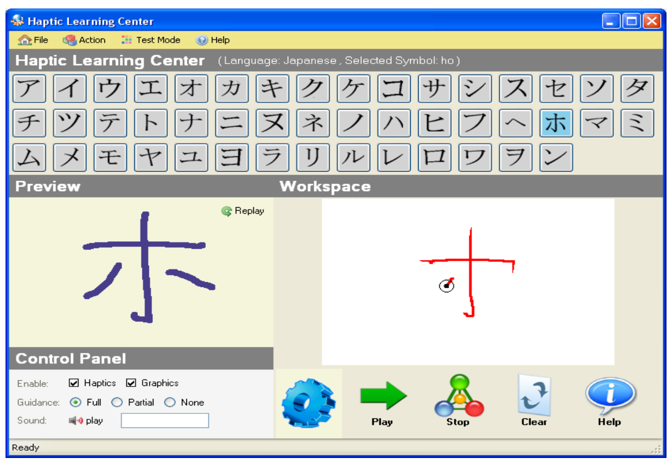

Figure 6.1. Haptic handwriting learning tool. The figure displays the different panes the user is viewing. The right pane shows the graphical tracking while the user manipulates the haptic instrument.

6.2 Force Feedback QoE Experiment

6.2.1 Application Description

The haptic handwriting learning system (Eid et al. 2007) is a haptic GUI application designed to facilitate learning of alphabetic handwriting of various languages by incorporating visual, auditory, and haptic feedback. The application is divided functionally into four blocks: the alphabets keyboard, the preview window, the workspace area, and the control panel (Figure 6.1). The alphabets keyboard contains all the characters of the selected language. The system supports in its current state five languages: Arabic, English, Chinese, Japanese, and French. The alphabets are stored in a language repository that stores the characters using an XML-based description. Users can select their desired language through the File menu which prompts the system to load the corresponding alphabets from the repository. The preview window replays graphically the selected character in vivid fonts to act as a reference for the user when he/she is practicing writing the letter in the workspace. The workspace in turn enables users to experience the handwriting with haptic and graphic feedback. As users are grasping the haptic device and moving the virtual cursor along the whiteboard plane, the graphic feedback allows them to see their progress whereas the haptic feedback allows them to physically experience it. Finally, the control panel controls the playback mode that the user wants to permit in the workspace. Depending on the user's selection, graphic or haptic feedback can be enabled as well as setting the playback guidance level such as no guidance for confident users, partial guidance for users in doubt, or full guidance for novice users. The guidance (if set to full) will move the haptic device in real time according to the selected character, and will display the graphic or/and haptic feedback according to the selected options. If no guidance is selected the haptic device will be transparent allowing the free movement of the user. The partial guidance is currently not implemented.

6.2.2 Experimental Setup

The application ran on a Pentium 4 dual processor machine (3.40 GHz and 3.39 GHz) with one GB of RAM, and an NVIDIA Quadro FX 1000 video card. The screen was placed in front of the users while the haptic device was placed on their working hand side as shown in Figure 6.2. The PHANTOM Omni was the haptic device of choice since it has six degrees of freedom position which allows the users to handle the haptic stylus like handling a pen. Moreover, the three degrees of freedom force feedback provides the desired force in the x, y, and z directions as if writing on paper.

Figure 6.2. Experimental setup at the DISCOVER Lab. The user is manipulating the haptic device to write a particular letter on the writing pane.

Twenty test subjects (19 males, 1 female) were asked to participate in the evaluation study of the haptic handwriting learning tool. Some of the subjects did have working experience with haptics, while others did not. However, none of the participants had Japanese writing background before. The users were introduced to the haptic device if it was unfamiliar to them. They were also introduced to the application and shown the different features and aspects including the various languages that can be loaded and the guidance option which was not enabled beyond the introduction of the application. The subjects were asked to practice the handwriting of three Japanese characters (ho, ki, and yo), two times each with only graphic feedback enabled using the mouse first and then PHANTOM Omni haptic device. The graphic only feedback with the PHANTOM Omni mimics the mouse operation while keeping the hardware setup similar to the force feedback testing. Afterwards, using the PHANTOM Omni haptic device, the users practiced writing the characters three times but this time the haptic and graphic feedbacks were enabled. The playback guidance was set to none to allow the free movement of the users according to their own pace. Following that, the users were asked to write the three Japanese characters on a sheet of paper and a score was given for each. The subjects were reminded that we are evaluating the application and not the users themselves. Finally, the subjects were asked to complete a Likert Scale questionnaire.

6.2.3 The Questionnaire

In the final stage of the experimentation, the users were given a questionnaire to reflect on their own experience with the haptic device when force feedback is enabled. The questionnaire provided adheres to the Likert scale format (Likert questionnaire is also described in Chapter 3). Essentially, most questions are followed by a five point scale in which the users are required to circle the point that is closest to their subjective consideration. Each extreme poles of the scale is marked by opposing descriptive labels based on the content of the questions. The descriptive labels help the user reflect on the question as he/she is writing the questionnaire. For example, the question below asks the users about their rating of the usefulness of the haptic feedback:

Q. To what extent do you think the haptic feedback was useful?

Not Useful   Completely Useful

1   2   3   4   5

The questionnaire consisted of the following questions:

Q1. How would you rank yourself in Japanese handwriting?

Q2. To what extent do you think the haptic feedback was realistic?

Q3. To what extent do you think the haptic feedback was useful?

Q4. To what extent, if any, did using the haptic device increase interaction intuitivism?

Q5. To what extent, if any, did using the haptic device cause fatigue?

Q6. To what extent did using the haptic device make for a better experience than using the mouse?

Q7. How would you rank your experience using the haptic device?

Q9. Have you ever used a haptic device before?

Q11. Give a grade, over 100, for the overall quality of the application? /100

Questions 2 to 7 all followed the above mentioned format (Likert). Questions 8 and 10 were omitted because they are related to features that were not implemented yet. Questions 1 and 9 ask the users about their background in haptic devices and Japanese characters. Question 11 asks the user to evaluate the quality of the application. This is used as the QoE rating of the user. This rating is in accordance with the official definition given by the International Communication Union (ITU) which states that QoE is "The overall acceptability of an application or service, as perceived subjectively by the end-user" (ITU-T Rec. 2008).

6.2.4 Results and analysis

The results of the questionnaire are summarized in Table 6.1. The table's headings are related to haptic parameters. For instance, Realism refers to haptic realism.

TABLE 6.1- RESULTS OF THE QUESTIONNAIRE CONDUCTED FOR THE FORCE FEEDBACK HAPTIC APPLICATION

Parameter	Mean	Standard Deviation
Realism	3.85	0.93
Usefulness	3.95	1.19
Intuitiveness	3.9	1.12
Fatigue	2	1.21
Haptic Vs. Mouse	3.95	0.94
Overall QoE	81.1	14.94

The parameters are the average of the users' responses along with the standard deviation. All the parameters except for the Overall QoE are out of 5. The overall QoE is a percentage scale.

The most relative question in the questionnaire that relates to the effect of haptics on the QoE is displayed in row six labeled Haptic vs. Mouse. Looking back at the questionnaire, Q6 investigates the users experience with haptic device compared with their experience using the mouse. The descriptors on the opposite labels of the five point Likert scale are 'Not at all' and 'Completely'. The mean value of the users' selection is 3.94 ± 0.94 which suggests a preference towards the haptic device as the mean is in the high range. The results are visually shown in Figure 6.3.

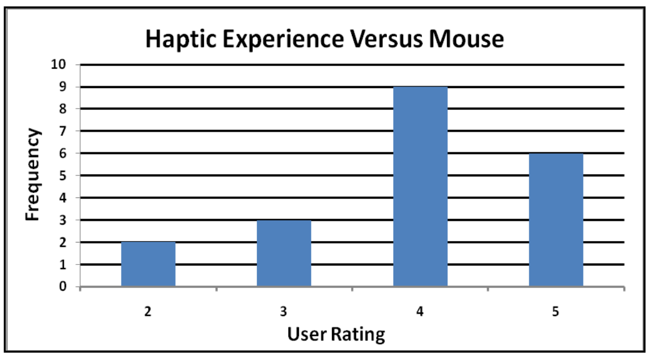

Figure 6.3. Frequency diagram of haptic versus mouse experience. The figure displays the number of users who rated their preference choice for haptic instruments in oppose to the mouse.

Six users concurred that their haptic experience is completely better than their previous mouse experience, in this given context. Only two users preferred using the mouse over the haptic device. Three users found their haptic experience to be roughly the same as the mouse experience as their rating was in the middle of the Likert Scale. The rest of the users lay above average which further supports the favoritism towards the haptic device compared to the mouse. It can be noticed that the majority of the users (15 out of 20) have ranked their preference high (4 or 5).

Nonetheless, the results seem ambiguous when viewed unaided by supporting data. The term 'better experience' certainly suggests users' tendency towards haptic devices, but does not necessary mean that the haptic device would be their desired hardware interface if they had to choose. To further investigate the claim of haptics providing a better experience, we have to observe the last row of Table 6.1, entitled Overall QoE. Here, users rated the quality of the application with a score out of a 100. If indeed haptics increase the overall QoE then users who strongly agree that their haptic experience exceeds their mouse experience will rate the overall quality higher than others. The previous prediction – that haptics add to the QoE – is valid since the majority of users established their strong affinity towards haptics.

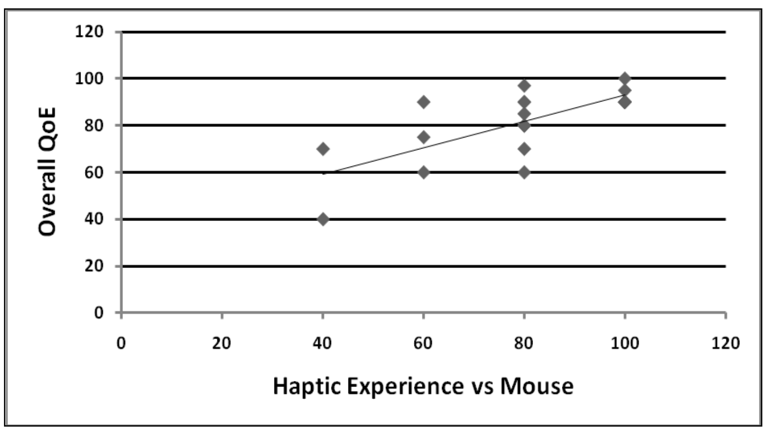

Figure 6.4. Scatter diagram of the results. The correlation mapping between the two parameters is shown by the straight line that represents the best linear fit.

Figure 6.4 displays the correlation between the two variables. The diagram suggests a linear, albeit not perfect, correlation as indicated by the straight line that reveals the overall trend of the scattered dots. We calculated the exact correlation of the two data series according to

$$r = \frac{\sum[(x - Mx)(y - My)]}{\sqrt{(SSx)(SSy)}} \qquad (6.1)$$

where X is the overall QoE series and Y is the Haptic Experience vs. Mouse series (with means of Mx, My and sum of squared deviation of SSx, SSy respectively).

The correlation between the haptic/mouse experience and the overall QoE is found to be 0.713 (degrees of freedom (DF) = 20 - 2 = 18, two-tailed significance at $p<0.001$). This is a high correlation value, which suggests that if users found their haptic experience better than the mouse experience (which is the majority of users) then they are more likely to have a better overall QoE when using the haptic device.

In accordance, we have divided our users into two populations. The first group rated the Mouse vs. Haptics category in the high range; 4 or 5. The other population is the rest of the users. We formulate our hypothesis as the following: The population with high Mouse vs. Haptics rating will rate QoE higher on average. The null hypothesis is that there is no real difference between the two populations, i.e. preferring haptics have no real effect on the QoE.

Performing a t-test on the results we can accept the hypothesis and reject the null hypothesis ($t = 2.86$, one-tailed test, $p<0.01$). Hence, the two groups of subjects significantly differ, which indicates that preference for the haptic device does actually add to the QoE.

It would be possible to examine other metrics that were also included in the questionnaire. Table 6.2 summarizes the correlation results of each parameter.

The realism of the haptic rendering correlates by a relatively high level with the overall QoE. Although it is not as significant as other factors, it is still a high correlation and could be due to the importance of haptic rendering quality in haptic-based applications. If users feel degradation in haptic rendering they would also feel a loss of reality in their touch feedback which would lead to lower perceived QoE (Wu et al. 2009).

It is also important for users to appreciate the purpose of the haptic interface. If they feel that the device is an unnecessary accessory or a burden then that would expectedly affect the overall quality. On the other hand, if they find that the interface is useful in delivering its purpose then the overall quality would be higher. Evidently, the correlation between haptic usefulness and overall quality is 0.527, significant at the $p<0.05$ level.

Surprisingly, there is no significant correlation between intuitiveness and QoE. Some of the users did not have any prior experience with haptic devices while others might not be completely used to them, nevertheless, this does not seem to correlate significantly with the overall QoE. Eventually, people will get used to the interface they are using similar to when the mouse was introduced to the public and people had to adjust for the buttons. However, the results do indicate that the users found the haptic device to be intuitive. Indeed, looking at Table 6.1, the users have chosen high values for intuitiveness ($\mu = 3.9$, sd = 1.12). The haptic device was intuitive for the majority of users but it did not correlate with the overall QoE. Hence, it is not a major factor in our analysis given that the user will have necessary time for adjustment.

The last factor is fatigue. It has a high level of correlation in magnitude but with an inverse direction (-0.615, $p<0.01$). Given that fatigue correlates highly with perceived QoE it is an important factor in determining the desired result of haptic interfaces and whether they would actually increase the application's QoE, or decrease it.

It is also possible to compute the statistical significance of the haptic parameters. The t-tests results are displayed in Table 6.3. Most haptic factors investigated here are significant except for the haptic intuitiveness. This reinforces the correlation results since all factors correlated significantly with QoE except for haptic intuitiveness.

The idea that haptics is one aspect of causality for increasing overall QoE perceived by users is emphasized by the results. The majority of the application's users who found the haptic interface to be realistic, useful, causing less fatigue, and on top of that has a better experience than the mouse also found the application to be more rewarding with a higher QoE.

TABLE 6.2- HAPTIC FORCE FEEDBACK FACTORS AND THEIR CORRELATION WITH QOE

Factor	Correlation	Significance Level (DF = 18)
Haptic Realism	0.408	p< 0.1
Haptic Usefulness	0.527	P< 0.05
Haptic Intuitiveness	0.311	Not significant
Fatigue	-0.615	P< 0.01
Haptic Vs. Mouse	0.713	P<0.001

TABLE 6.3- STATISTICAL ANALYSIS RESULTS FOR FORCE FEEDBACK PARAMETERS

Factor	T-Value	Significance Level (DF = 18)
Haptic Realism	1.99	p< 0.05
Haptic Usefulness	1.85	P< 0.05
Haptic Intuitiveness	0.696	Not significant
Fatigue	2.7	P< 0.01
Haptic vs. Mouse	2.86	P<0.01

6.3 Tactile QoE Experiment

6.3.1 Application Description

The application is composed of a client browser (implemented using Java-based SWT) and the tactile arm band device (Abdur Rahman et al. 2010). The haptic rendering logic is embedded in the client browser and necessary Bluetooth communication module is used to connect the arm band to the computer.

The arm band device embeds vibro-tactile motors that generate vibrations at controllable amplitude, frequency, and duration to simulate different tactile feedback (shown in Figure 6.5). The application streams a YouTube video onto the local machine and presents the video to the user via the tactile player. At the server, the YouTube video is annotated with tactile feedback using XML notation. This tactile content is stored in XML file with timestamps that specify when the actuator is triggered.

The video annotation can be done by the owners wishing to add the tactile feedback to their online video. At the desired time ranges, the authors can add the tactile content with a specific amplitude and frequency for the given duration. Those desired time ranges are decided by the owners of the video at interactive or intense peak moments to add dramatic or entertaining effects.

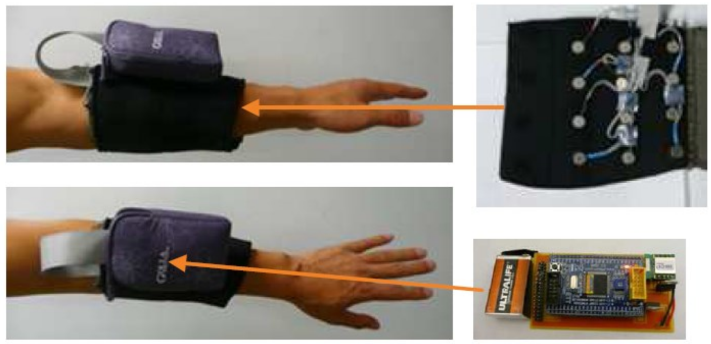

Figure 6.5. The haptic arm band tactile device.

The user wearing the arm band will feel the tactile sensation as series of electrical vibration that is gently stroking through his/her arm. The vibrations intensity will be influenced by the annotation that the author of the video has created and stored on the server using a haptic authoring interface. The vibrations time range would occur at certain events that the author deems worthy of conveying certain tactile simulation to the user. Hence the YouTube annotation is dependent on the scenario of the video.

6.3.2 Experiment Setup

A snapshot of the experimental setup is shown in Figure 6.6. The application ran on a Pentium 4 dual processor laptop (3.40 GHz and 3.39 GHz) with one GB of RAM. The screen was placed in front of the users while the haptic device was placed on their working hand side as shown in Figure 6.6. The haptic arm band tactile device, that was developed as a prototype at the MCRlab, University of Ottawa, was the tactile device of choice due to its availability and suitability for the experimental application.

Figure 6.6. Experimental setup at the DISCOVER Lab. The user is watching a video and receiving synchronous tactile feedback via the arm bend device.

In this particular experiment, the video that was shown contained several parts. The video was created as a demo for the haptic jacket (Eid et al. 2008). The main scenario was a distant husband being reprimanded by his spouse. The two are communicating through a webcam and the husband is wearing a haptic tactile jacket. As the wife is scolding him she humorously punches him. Meanwhile, as the husband amicably ‘feels’ the punch through his haptic jacket, the user watching the video feels the tactile sensation on the wrist arm band (the video was annotated during the punches and during the virtual hugs as well).

Twenty subjects (14 males, 6 females) took part in the tactile YouTube experiment; all of them were students from the School of Information Technology and Engineering at the University of Ottawa. No particular reward was given to them for their collaboration. Prior to the start of the experimental session, the application is introduced to the users along with the objectives of the experiment. The experiment video was shown and explained. The participants were provided with assistance for wearing the armband. All the users were presented with the same tactile video material (described above) and were asked to complete a Likert Scale questionnaire immediately after the experiment (to get as instantaneous feedback as possible). The users were debriefed briefly after the experiment. The subjects were reminded that we are evaluating the application and not the users themselves.

6.3.3 The Questionnaire

The questionnaire was similar to the one given in the force feedback experimentation. However the two experiments were not designed in parallel and there were some changes. Most importantly the scale chosen for the tactile experiment was a seven point Likert-type scale in oppose to five points chosen for the kinesthetic experiment. Both questionnaires are considered a Likert-type scale since there is a middle point and therefore an equal amount of positive and negative items in most questions.

Moreover, there are few choice differences between tactile and kinesthetic parameters to reflect the variation of the haptic devices used. Intuitiveness was replaced by excitement and fatigue by discomfort since the users are not manipulating the haptic device, instead they are wearing it.

The following question constituted the core of the questionnaire:

Q1. Do you prefer using a haptic device while watching an online video (such as YouTube)?

Q2. How realistic is the haptic feedback?

Q3. To what extent, if any, did using the haptic device create discomfort?

Q4. To what extent do you think the haptic feedback was useful?

Q5. To what extent, if any, did using the haptic device increase excitement?

Q6. Rate the overall experience you had during the demo

Questions 1 to 5 are a seven-point Likert-type scale. Question 6 asks users to rate their overall experience during the demo. This was the overall QoE rating. In this case also, the QoE rating is in agreement with the ITU definition.

6.3.4 Results and Analysis

The result of the questionnaire conducted is displayed in Table 6.4. The correlation between the QoE and the different parameter are presented in Table 6.5. Most parameters correlated significantly with QoE except for excitement which did not have a significant correlation. The parameter of focus in the tactile experimentation is the haptic preference which has a correlation of 0.77, $p<0.001$ with QoE. This is the user's answer to the question, "Do you prefer using a haptic device while watching an online video (such as YouTube)?". The frequency distribution of the results is shown in Figure 6.7. Fifteen out of the twenty users rated their preference high (five or above), while five users were either neutral or preferred to watch online video without haptic feedback.

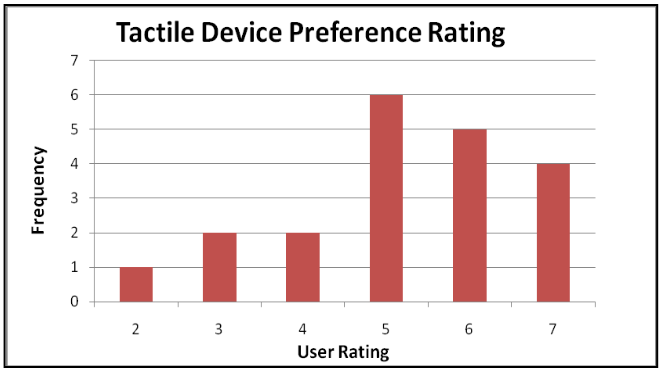

Figure 6.7. Frequency diagram of tactile haptic device preference rating. The diagram represents the subjects' ratings to their preference of using a haptic device while watching an online video.

In the tactile case as well, discomfort had a significant inverse correlation, while realism and usefulness had high correlation values with QoE. Discomfort parameter is an important factor in the user perceived QoE, since any slight discomfort will reduce the quality of the application. The realism and usefulness in this case is dependent on the scenario of the video the users were subjected to. If the video has some action elements that are suited to be represented by tactile feedback then users will find it useful. The video we demonstrated to the subjects contained such a scenario, which clarifies why usefulness and realism had high correlation values.

All factors including excitement were statistically significant (Table 6.6). Excitement did not correlate significantly with QoE, but the population of users who rated excitement higher also viewed QoE higher significantly than the rest of the users. Therefore we cannot exclude the excitement factor from the analysis, as we did with intuitiveness in the force feedback analysis. Intuitiveness effect will fade as users get used to the application, but excitement should be present, although in certain cases users get excited about new technologies but without having a decent QoE.

TABLE 6.4- RESULTS OF THE QUESTIONNAIRE CONDUCTED FOR THE TACTILE HAPTIC APPLICATION

Parameter	Mean	Standard Deviation
Realism	4.9	1.55
Usefulness	5.7	1.03
Excitement	5.9	1.55
Discomfort	2.6	1.05
Haptic Preference	5.2	1.44
Overall QoE	81.0	15.53

The parameters are the average of the users' responses along with the standard deviation. All the parameters except for the Overall QoE are out of 7. The overall QoE is a percentage scale.

TABLE 6.5- TACTILE FACTORS AND THEIR CORRELATION WITH QOE

Factor	Correlation	Significance Level (DF = 18)
Haptic Realism	0.86	p< 0.001
Haptic Usefulness	0.68	P< 0.01
Excitement	0.38	Not significant
Discomfort	- 0.75	P< 0.001
Haptic Preference	0.77	P< 0.001

TABLE 6.6- STATISTICAL ANALYSIS RESULTS FOR TACTILE HAPTIC PARAMETERS

Factor	T-Value	Significance Level (DF = 18)
Haptic Realism	3.84	p< 0.001
Haptic Usefulness	2.68	P< 0.01
Excitement	1.96	P< 0.05
Discomfort	3.7	P< 0.001
Haptic Preference	4.63	P< 0.0005

6.4 Kinesthetic and Tactile Results Comparison

In this section, the results obtained from Section 6.2.4 and Section 6.3.4 are compared. The basis of comparison is to detect similarities between the two sets of results. Since the experiments were different in content and tasks, it cannot be deduced that one type of feedback is more positively or negatively perceived by the user. The separate effect of each type of feedback on the user is the desired basis for the analogy.

Comparing both the force feedback analysis and the tactile analysis, it is remarkable the similarity of the results between the two. The QoE correlation results followed a strikingly similar pattern with intuitiveness parameter in the force feedback application being replaced by the excitement parameter. Discomfort/fatigue had an inverse high correlation, while realism and usefulness had a positive high correlation.

Fifteen out of twenty users in each scenario had a preference for haptic devices over traditional methods such as the mouse or over online videos without tactile feedback. Although the two groups which performed the kinesthetic experiment and the tactile experiment were independent of each other still they have managed to keep tight results. This is furthermore manifested in the mean QoE value for both applications along with the standard deviation. For force feedback experiment, where users were asked to enter an overall experiment rating out of a 100, it was 81.1 ± 14.94 while for the tactile experiment where users rated the application from 1 to 10, it was 81 ± 15.53.

Statistical analysis presented a different way than correlation to look at the data. Correlation shows the direction of the results and the relationship between a given parameter and the

QoE. On the other hand, statistical analysis provides causality and an indication that the factors presented to the user indeed made a difference in their perceived QoE.

More importantly is that the majority of the users (75% in each case) were in favor of utilizing haptic devices. And since they had a higher QoE then it is an indication that for the majority of our sample population haptic did cause a higher QoE and the results are statistically significant in both tactile an kinesthetic case.

Final thought on the comparison between tactile and kinesthetic results is about the statistical effect size that has become a preferred element in recent years to be included with hypothesis testing (Aron and Aron 2007). Effect size is calculated according to the following equation

$$r = \frac{Mx1 - Mx2}{Sx} \qquad (6.2)$$

which is the difference of means between the two populations of the hypothesis over the standard deviation of the population.

It seems that the tactile experimental group was more emphatic than the kinesthetic group with their results. Looking at Table 6.3 compared with Table 6.6, the t-values in the latter are much higher for most significant parameters. Since both groups have the same degrees of freedom (20 users in each group) and similar standard of deviation for the overall QoE value, the higher t-values (which depends on the difference of means as well) indicate a higher effect size in terms that the null hypotheses can be rejected with a higher probability when considering the significant metrics.

6.5 Conclusion

In this chapter we have presented two user studies steered towards determining the effect of haptics on the perceived QoE by the users. One user study was aimed to examine force feedback (kinesthetic) haptic metrics and their correlation with the overall QoE as well as the statistical significance of these metrics. The other study considered tactile haptic metrics and similarly examined correlation and statistical significance.

The conclusion for both studies is that haptic devices, whether tactile or kinesthetic, will add to the QoE if users prefer to use these devices over traditional media hardware. The majority of the users did state their preference towards haptic devices and based on the results of the user studies, it can be statistically assured that haptics will improve the QoE of these users.

CHAPTER 7

DEDUCING USER'S FATIGUE FROM HAPTIC DATA

It is more apparent from the previous chapter that in order to assess the overall satisfaction of the user of multimedia applications utilizing haptic interfaces, it is necessary to evaluate the user's fatigue factor. One effect that correlates strongly with QoE of haptic-based applications is fatigue. Chapter 6 showed that perceived fatigue and comfort could impact the way users self-evaluate the new interface when they use it in various applications.

Fatigue could be an undesired by-product of haptic-based applications, as in the previous chapter. In some cases, however, fatigue can increase the quality of the application if the desired goal is, for example, to increase the endurance of the user in virtual exercise training. In both cases, fatigue assessment is important and can lead to better evaluation of QoE of haptic-audio-visual applications. In (Hamam et al. 2008b), we have shown the role of fatigue in shaping the QoE evaluation of a haptic-based application. Fatigue was determined based on a questionnaire administered to users of the system as well as observations done during usage of the haptic device. Reported results of the questionnaire were transformed into a percentage quantity for each user.

Fatigue is classified into mental fatigue and physical fatigue. Physical fatigue is of importance when it comes to haptic manipulation as the arm is moving continuously. Physical fatigue is defined as the inability to work certain muscles according to the capability of the individual (Lou et al. 2001). Physical fatigue is associated with lack of energy which indicates that the ability of individuals to continue the task diminishes with time as they continue their activity without rest. On the other hand, mental fatigue is not correlated with muscle movement and thus it is of less concern to haptic manipulation since it is application dependent. In this chapter, we will be referring to physical fatigue as fatigue by itself.

The definition of fatigue is not universally agreed upon. Some authors argue that although early definitions of fatigue relate the concept to the lack of ability in maintaining the required or expected force or a reduction in peak force of a muscle contraction, these definitions do not account for variations in muscle movement and the user sensation of force (Enoka and Stuart 1992; Barry and Enoka 2007). The authors adopt the following definition: "acute impairment of performance that includes both an increase in the perceived effort necessary to exert a desired force and an eventual inability to produce this force".

The above definition is excessive and perhaps could not be reproduced with user experimentation involving haptic devices. Regular haptic users will not experience the acute impairment described above. However the same authors state that any activity involving the neuromuscular system at any intensity will eventually produce fatigue. The higher the force exerted during a task, the faster a muscle fatigues.

Specific to repetitive tasks, fatigue does influence the movement of the user (Gates and Dinwell 2008). Perhaps it is subtle changes in the data that may influence the users' profile (speed, position, force, among others). The effect of fatigue may also be subjective and perceived by users as they perform the repetitive tasks.

The analysis of fatigue through questionnaire may fit into certain contexts such as (Hamam et al. 2008b). However, in certain cases we might need to compute fatigue objectively without a questionnaire and without involving the user at all. Questionnaire evaluation, though useful in many scenarios, has been criticized by directly engaging the user in the evaluation leading the results to be slightly biased according to users' interpretation (Whalen et al. 2003). An indirect measurement is harder to define but will get the intended results with the advantage of the user being unconscious of the whole procedure and what is being measured.

In this chapter, we use indirect measurement to objectively detect users' fatigue when working with a haptic-enabled application. The data collected during the interaction with the application are analyzed to see if there are patterns that would indicate signs of fatigue. These patterns can then be used by the designers of the application to control the flow of movement according to the context of the application.

Our idea is based on how the repetitive movement of haptic device users may experience fatigue. Typically, during the utilization of a haptic-based application, the user manipulates the haptic device in a repetitive manner. In essence, we look into the users' data and look into their profile while performing repetitive tasks on a virtual reality setup. We ask the following questions: Is the users' velocity an indication of the level of energy and thus level of their fatigue? Does the repetitive motion of users generate a specific force profile? Does users' perception of fatigue match their force profile? We first start by describing ways of fatigue assessment in the literature followed by the user study and experimental setup.

7.1 Fatigue Assessment

Objective fatigue assessment during an evaluation can take three forms: physiological measures, behavioral measures and visual measures (Qiang et al. 2006). Physiological measures are intrusive to the user, while visual measures could be error prone. Behavioral measures are gaining popularity with the disadvantage of being computationally intensive. Subjective measures can also be used to assess users' perceived fatigue (Potkonjak et al. 2002).

Fatigue traditionally has been assessed using a questionnaire. For instance, some researchers (Lee et al. 1991, Beurskens 2001) use a questionnaire to detect the fatigue among users during different situations and lack of sleep scenarios.

In (Kondo et al. 2005), the authors analyze the level of familiarization and fatigue for different contact states using the CyberGlove and cylindrical objects. The analysis is based on the tasks that will cause fatigue such as moving from light to heavy grasps. In (Ito and Yokokohji 2009), the authors measure the effect of an armrest when maneuvering master control devices. The assumption is that the arm rest will reduce muscle fatigue by reducing the gripping force necessary to maneuver the device.

In a different direction, (Kahol et al. 2008) measure the fatigue of surgical residents by assessing their psychomotor and cognitive skill evaluation using a virtual reality simulator with haptic feedback. Fatigue via surgical skills was assessed by task completion time, hand-and-tool-movement smoothness, and cognitive errors.

(Ma et al. 2010) provides a framework for modeling a digital human in virtual reality environment. The virtual human has a set of parameters that would describe the motion but unique in this paper is the modeling of fatigue and incorporating it in the framework. Joint fatigue is evaluated by the decrease of strength in the joints. The virtual framework is tested using a virtual human undergoing a hole-drilling task.

7.2 Haptic Signature Application

In (Alsulaiman et al. 2008), we have developed a haptic application that allows users to hapticaly write on a virtual background. In that particular case, the users were required to hapticaly sign the virtual application. The data acquired was used to identify users based on haptic features collected. In this chapter, we use the same application running in three dimensions on different hardware to infer users' fatigue using similar sets of data collected.

7.2.1 Application Description

The haptic writing environment provides a virtual environment where users can perform various writing tasks including writing their own signature on a virtual plate. As can be seen in Figure 7.1, the users manipulate the haptic device as a pen and its 3-dimensional position is mapped to a cursor in the virtual environment. When the cursor collides against a white rectangular virtual plate, the users can feel the repulsive force based on the penalty-method and blue dots are drawn on the collision position. A Phantom Desktop haptic device was the haptic device of choice for this application since it has six degrees of freedom related to positional and rotational movement and three degrees of freedom related to force feedback. Most importantly, it can measure 3-dimensional position and orientation of the end-effector. The hardware setup displayed in Figure 7.1 is manufactured by Reachin Technologies. It allows users to work in a three dimensional environment adding more realism to the VR application.

Figure 7.1. Haptic signature application. This figure shows a user signing his name on virtual background using a haptic device.

7.2.2 Experimental Setup

Fifteen male users of different ages (25-35) have volunteered to participate in the experiment. The level of haptic experience varied between users. We requested every user to provide sixty handwritten signatures using our system. We did not start the experiment until the user felt comfortable with the environment and after virtually signing at least once without any complications. For users who have experience with haptic devices, we started capturing their signatures from the second trial. Each user performed 60 recorded trials. Users were asked to fill out a short questionnaire after the 30th trial as well as the 60th trial. The questions reflect the users' mental state during the middle and end of the experiment regarding their fatigue level. The questionnaire provides the subjective results required to reinforce and validate the fatigue inference we conducted on the data collected.

7.2.3 Data Collection

Many attributes have been recorded during the performance of the trials. When a user writes his/her signature on a virtual plate, the 3-dimensional position (p), force applied (f), velocity (v), and angular rotation (a) of the virtual pen-tip were measured and recorded in a csv file at each timestamp (t). A simple element that represents a state in our system can be described as the vector s = {px, py, pz, fx, fy, fz, vx, vy, vz, ax, ay, az, t} where subscript x, y, and z represent spatial dimensions. Each trial consists of thousands of s elements.

7.2.4 Questionnaire Results

The results of the questionnaire given to users right after trial 30 and trial 60 (last trial for each user) are displayed in Figure 7.2.

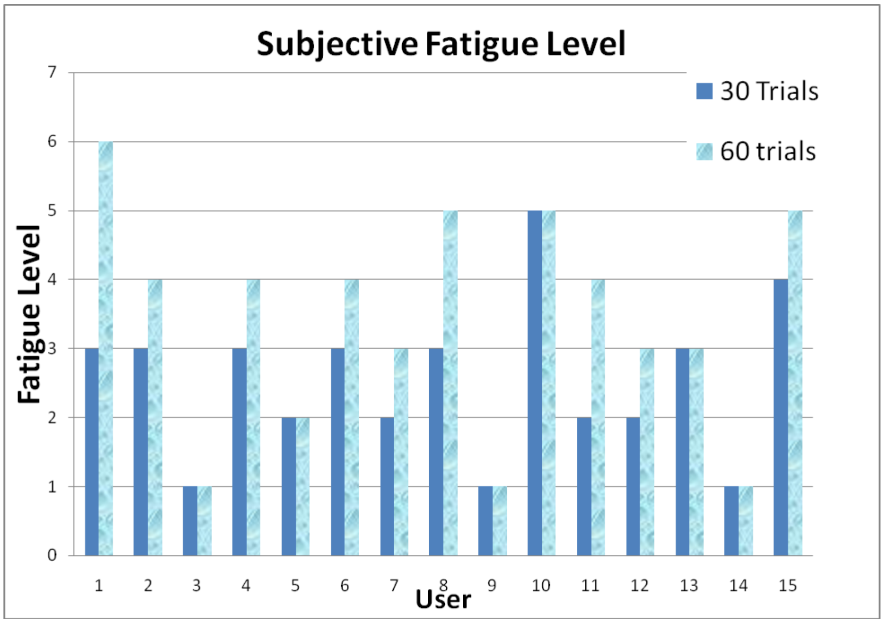

Figure 7.2. Fatigue level questionnaire results.

Figure 7.2 displays the result of the Likert questionnaire administered to the users which investigate the level of fatigue they are experiencing during their haptic usage. The Likert scale ranged from one to seven (seven donating high fatigue).

Based on the results, we can notice that most users did feel more tired at the end of the experiment when compared to the middle of the experiment. Other users indicated that their fatigue level remained the same throughout the trials 30 to 60.

7.2.5 Task Completion Time

In order to analyze the data objectively, the leaning effect on the users must be considered and dealt with, such that the data analysis avoid that consequence.

When dealing with user experimentation, users will take time to adapt to the hardware, experimental setup, and the task at hand. Learning effect is the bias in data that occurs due to users performing the experiment while still learning the subtleties of the experiment.

In other words learning effect is unavoidable in every experiment. We have assumed that it will take the user few trials to get used to the hardware setup. We have averaged the users' task completion time (TCT) for each trial, and monitored when it will stabilize, as Figure 7.3 displays.

The average TCT value keeps decreasing until trial 30 is approached, in which the TCT value stabilizes and fluctuates around 4 seconds. This indicates that at the beginning of the experiment, users are getting accustomed to using the haptic device and signing the virtual cheque. At around trial 30, the learning effect would be at a minimum and users are comfortable with the procedure.

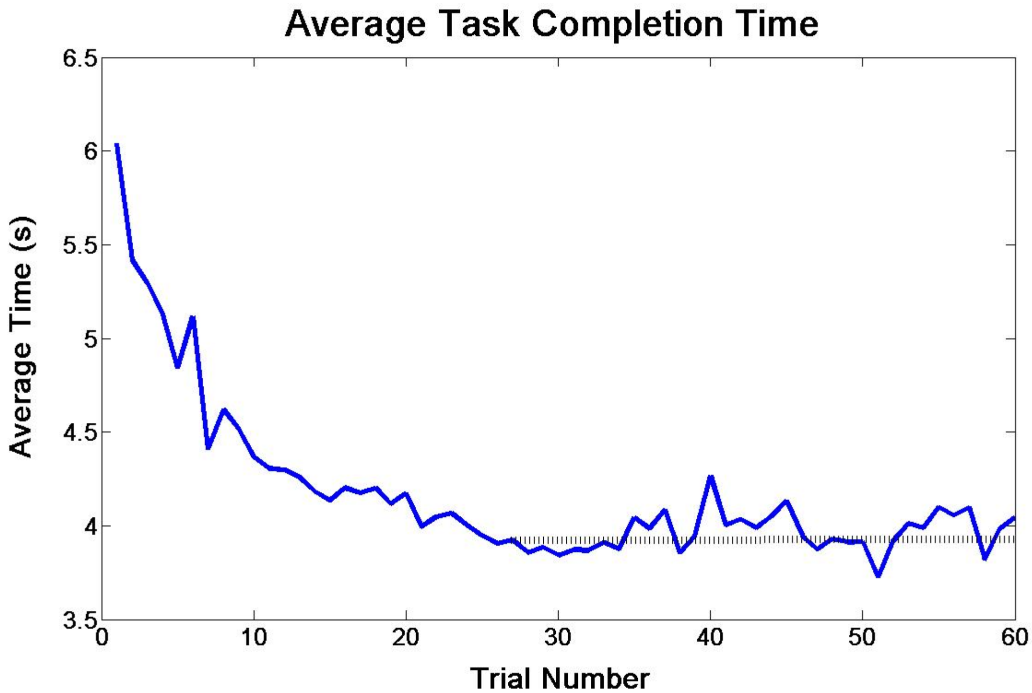

Figure 7.3. TCT average value per trial.

7.3 Velocity Based Energy Analysis

7.3.1 Energy Calculation

Users' constant motion can lead to users being tired after certain amount of time. The more time they spend utilizing their muscle forces the more tired they are going to get. Fatigue has been linked to the lack of energy in users (Lou et al. 2001). Our assumption here states that if the user manipulating the haptic device gets tired by maneuvering the device over a certain period of time then his/her energy level would decrease in magnitude.

Energy is divided into potential energy and kinetic energy [http://www.physicsclassroom.com/Class/energy/U5l1c.cfm]. Since the arm movement that manipulates the haptic device involves position displacement, we focused on examining the kinetic energy and its relation with fatigue. The following formula defines the kinetic energy where m is the mass and v is the velocity magnitude,

$$E_k = \frac{1}{2} mv^2 \qquad (7.1)$$

For a given user, the mass of the haptic device and the user's arm is constant. Looking at eq. (7.1), if we want to compute the difference in energy for the same user, then, the only variable that is changing is the velocity magnitude since the mass is constant. Therefore the difference in energy at any given time can be formulated as follows

$$E_{k1} - E_{k2} = \frac{1}{2} mv_1^2 - \frac{1}{2} mv_2 2 = \frac{1}{2} m (v_1^2 - v_2^2) \qquad (7.2)$$

Consequently, the magnitude of the velocity is the deciding factor to examine the change in kinetic energy for a given user at any point in time. For a user x, the task performed is the repetitive hand signature task for 60 trials. For that specific user, the signature is constant and hence the task is constant relative to the user. The force F applied by the user is divided into fx, fy, and fz according to the Euclidean space. Each force is associated with a Euclidean space displacement given by vector D which can also be divided into dx, dy, and dz. From displacement over time the velocity in each direction can be computed and the velocity magnitude can then be calculated.

Since the results constituted hundreds of data vectors, for each trial, sampled at fractions of a second, we averaged each 30 together to get the velocity over a bigger time frame. Our algorithm for computation is shown in Figure 7.4.

```
for each user
 for each trial
  for each sample greater than 1
    calculate deltaX , add it to displX
    calculate deltaY, add it to displY
    calculate deltaZ, add it to displZ
    if (sample number reaches 30)
            velX = displX/(deltaTime for the 30 samples)
            velY = displY/(deltaTime for the 30 samples)
            velZ = displZ/(deltaTime for the 30 samples)
            Ei = velX ^ 2 + velY ^2 + velZ^2
            Store Ei, reset values
for each user
      for each trial greater than 30
            Ediff = Ei at 30 – Ei at current trial
            Sum = Sum + Ediff
      store sum
for each user plot sum vs trial (start from trial 31)
```

Figure 7.4. Algorithm for computation of energy.

Our goal is to compute the sum of the energy differences between different trials and to determine if the energy is decreasing or increasing with successive trials.

From the algorithm above, it can be noticed that we started our base trial at trial 30 not at trial 1. That is in accordance with TCT time stabilizing at around 30. This way we are able to distinguish between learning time and actual velocity changes due to decrease in energy as practice might affect the velocity of the user at the beginning.

7.3.2 Energy Difference

As mentioned, we took the energy difference as a sum of energy differences between two trials at various times during the experiment. We took trial 30 as the base trial and compared the differences of later trials with the energy at trial 30.

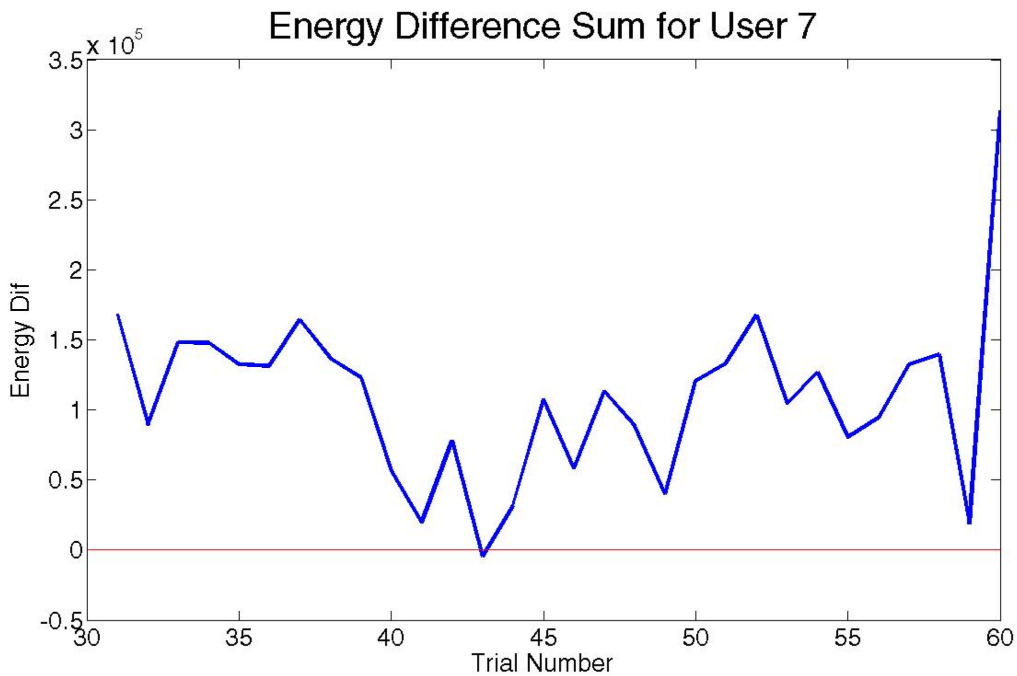

Figure 7.5. Energy difference between trials 31 to 60 and trial 30 for user 7.

Figure 7.5 shows the result of one user of the application. The graph specifies that the user energy level in most cases is above zero. This indicates that for user 7, the sum of energy of trial 30 was greater than the sum of energy of most trials afterwards till the end of the experiment. Relating this result to the results depicted in Figure 7.2, we observe that user 7 indeed indicated that he is experiencing more fatigue at the end of the experiment rather than in the middle of the experiment.

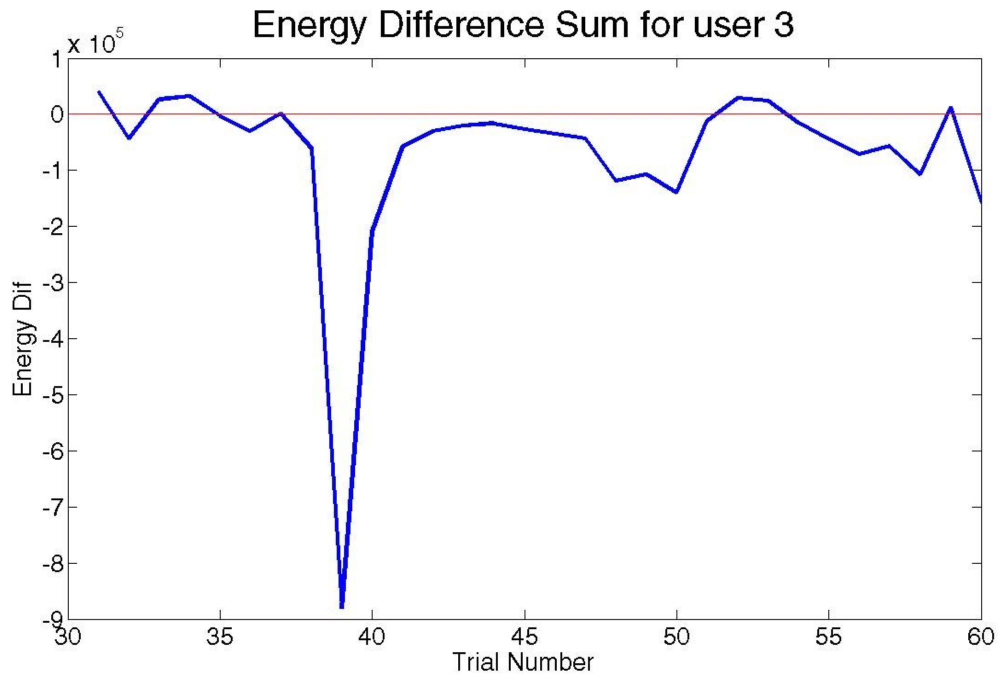

Figure 7.6. Energy difference between trials 31 to 60 and trial 30 for user 3.

Figure 7.6 shows the result of another user of the application who reported no difference in his perceived fatigue level (see Figure 7.2). User 3 energy differences between trial 30 and subsequent trials revolve around zero and dip into the negative in some cases. This means that his kinetic energy did not decrease overall and in some cases increased, which explains why he did not feel more tired by the end of the experiment.

Based on that analysis above we get 73.3% of the users who comply with that assumption. Their kinetic energy difference was in the positive compared to the base trial when they reported feeling fatigue by the end of the experiment, or their kinetic energy difference compared to the base trial was in the negative when they reported no change in fatigue between the middle and the end of the experiment.

7.4 Force Base Analysis

Velocity is a unique signature for the user and can be utilized to indicate the level of the energy as shown in the previous section. Nonetheless, the velocity of the user is not exclusive for haptic devices as velocity can be acquired with other types of interfaces. There are other attributes that we have collected from users that would be exclusive for haptic devices. Among these attributes are forces in the x, y, and z direction as described in the experimentation procedure. This section discusses the relation between force attributes and fatigue. We start by creating force profiles for the users.

7.4.1 Force Profiling

During the user study, we have collected huge amounts of data at frequent time stamps. Analyzing forces among these untreated data would be cumbersome. The graph of Figure 7.7 shows the force results for user one.

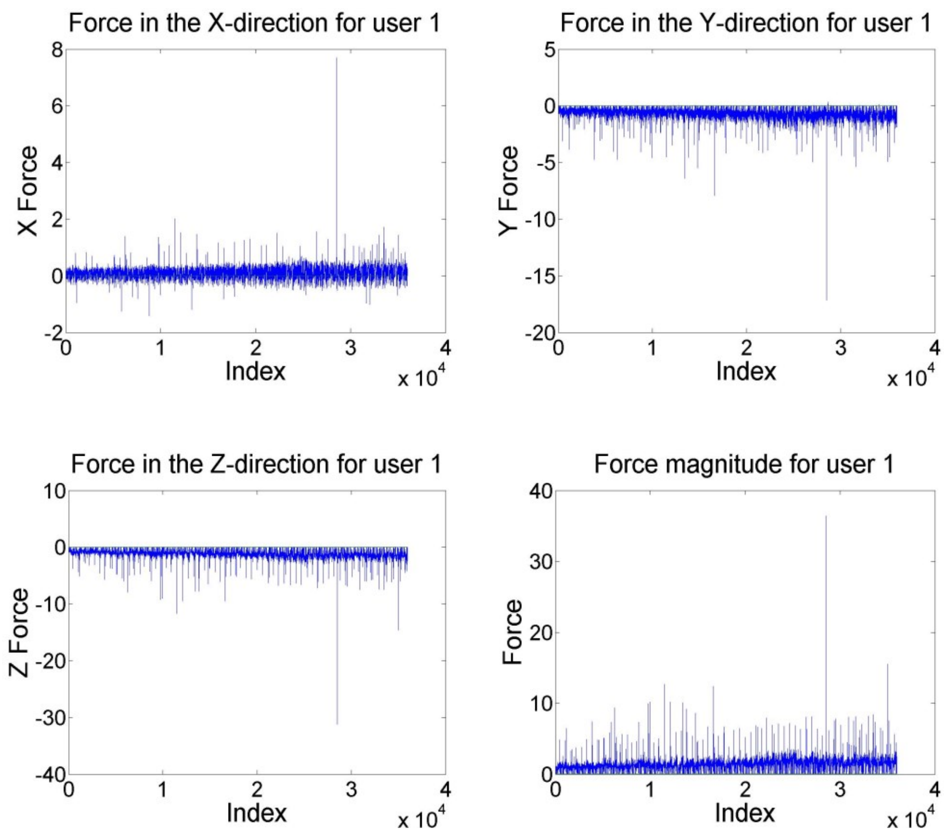

Figure 7.7. Force profile for user 1.

The figure displays the force profile for the given user which is the force magnitude and direction along the x, y, and z axis. It also shows the force magnitude for user 1 computed at each timestamp (F = √ (Fx2 + Fy2 + Fz2)). The force profile represents the force at all the timestamps from all the trials. There are close to 40,000 timestamps for user 1 as can be seen from the value of the x-axis in the figure.

7.4.1.1 Normalization and Quantization

The force profile provides a unique print for the user. However raw data need to be treated in order to find patterns and to be analyzed properly. In order to get a sense of the data we normalized the data then quantized it according to the following algorithm:

```
FOR EACH USER
   BUILD FORCE PROFILE MATRICES
   NORMALIZE DATA
   SAMPLESIZE = TOTAL_SIZE/40
   FOR ALL DATA
       SUM TOTAL FORCE MAG
       SUM FORCE X MAG
       SUM FORCE Y MAG
       SUM FORCE Z MAG
       IF CURRENTSAMPLENO MOD SAMPLESIZE IS ZERO
               AVERAGE RESULTS OVER THE SAMPLE SIZE
               STORE RESULTS
```

This way the results will be grouped and we will have 40 subsamples for each force direction for every user. This will provide a lower jitter in the data. We have used force magnitude in our algorithm since force can be positive or negative according to the direction along a certain access. With magnitude all the values are going to be positive and we can get a true average over the sample size. The results for user 1 are displayed in Figure 7.8.

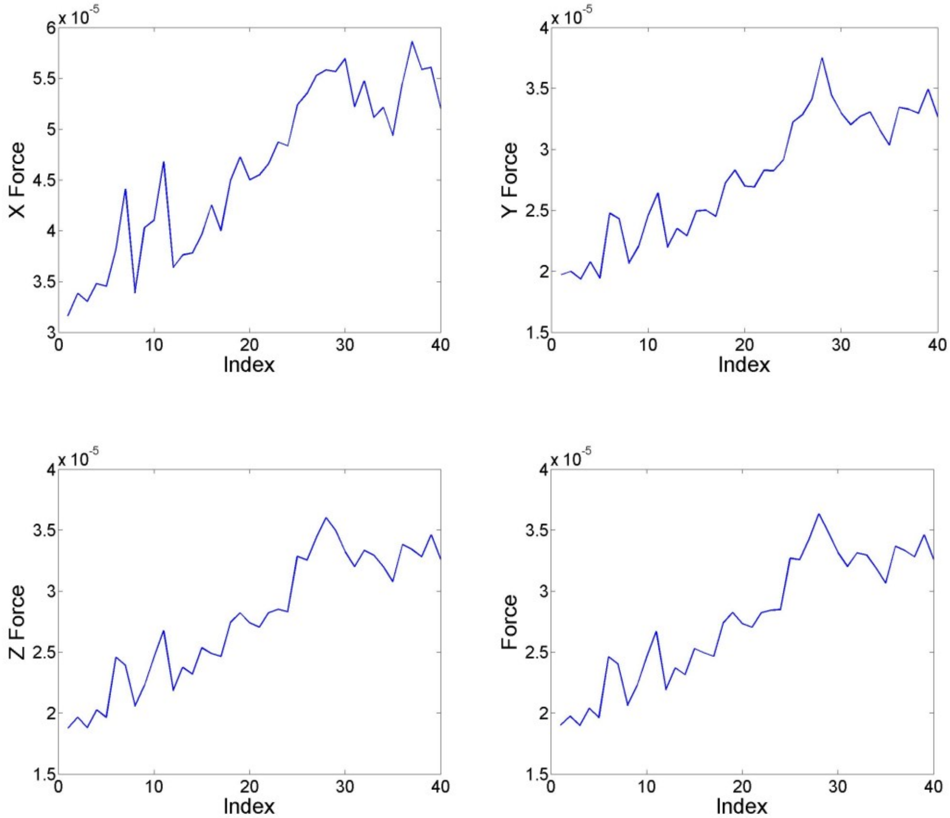

Figure 7.8. Normalized and quantized force profile for user 1.

7.4.1.2 Cumulative Force-Average Force-Work

Another addition to the force profile of the user is the cumulative force, average force and energy. For each user, the cumulative and average force magnitudes are calculated for each trial. Cumulative force is the total force by the unit time (timestamps) the user exerts in one trial. It was calculated by eq. (7.3). The magnitude of force can also be averaged for each trial which is the average force of the user.

$$CF= \sum(\text{force} * \Delta \text{ time}) \tag{7.3}$$

Work and kinetic energy are related according to the following equation:

$$W = E_{kf} - E_{ki} = \int F.dx \tag{7.4}$$

where E_{kf} is the final kinetic energy E_{ki} is the initial kinetic energy, F is the force, and dx is the minute change in position.

Since we have recorded the position and force at small time stamps we can calculate the work for each trial. The work value will provide us with the average change in kinetic energy. For user 1, the graphs are presented in Figure 7.9.

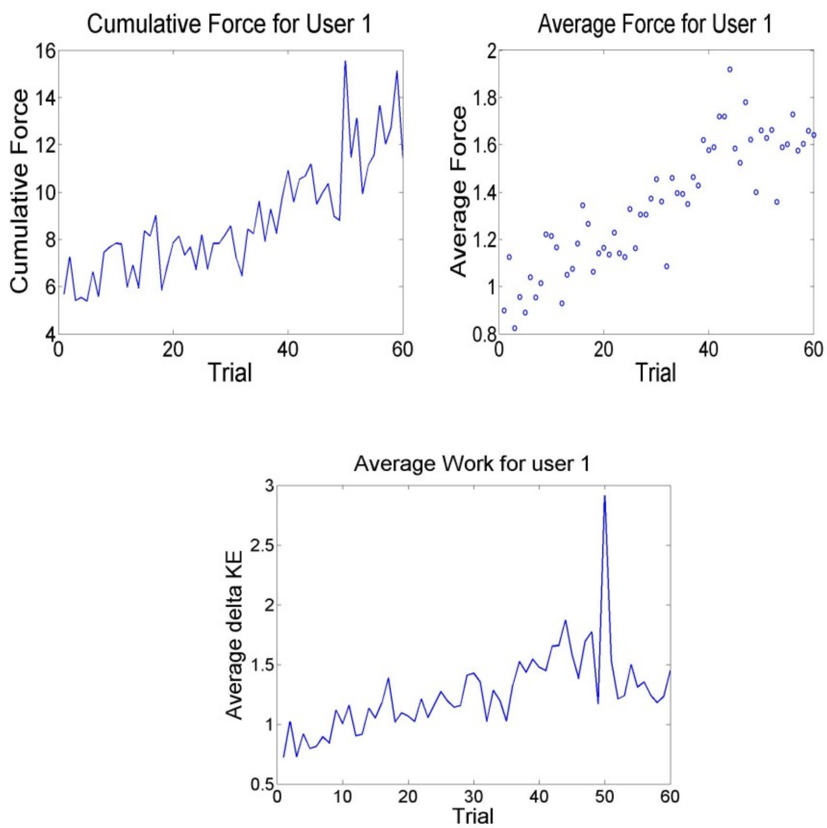

Figure 7.9. Cumulative force, average force, average work for user 1.

7.4.2 User Force Patterns

The analysis in this section is based on data collected after trial 30. In other words, trial 30 will be the base trial according to the TCT values in Section 7.2.5.

After building the force profile for each user, we wanted to establish the relation between that force profile and the user perceived fatigue. Since each user has a unique profile we can examine that profile against the user perceived level of fatigue.

7.4.2.1 Correlation Results

We have calculated correlation values for the quantized force profile of the users along with their subjective evaluation at the end of the experiment. The quantized forces values were based on the direction and magnitude of the forces. This was achieved by using linear regression to obtain the proper value. For example, the value representing quantized force in the x direction for some user would be the slope of the linear regression plot, represented by the red line in Figure 7.10.

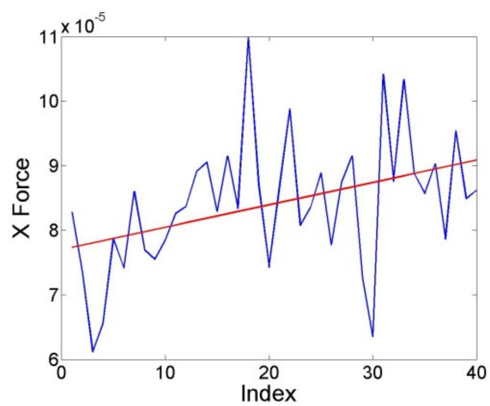

Figure 7.10. Linear regression coefficients displayed visually.

The slope coefficients of the linear regression of the users' profile were mapped with their subjective ratings through linear correlation (Table 7.1). The users' subjective rating at trial 60 was converted to normalized form by eq. (7.5) (Preston and Colman 2000) to calculate the correlation:

$$(\text{rating} - 1) / (\text{number of response categories} - 1) \times 100 \qquad (7.5)$$

The correlation value in bold indicates a significant correlation ($p<0.05$). The results suggest that only cumulative force and quantized force in the x direction do not correlate significantly with users' ratings. The negative correlation value indicates an inverse relationship between the subjective ratings and the force profile element value. The higher the users are rating their perceived fatigue the lower are their force element values, which indicates that users perceiving fatigue will reduce their force elements accompanying their force profile, due to that perceived fatigue.

TABLE 7.1- FORCE PROFILE AND SUBJECTIVE RATINGS CORRELATION

Force Profile Element	Correlation Value
Cumulative Force	-0.161
Average Force	**-0.583**
Work	**-0.466**
Quantized Force	**-0.487**
Quantized X Force	-0.379
Quantized Y Force	**-0.517**
Quantized Z Force	**-0.481**

The correlation values in bold are significant correlations ($p<0.05$)

7.4.2.2 Visual trends in the force profile

The trends in each user force profile are consistent for that particular user. For example if a user has an increasing trend in quantized total force, then his quantized forces in the x, y, and z direction would also have an increasing trend. Figure 7.11 represents this visually for user 1.

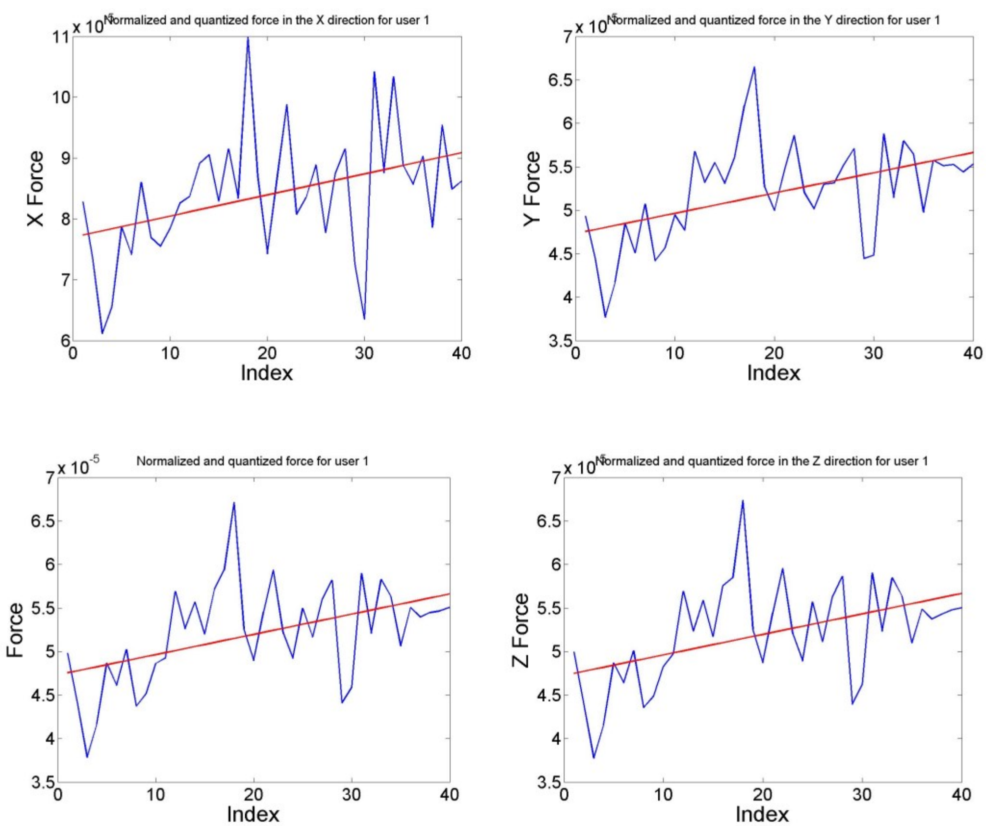

Figure 7.11. Trends in quantized force profile for user 1.

Beside the incremental trend of all quantized forces of user 1, it can be noticed that the rate of increase looks visually similar. Please note that Figure 7.11 is different from Figure 7.8 although both represent the quantized elements of user 1. As mentioned before, all the analysis and visual trends were conducted with trial 30 being the base, so the force profile was rebuilt for each user starting with that trial.

Other interesting visual trend discovered from users force profiles is that cumulative force has a negative (decreasing) trend for almost all users. Average force and user's work follow suit with quantized force elements for each user.

7.4.2.3 Foreseeing user subjective rating from their work profile

Since work represents the change in kinetic energy (eq. (7.4)) the higher the work value the higher is the change in energy. Higher work values indicate larger force or higher distance values which leads to more expenditure in kinetic energy. We therefore expect that users with lower perceived fatigue to apply more forces as trials increases because they are not experiencing fatigue according to their own self-evaluation.

The results from the data reinforce our expectations. There is a downward trend of the work value as the perceived fatigue rating at trial 60 increases (Figure 7.12). The work value is the linear regression slope obtained from the users' profile.

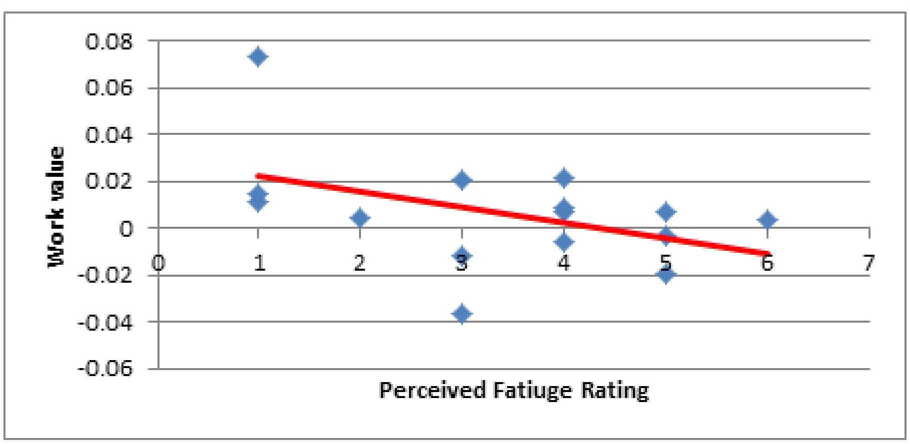

Figure 7.12. Work value vs. perceived fatigue rating.

7.5 Conclusion

This chapter presented a way of detecting fatigue for haptic-based applications. Users were asked to perform signing a virtual cheque 60 times using the haptic device. Their data were recorded at numerous timestamps throughout the experiment. Fatigue was detected by calculating the variation in user's energy in a time interval based on the velocity of the user. The magnitude of the variation is out of scope of this chapter but should be addressed in the future.

The chapter also presented trends in users' force profile generated from repetitive tasks using the Phantom Desktop haptic device. The user force profile is constituted from cumulative force, average force, quantized force (total force and x, y, and z directional forces), and average work. Each user profile is unique. From the users' profile we were able to infer certain trends and visualize them through linear regression analysis.

This research serves as a building block in evaluating the QoE of haptic-based applications. Fatigue is an important quality parameter and by mapping perceived fatigue with energy and force profile elements we will be able to use these findings in a QoE evaluation engine in the future. One limitation that could be generated with the fatigue examination is boredom. Boredom can affect results of the experiment if not controlled properly. Ways to detect and control boredom should be addressed in the future.

Undesired fatigue reduction can increase the QoE of a given application. Ways to reduce fatigue resulting from haptic devices is an open area of research and should be considered for future work. One suggestion is to use armrests but that is not always feasible depending on the application. Another suggestion is to use rest intermissions during usage of the haptic device but again that is application dependent.

CHAPTER 8

CONCLUSION AND FUTURE DIRECTIONS

8.1 Concluding Remarks

In this thesis we have discussed various topics related to QoE and VR applications. QoE is a relatively novel topic with new work emerging constantly. We have introduced the topic and given a detailed background on QoE in multimedia applications especially in the VR domain. The background consisted of QoE definition, relation between QoS and QoE, and QoE in the field of VR. Moreover QoE outside the VR domain was introduced briefly to give the reader a sense of work done in other areas.

In Chapter 3, we classify parameters that describe a VR environment. Those parameters are the bricks for any QoE evaluation engine. The parameters are numerous and are organized into a hierarchy. The top level of the hierarchy is the adopted definition of the QoE which is the QoS and the UX. The categories under the hierarchy branch out, and parameters are systemized according to the subcategories. We introduce a case study that uses parameters from the classification above to fuel our QoE evaluation engine. The study is conducted according to a specific protocol, using a haptic-based multimedia application.

There are two types of QoE evaluation paradigms adopted: mathematical modeling evaluation and fuzzy logic evaluation. In Chapter 4 we describe the mathematical modeling paradigm. With mathematical modeling a straightforward weighted average method is assumed. The mathematical model deals with the QoE according to the user study and the parameters presented in the previous chapter. Thus the equations of the mathematical model are broken down similar to the hierarchy of the classification. The determination of weights involved was a challenge and different approaches were conducted for that particular task. The different approaches performed similarly with some approaches being more accurately than others. Our overall observation, however, is that modeling human behaviour by a mathematical model will be limited by the nonlinearity of the human behavior. Whichever approach is utilized to determine the parameters and weights of the mathematical model, there will be a certain range of error expected due to the shortcomings of the mathematical model.

In light of the previous observation, and to tackle the shortcoming of the mathematical model we introduce the fuzzy logic paradigm in Chapter 5. We built the fuzzy logic inference system to assess the QoE from the subjective input parameters of the case study presented earlier. Fuzzy logic was needed because the mathematical model limitations can be handled by the fuzzy logic system, especially the nonlinearity of the input and output. Our fuzzy logic engine produced more accurate results than the mathematical model.

Chapter 6 of this thesis focused on a very important media in VR applications, namely haptics. The chapter details the difference and similarity of the effect of both types of haptic technology, kinesthetic feedback and tactile feedback, on the QoE. Haptic technology provides the user with additional perception through the sense of touch. The advancement of the human touch and the bidirectional exchange of forces between the user and the haptic interface create a complex and stimulating experience for the user. Thus the goal of the chapter was to evaluate QoE of the user when haptic interfaces replace input methods that we are accustomed to, such as the mouse and keyboard. The idea was to assess what the haptic interface provides for the user. It was found that kinesthetic and tactile haptics elevate the QoE of application in the absence of fatigue.

The high correlation of fatigue with the QoE involving haptic interfaces paved the way for Chapter 7. In this chapter we conduct a detailed study on ways to detect fatigue from the user haptic data. This special fatigue investigation enables the detection of fatigue without involving the user directly. The haptic data is the leading source for indicating whether the user is experiencing fatigue. By involving the user in repetitive tasks, there were different types of data that could be used in this research. We started by modeling the energy of the user from the velocity data of the user to indicate fatigue. The energy model holds true given that the mass specific to a certain user remains constant in the case of repetitive task. Another type of data which is exclusive to haptic devices is the force of the user. We have built force profiles for each user and deduced patterns indicating fatigue from the force profile. The validation of both types of models, velocity and force, was done by subjective measures of the user.

8.2 Possibility of Future Work

The thesis tackles obstacles that relate to the QoE of VR applications. The human behavior is complex and vast. We try to objectify this behavior by setting equations and rules under given circumstances. There is always room for improvement.

The taxonomy could be rearranged to divide the user state into three categories: perception measures, psychological measures, and physiological measures. In this case, perception measures and psychological measures would be merged into one subcategory as a subjective user state, while physiological measures would be the biological indicator of the user state. This would depend on the advancement of the psychophysiological field in VR that would map the biological indicators with the subjective ones.

Different methodologies were conducted to reduce the error generated by the mathematical model paradigm; among these methodologies is linear regression which produced the highest error rates. Studying the behavior of the data, we assumed a linear regression fit between the input (parameter evaluation) and the QoE output of the user. Human behavior is not linear, but this is the best fit that the data provided, and the most convenient given the linear relationship of the weighted average methodology. Other forms of regression analysis might be performed in the future, while comparing the results to the linear regression fit.

In the fuzzy logic evaluation system the rules were generated from the data of the user study. Expert rule generation will provide an alternate strategy to feed the fuzzy inference engine with a set of rules and standards to calculate the QoE of an application. Expert rule generation is application dependent and requires stringent results monitoring at the beginning of the process because it is error prone given the human error tendencies of the experts. However if this methodology is proven efficient, then it can save cost and time in the future by building an FIS faster.

The fatigue analysis showed ways to detect fatigue while utilizing a haptic device. The fatigue magnitude generated from velocity or force profile is left for future work.

Moreover, ways to reduce fatigue during the usage of haptic applications is still desirable to investigate.

Future direction can lead to automation of the QoE evaluation, thus eliminating the need for the user. This would be done in steps, such as the fatigue investigation we did in this thesis. Other parameters can be standardized and detected as well from the haptic data.

REFERENCES

Abdur Rahman, M., Alkhaldi, A., Cha, J., & El Saddik, A. (2010). Adding haptic feature to YouTube. *Proceedings of the International Conference on Multimedia,* Firenze, Italy. 1643-1646.

Abu-Tair, M., Marshall, A., & Arroniz, B. M. (2011). Performance evaluation of QoS technique for supporting multiple contact point haptic device interactions. *Consumer Electronics (ICCE), 2011 IEEE International Conference on,* 915-916.

Al Osman, H., Eid, M., & El Saddik, A. (2008). Evaluating ALPHAN: A network protocol for haptic interaction. *The 16th Symposium on Haptic Interfaces for Virtual Environments and Teleoperator Systems,* Reno, Nevada, USA.

Alamri, A., Eid, M., & El Saddik, A. (2007). A haptic enabled UML case tool. *Multimedia and Expo, 2007 IEEE International Conference on,* 1023-1026.

Alamri, A., Eid, M., & Saddik, A. E. (2008). A quality of performance model for evaluating post-stroke patients. *Computational Intelligence for Measurement Systems and Applications, 2008. CIMSA 2008. 2008 IEEE International Conference on,* 14-18.

Alsulaiman, F., Cha, J., & El Saddik, A. (2008). User identification based on handwritten signatures with haptic information. 114-121.

Andersson, U., Lyxell, B., Rönnberg, J., & Spens, K. E. (2001). Effects of tactile training on visual speechreading: Performance changes related to individual differences in cognitive skills. *Journal of Deaf Studies and Deaf Education, 6*(2), 116.

Aron, A., & Aron, E. N. (2003). *Statistics for psychology* (3rd ed. ed.) Prentice Hall/Pearson Education.

Asghar, J., Le Faucheur, F., & Hood, I. (2009). Preserving video quality in IPTV networks. *Broadcasting, IEEE Transactions on, 55*(2), 386-395.

Barry, B. K., & Enoka, R. M. (2007). The neurobiology of muscle fatigue: 15 years later. *Integrative and Comparative Biology, 47*(4), 465-473.

Basdogan, C., Ho, C. H., Srinivasan, M. A., & Slater, M. (2000). An experimental study on the role of touch in shared virtual environments. *ACM Transactions on Computer-Human Interaction (TOCHI), 7*(4), 443-460.

Beauregard, R., & Corriveau, P. (2007). User experience quality: A conceptual framework for goal setting and measurement. *Proceedings of the 1st International Conference on Digital Human Modeling,* 325-332.

Behr, K., Nosper, A., Klimmt, C., & Hartmann, T. (2005). Some practical considerations of ethical issues in VR research. *Presence: Teleoperators & Virtual Environments, 14*(6), 668-676.

Bernardo, F., Vucevic, N., Umbert, A., & Lopez-Benitez, M. (2008). Quality of experience evaluation under QoS-aware mobility mechanisms. *Wireless Conference, 2008. EW 2008. 14th European,* 1-7.

Beurskens, A. J. H. M., Bültmann, U., Kant, I. J., Vercoulen, J. H. M. M., Bleijenberg, G., & Swaen, G. M. H. (2000). Fatigue among working people: Validity of a questionnaire measure. *Occupational and Environmental Medicine, 57*(5), 353.

Black Ken. (c2004.). *Business statistics: for contemporary decision making /* (4th ed.) Leyh Publishing.

Bo Wang, Xiangmin Wen, Sun Yong, & Zheng Wei. (2009). A new approach measuring users' QoE in the IPTV. *Circuits, Communications and Systems, 2009. PACCS '09. Pacific-Asia Conference on,* 453-456.

Bolas, M. T. (1994). Human factors in the design of an immersive display. *Computer Graphics and Applications, IEEE, 14*(1), 55-59.

Bouch, A., & Sasse, M. A. (1999). Network quality of service: What do users need. *Proceedings of the 4th International Distributed Conference.*

Brown, E., & Cairns, P. (2004). A grounded investigation of game immersion. *CHI '04 Extended Abstracts on Human Factors in Computing Systems,* Vienna, Austria. 1297-1300.

Burdea, G., & Coiffet, P. (2003). Virtual reality technology. *Presence: Teleoperators and Virtual Environments, 12*(6), 663-664.

Chen, C., Czerwinski, M., & Macredie, R. (1998). Human factors in virtual environments. *Virtual Reality, 3*(4), 223-225.

Chen, J., & Terrence, P. (2008). Effects of tactile cueing on concurrent performance of military and robotics tasks in a simulated multitasking environment. *Ergonomics, 51*(8), 1137-1152.

Coles, T. R., Meglan, D., & John, N. W. (2011). The role of haptics in medical training simulators: A survey of the state of the art. *Haptics, IEEE Transactions on, 4*(1), 51-66.

Cribbie, R. A., Gruman, J. A., & Arpin - Cribbie, C. A. (2004). Recommendations for applying tests of equivalence. *Journal of Clinical Psychology, 60*(1), 1-10.

De Marez, L., & De Moor, K. (2007). The challenge of user-and QoE-centric research and product development in today's ICT-environment. *Observatorio (OBS*), 1*(3).

Deml, B. (2007). Human factors issues on the design of telepresence systems. *Presence: Teleoperators and Virtual Environments, 16*(5), 471-487.

Durlach, N., & Slater, M. (2000). Presence in shared virtual environments and virtual togetherness. *Presence: Teleoperators and Virtual Environments, 9*(2), 214-217.

Ebrahimi, T. (2009). Quality of multimedia experience: Past, present and future. *Proceedings of the 17th ACM International Conference on Multimedia,* Beijing, China. 3-4.

Eid, M., Orozco, M., & El Saddik, A. (2007). A guided tour in haptic audio visual environments and applications. *International Journal of Advanced Media and Communication, 1*(3), 265-297.

Eid, M. A., Mansour, M., Saddik, A. H. E., & Iglesias, R. (2007). A haptic multimedia handwriting learning system. *Emme '07: Proceedings of the International Workshop on Educational Multimedia and Multimedia Education,* Augsburg, Bavaria, Germany. 103-108.

Eid, M., Jongeun Cha, & El Saddik, A. (2008). HugMe: A haptic videoconferencing system for interpersonal communication. *Virtual Environments, Human-Computer Interfaces and Measurement Systems, 2008. VECIMS 2008. IEEE Conference on,* pp. 5-9.

El Saddik. (2007). The potential of haptics technologies. *Instrumentation & Measurement Magazine, IEEE, 10*(1), 10-17.

El Saddik, A., Orozco, M., Eid, M., & Cha, J. (2011). Haptics technologies: Bringing touch to multimedia. Springer.

Ellis, R., Ismaeil, O., & Lipsett, M. (1996). Design and evaluation of a high-performance haptic interface. *Robotica, 14*(03), 321-327.

Enoka, R. M., & Stuart, D. G. (1992). Neurobiology of muscle fatigue. *Journal of Applied Physiology, 72*(5), 1631-1648.

Fiedler, M., Hossfeld, T., & Phuoc Tran-Gia. (2010). A generic quantitative relationship between quality of experience and quality of service. *Network, IEEE, 24*(2), 36-41.

Fiedler, M., Kilkki, K., & Reichl, P. (2009). 09192 abstracts collection -- from quality of service to quality of experience. *From Quality of Service to Quality of Experience,* (09192).

Fiedler, M., Kilkki, K., & Reichl, P. (2009). 09192 executive summary -- from quality of service to quality of experience. *From Quality of Service to Quality of Experience,* (09192).

Flach, J. M., & Holden, J. G. (1998). The reality of experience: Gibson's way. *Presence: Teleoperators and Virtual Environments, 7*(1), 90-95.

Forlizzi, J., & Battarbee, K. (2004). Understanding experience in interactive systems. *Proceedings of the 5th Conference on Designing Interactive Systems: Processes, Practices, Methods, and Techniques,* Cambridge, MA, USA. 261-268.

Fu, F. L., Su, R. C., & Yu, S. C. (2009). EGameFlow: A scale to measure learners' enjoyment of e-learning games. *Computers & Education, 52*(1), 101-112.

Gabbard, J. L., Hix, D., & Swan, J. E.,II. (1999). User-centered design and evaluation of virtual environments. *Computer Graphics and Applications, IEEE, 19*(6), 51-59.

Gaggioli, A., Bassi, M., & Fave, A. (2003). Quality of experience in virtual environments. *Emergin Ccommuniction, 5*, 121-136.

Gates, D. H., & Dingwell, J. B. (2008). The effects of neuromuscular fatigue on task performance during repetitive goal-directed movements. *Experimental Brain Research, 187*(4), 573-585.

Greengrass, J., Evans, J., & Begen, A. C. (2009a). Not all packets are equal, part I: Streaming video coding and SLA requirements. *Internet Computing, IEEE, 13*(1), 70-75.

Greengrass, J., Evans, J., & Begen, A. C. (2009b). Not all packets are equal, part 2: The impact of network packet loss on video quality. *Internet Computing, IEEE, 13*(2), 74-82.

Greenway, R. B., & Faddis, T. N. (1993). Effects of tactile information on telerobotic performance. *Proc. SPIE, 1833*, 317-324.

Grigore, B., & Coiffet, P. (2003). *Virtual reality technology* (2nd edition ed.).

Guerraz, A., Loscos, C., & Ritter Widenfeld, H. (2006). *How to use physical parameters coming from the haptic device itself to enhance the evaluation of haptic benefits in user interface?* HAL - CCSD.

Gutierrez, F., Pierce, J., Vergara, V. M., Coulter, R., Saland, L., Caudell, T. P., et al. (2007). The effect of degree of immersion upon learning performance in virtual reality simulations for medical education. *Studies in Health Technology and Informatics, 125*, 155-160.

Gwilliam, J. C., Mahvash, M., Vagvolgyi, B., Vacharat, A., Yuh, D. D., & Okamura, A. M. (2009). Effects of haptic and graphical force feedback on teleoperated palpation. *Robotics and Automation, 2009. ICRA '09. IEEE International Conference on,* pp. 677-682.

Haiqing Du, Chang Guo, Yixi Liu, & Yong Liu. (2009). Research on relationship between QoE and QoS based on BP neural network. *Network Infrastructure and Digital Content, 2009. IC-NIDC 2009. IEEE International Conference on,* 312-315.

Hajshirmohammadi, I., & Payandeh, S. (2007). Fuzzy set theory for performance evaluation in a surgical simulator. *Presence: Teleoperators and Virtual Environments, 16*(6), 603-622.

Hamam, A., Eid, M., El Saddik, A., & Georganas, N. D. (2008a). A quality of experience model for haptic user interfaces. *HAS '08: Proceedings of the 2008 Ambi-Sys Workshop on Haptic User Interfaces in Ambient Media Systems,* Quebec City, Canada. 1-6.

Hamam, A., Eid, M., El Saddik, A., & Georganas, N. D. (2008b). A fuzzy logic system for evaluating quality of experience of haptic-based applications. *Eurohaptics,* Madrid. *, 5024* 129-138.

Hamam, A., & Georganas, N. D. (2008c). A comparison of mamdani and sugeno fuzzy inference systems for evaluating the quality of experience of hapto-audio-visual applications, *HAVE'08.*

Hamam, A., Georganas, N. D., & El Saddik, A. (2010). Effect of haptics on the quality of experience. *Haptic Audio-Visual Environments and Games (HAVE), 2010 IEEE International Symposium on,* 1-6.

Hamam, A., Georganas, N. D., Alsulaiman, F., & El Saddik, A. (2010). Deducing user's fatigue from haptic data. *Proceedings of the International Conference on Multimedia,* Firenze, Italy. 1083-1086.

Heikkinen, J., Olsson, T., & Vaananen-Vainio-Mattila, K. (2009). Expectations for user experience in haptic communication with mobile devices. *Proceedings of the 11th*

International Conference on Human-Computer Interaction with Mobile Devices and Services, Bonn, Germany. 28:1-28:10.

Hestnes, B., Brooks, P., Heiestad, S., Ulseth, T., & Aaby, C. (2003). Quality of experience in real-time person-person communication-user based QoS expressed in technical network QoS terms. *19th International Symposium on Human Factors in Telecommunication (HFT03),* 1-4.

Hoffman, H. G., Hollander, A., Schroder, K., Rousseau, S., & Furness, T. (1998). Physically touching and tasting virtual objects enhances the realism of virtual experiences. *Virtual Reality, 3*(4), 226-234.

Hoßfeld, T., Tran-Gia, P., & Fiedler, M. (2007). Quantification of quality of experience for edge-based applications. *Proceedings of the 20th International Teletraffic Conference on Managing Traffic Performance in Converged Networks,* 361-373.

Huadong W., Siegel, M., & Khosla, P. (1999). Vehicle sound signature recognition by frequency vector principal component analysis. *Instrumentation and Measurement, IEEE Transactions on, 48*(5), 1005-1009.

Hyun J.K., Byung K.K., Hyun H.C., & Seong G.C. (2010). Implementation of QoS control system with QoE parameters on multimedia services. *Advanced Communication Technology (ICACT), 2010 the 12th International Conference on,* 1035-1040.

Hyun J.K., Dong H.L., Jong M.L., Kyoung H.L., Won L., & Seong G.C. (2008). The QoE evaluation method through the QoS-QoE correlation model. *Networked Computing and Advanced Information Management, 2008. NCM '08. Fourth International Conference on, , 2* 719-725.

Hyun J.K., & Seong G. C. (2010). A study on a QoS/QoE correlation model for QoE evaluation on IPTV service. *Advanced Communication Technology (ICACT), 2010 the 12th International Conference on,* 1377-1382.

Ida, Y., Ishibashi, Y., Fukushima, N., & Sugawara, S. (2010). QoE assessment of interactivity and fairness in first person shooting with group synchronization control. *Network and Systems Support for Games (NetGames), 2010 9th Annual Workshop on,* 1-2.

IJsselsteijn, W., de Kort, Y., Poels, K., Jurgelionis, A., & Bellotti, F. (2007). Characterising and measuring user experiences in digital games. *International Conference on Advances in Computer Entertainment Technology.*

Insko, B. E. (2003). Measuring presence: Subjective, behavioral and physiological methods. *EMERGING COMMUNICATION, 5*, 109-120.

Ito, S., & Yokokohji, Y. (2009). Maneuverability of master control devices considering the musculo-skeletal model of an operator. *EuroHaptics Conference, 2009 and Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems. World Haptics 2009. Third Joint,* 57-62.

Ito, Y., & Tasaka, S. (2005). Quantitative assessment of user-level QoS and its mapping. *Multimedia, IEEE Transactions on, 7*(3), 572-584.

ITU-T Rec. (2006). P. 800.1 "Mean opinion score (MOS) terminology ". *International Telecommunication Union, Geneva.*

ITU-T Rec. (2008). P. 10/G. 100 (incl. amendment 2) "Vocabulary for Performance and Quality of Service", *International Telecommunication Union, Geneva.*

Iwata, K., Ishibashi, Y., Fukushima, N., & Sugawara, S. (2010). QoE assessment in haptic media, sound, and video transmission: Effect of playout buffering control. *Comput.Entertain., 8*(2), 12:1-12:14.

Jain, R. (2004). Quality of experience. *Multimedia, IEEE, 11*(1), 96-95.

Jandura, L., & Srinivasan, M. A. (1994). Experiments on human performance in torque discrimination and control. *Dynamic Systems and Control, ASME, DSC-55, 1*, 369-375.

Jassbi, J. J., Serra, P. J. A., Ribeiro, R. A., & Donati, A. (2006). A comparison of mandani and sugeno inference systems for a space fault detection application. *World Automation Congress, 2006. WAC'06,* 1-8.

Jassbi, J., Alavi, S. H., Serra, P. J. A., & Ribeiro, R. A. (2007). Transformation of a mamdani FIS to first order sugeno FIS. *Fuzzy Systems Conference, 2007. FUZZ-IEEE 2007. IEEE International,* 1-6.

Jennett, C., Cox, A. L., Cairns, P., Dhoparee, S., Epps, A., Tijs, T., et al. (2008). Measuring and defining the experience of immersion in games. *International Journal of Human-Computer Studies, 66*(9), 641-661.

Kahol, K., Leyba, M. J., Deka, M., Deka, V., Mayes, S., Smith, M., et al. (2008). Effect of fatigue on psychomotor and cognitive skills. *The American Journal of Surgery, 195*(2), 195-204.

Khirman, S., & Henriksen, P. (2002). Relationship between quality-of-service and quality-of-experience for public internet service. *In Proc. of the 3rd Workshop on Passive and Active Measurement.*

King, C., Culjat, M. O., Franco, M. L., Lewis, C. E., Dutson, E. P., Grundfest, W. S., et al. (2009). Tactile feedback induces reduced grasping force in robot-assisted surgery. *Haptics, IEEE Transactions on, 2*(2), 103-110.

Kondo, M., Ueda, J., Matsumoto, Y., & Ogasawara, T. (2005). Evaluation of manipulative familiarization and fatigue based on contact state transition. *Eurohaptics Conference, 2005 and Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems, 2005. World Haptics 2005. First Joint,* 102-107.

Kumar, K. (2005). A marriage made in QoE heaven. *CED Magazine,* 37-39.

Kusunose, Y., Ishibashi, Y., Fukushima, N., & Sugawara, S. (2010). QoE assessment in networked air hockey game with haptic media. *Network and Systems Support for Games (NetGames), 2010 9th Annual Workshop on,* 1-2.

Kyoung S.P., & Kenyon, R. V. (1999). *Effects of network characteristics on human performance in a collaborative virtual environment*. Proceeding of Virtual Reality, 104-111.

Law, E. L., Roto, V., Hassenzahl, M., Vermeeren, A. P. O. S., & Kort, J. (2009). Understanding, scoping and defining user experience: A survey approach. *Proceedings of the 27th*

International Conference on Human Factors in Computing Systems, Boston, MA, USA. 719-728.

Law, E., Roto, V., Vermeeren, A. P. O. S., Kort, J., & Hassenzahl, M. (2008). Towards a shared definition of user experience. *CHI '08 Extended Abstracts on Human Factors in Computing Systems,* Florence, Italy. 2395-2398.

Lee, K. A., Hicks, G., & Nino-Murcia, G. (1991). Validity and reliability of a scale to assess fatigue. *Psychiatry Research, 36*(3), 291-298.

Lewis, C. H., & Griffin, M. J. (1997). Human factors consideration in clinical applications of virtual reality. *Virtual Reality in Neuro-Psycho-Physiology Cognitive, Clinical and Methodological Issues in Assessment and Rehabilitation, 44*.

Lin, J. J. W., & Parker, D. E. (2007). User experience modeling and enhancement for virtual environments that employ wide-field displays. *Proceedings of the 1st International Conference on Digital Human Modeling,* Beijing, China. 423-433.

Lou, J. S., Kearns, G., Oken, B., Sexton, G., & Nutt, J. (2001). Exacerbated physical fatigue and mental fatigue in parkinson's disease. *Movement Disorders, 16*(2), 190-196.

Ma, L., Chablat, D., Bennis, F., Zhang, W., Hu, B., & Guillaume, F. (2010). Fatigue evaluation in maintenance and assembly operations by digital human simulation in virtual environment. *Virtual Reality,* 1-14.

Mahlke, S. (2005). Understanding users' experience of interaction. *Proceedings of the 2005 Annual Conference on European Association of Cognitive Ergonomics,* 251-254.

Mahlke, S. (2006). Studying user experience with digital audio players. *Entertainment Computing-ICEC 2006,* 358-361.

Mamdani, E. H., & Assilian, S. (1999). An Experiment in Linguistic Synthesis with a Fuzzy Logic Controller. *International Journal of Human-Computer Studies, 51*(2), 135-147.

Mandryk, R. L., & Atkins, M. S. (2007). A fuzzy physiological approach for continuously modeling emotion during interaction with play technologies. *International Journal of Human-Computer Studies, 65*(4), 329-347.

Mania, K., & Hawkes, R. (2003). Human factors in virtual environments. Technical report.

Mansour, M., Eid, M., & El-Saddik, A. (2007). A multimedia handwriting learning and evaluation tool. *International Workshop on Educational Multimedia and Multimedia Education,* 103-108.

Marshall, A., Kian Meng Yap, & Wai Yu. (2008). Quality of service issues for distributed virtual environments with haptic interfaces. *Multimedia Signal Processing, 2008 IEEE 10th Workshop on,* 40-45.

Massie, T. H. (1996). Initial haptic explorations with the phantom: Virtual touch through point interaction. Massachusetts Institute of Technology.

Matijasevic, M., Gracanin, D., Valavanis, K. P., & Lovrek, I. (2000). Towards QoS specification for distributed virtual environments. *Electrotechnical Conference, 2000. MELECON 2000. 10th Mediterranean,* 132-135.

McNamara, N., & Kirakowski, J. (2005). Defining usability: Quality of use or quality of experience? *Professional Communication Conference, 2005. IPCC 2005. Proceedings. International,* 200-204.

Meehan, M., Insko, B., Whitton, M., & Brooks, J.,Frederick P. (2002). Physiological measures of presence in stressful virtual environments. *Proceedings of the 29th Annual Conference on Computer Graphics and Interactive Techniques,* San Antonio, Texas. 645-652.

Meitzler, T. J., & Sohn, E. (2005). A comparison of mamdani and sugeno methods for modeling visual perception lab data. *Fuzzy Information Processing Society, 2005. NAFIPS 2005. Annual Meeting of the North American,* 116-120.

Microsoft Inc. (2007). *Quality of experience: A strategic competitive advantage of microsoft unified communications*. White paper.

Miller, V. G. (1993). Measurement of self-perception of intuitiveness. *Western Journal of Nursing Research, 15*(5), 595.

Moller, S., Engelbrecht, K. , Kuhnel, C., Wechsung, I., & Weiss, B. (2009). A taxonomy of quality of service and quality of experience of multimodal human-machine interaction. *Quality of Multimedia Experience, 2009. QoMEx 2009. International Workshop on,* 7-12.

Monin, B., & Oppenheimer, D. M. (2005). Correlated averages vs. averaged correlations: Demonstrating the warm glow heuristic beyond aggregation. *Social Cognition, 23*(3), 257-278.

Montgomery, D. C., Peck, E. A., & Vining, G. G. (2012). *Introduction to linear regression analysis*, Wiley.

Mukherjea, S. (1998). Using VRML-based visualisations to facilitate information retrieval in the world wide web. *Virtual Reality, 3*(4), 245-258.

Muntean, C. H., Muntean, G. M., McManis, J., & Cristea, A. I. (2007). Quality of experience-LAOS: Create once, use many, use anywhere. *International Journal of Learning Technology, 3*(3), 209-229.

Nacke, L., & Lindley, C. A. (2008). Flow and immersion in first-person shooters: Measuring the player's gameplay experience. *Proceedings of the 2008 Conference on Future Play: Research, Play, Share,* 81-88.

Naud, M., Ullah, S., Richard, P., Otmane, S., & Mallem, M. (2009). Effect of tactile feedback and viewpoint on task performance in a collaborative virtual environmment.

Nichols, S., Haldane, C., & Wilson, J. R. (2000). Measurement of presence and its consequences in virtual environments. *International Journal of Human-Computer Studies, 52*(3), 471-491.

Nokia. (2006). Quality of experience (QoE) of mobile services: Can it be measured and improved? *White Paper.*

O'Brien, H. L., & MacLean, K. E. Measuring the user engagement process.

Okamoto, J., Watanabe, K., Honda, A., Uchida, M., & Hangai, S. (2009). HDTV objective video quality assessment method applying fuzzy measure. *International Workshop on Quality of Multimedia Experience, 2009. QoMEx 2009,* 168-173.

Pausch, R., Proffitt, D., & Williams, G. (1997). Quantifying immersion in virtual reality. *Proceedings of the 24th Annual Conference on Computer Graphics and Interactive Techniques,* 13-18.

Pawar, V. M., & Steed, A. (2009). Evaluating the influence of haptic force-feedback on 3D selection tasks using natural egocentric gestures. *Virtual Reality Conference, 2009. VR 2009. IEEE,* pp. 11-18.

Pongrac, H., Leupold, J., Behrendt, S., Färber, B., & Färber, G. (2007). Human factors for enhancing live video streams with virtual reality: Performance, situation awareness, and feeling of telepresence. *Presence: Teleoperators and Virtual Environments, 16*(5), 488-508.

Potkonjak, V., Kostic, D., Rasic, M., & Dordaevic, G. S. (2002). Motion in human and machine: A virtual fatigue approach. *Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on, 32*(5), 582-595.

Preston, C. C., & Colman, A. M. (2000). Optimal number of response categories in rating scales: Reliability, validity, discriminating power, and respondent preferences. *Acta Psychologica, 104*(1), 1-15.

Qiang Ji, Lan, P., & Looney, C. (2006). A probabilistic framework for modeling and real-time monitoring human fatigue. *Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on, 36*(5), 862-875.

Ramsey, A. (1997). Investigation of physiological measures relative to self-report of virtual reality induced sickness and effects (VRISE). *The International Workshop on Motion Sickness: Medical and Human Factors, Marbella, Spain, may,* Marbella, Spain. 26-28.

Reichl, P. (2007). From 'Quality-of-service' and 'Quality-of-design' to 'Quality-of-experience': A holistic view on future interactive telecommunication services. *Software, Telecommunications and Computer Networks, 2007. SoftCOM 2007. 15th International Conference on,* 1-6.

Reichl, P., Egger, S., Schatz, R., & D'Alconzo, A. (2010). The logarithmic nature of QoE and the role of the weber-fechner law in QoE assessment. *Communications (ICC), 2010 IEEE International Conference on,* 1-5.

Reiter, U. (2009). Perceived quality in consumer electronics - from quality of service to quality of experience. *Consumer Electronics, 2009. ISCE '09. IEEE 13th International Symposium on,* 958-961.

Robles-De-La-Torre, G. (2006). The importance of the sense of touch in virtual and real environments. *Multimedia, IEEE, 13*(3), 24-30.

Roid, G. H. (2004). Quality of performance and change-sensitive assessment of cognitive ability.

Rowe, L. A., & Jain, R. (2005). ACM SIGMM retreat report on future directions in multimedia research. *ACM Trans.Multimedia Comput.Commun.Appl., 1*(1), 3-13.

Sallnas, E., Rassmus-Gr\ohn, K., & Sj\ostr\om, C. (2000). Supporting presence in collaborative environments by haptic force feedback. *ACM Trans.Comput.-Hum.Interact., 7*(4), 461-476.

Sastry, L., & Boyd, D. R. S. (1998). Virtual environments for engineering applications. *Virtual Reality, 3*(4), 235-244.

Sauro, J., & Kindlund, E. (2005). A method to standardize usability metrics into a single score. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*.

Schubert, T., Friedmann, F., & Regenbrecht, H. (2001). The experience of presence: Factor analytic insights. *Presence: Teleoperators & Virtual Environments, 10*(3), 266-281.

Schuemie, M. J., Van Der Straaten, P., Krijn, M., & Van Der Mast, C. A. P. G. (2001). Research on presence in virtual reality: A survey. *CyberPsychology & Behavior, 4*(2), 183-201.

Segerstahl, K., & Oinas-Kukkonen, H. (2007). Distributed user experience in persuasive technology environments. *Proceedings of the 2nd International Conference on Persuasive Technology,* 80-91.

Seroussi, R., Krag, M. H., Wilder, P., & Pope, M. H. (1989). The design and use of a microcomputerized real-time muscle fatigue monitor based on the medical frequency shift in the electromyographic signal. *Biomedical Engineering, IEEE Transactions on, 36*(2), 284-286.

Siller, M., & Woods, J. (2003a). Improving quality of experience for multimedia services by QoS arbitration on a QoE framework. *In Proc. of the 13th Packed Video Workshop 2003.*

Siller, M., & Woods, J. C. (2003b). QoS arbitration for improving the QoE in multimedia transmission. *Visual Information Engineering, 2003. VIE 2003. International Conference on,* 238-241.

Slater, M., Linakis, V., Usoh, M., & Kooper, R. (1996). Immersion, presence, and performance in virtual environments: An experiment with tri-dimensional chess. *ACM Virtual Reality Software and Technology (VRST),* 163–172.

Slater, M. (1999). Measuring presence: A response to the witmer and singer presence questionnaire. *Presence: Teleoperators and Virtual Environments, 8*(5), 560-565.

Smith, C. M. (1997). Human factors in haptic interfaces. *Crossroads, 3*(3), 14-16.

Srinivasan, M. A., & Basdogan, C. (1997). Haptics in virtual environments: Taxonomy, research status, and challenges. *Computers & Graphics, 21*(4), 393-404.

Stanney, K. (1995). Realizing the full potential of virtual reality: Human factors issues that could stand in the way. *Virtual Reality Annual International Symposium, 1995.* 28-34.

Stanney, K. M., Mourant, R. R., & Kennedy, R. S. (1998). Human factors issues in virtual environments: A review of the literature. *Presence: Teleoperators and Virtual Environments, 7*(4), 327-351.

Streiner, D. L. (2003). Unicorns do exist: a tutorial on "proving" the null hypothesis. *Canadian Journal of Psychiatry.Revue Canadienne De Psychiatrie, 48*(11), 756-761.

Strickland, D., Hodges, L., North, M., & Weghorst, S. (1997). Overcoming phobias by virtual exposure. *Commun.ACM, 40*(8), 34-39.

Sweetser, P., & Wyeth, P. (2005). GameFlow: A model for evaluating player enjoyment in games. *Computers in Entertainment (CIE), 3*(3), 3-3.

Szabo, T., & Horvath, G. (1998). Finite word length computational effects of the principal component analysis networks. *Instrumentation and Measurement, IEEE Transactions on, 47*(5), 1218-1222.

Takagi, T., & Sugeno, M. (1985). Fuzzy identification of systems and its applications to modeling and control. *IEEE Transactions on Systems, Man, and Cybernetics, 15*(1), 116-132.

Takahashi, A., Hands, D., & Barriac, V. (2008). Standardization activities in the ITU for a QoE assessment of IPTV. *Communications Magazine, IEEE, 46*(2), 78-84.

Takatalo, J., Nyman, G., & Laaksonen, L. (2008). Components of human experience in virtual environments. *Computers in Human Behavior, 24*(1), 1-15.

Tan, H. Z., Srinivasan, M. A., Eberman, B., & Cheng, B. (1994). Human factors for the design of force-reflecting haptic interfaces. *Dynamic Systems and Control, 55*(1), 353-359.

Tatematsu, A., Ishibashi, Y., Fukushima, N., & Sugawara, S. (2010). QoE assessment in haptic media, sound and video transmission: Influences of network latency. *Communications Quality and Reliability (CQR), 2010 IEEE International Workshop Technical Committee on,* 1-6.

Tullis, T., & Albert, B. (2008). *Measuring the user experience: Collecting, analyzing, and presenting usability metrics* Morgan Kaufmann.

van der Putten, E. P. W., van den Dobbelsteen, J. J., Goossens, R. H. M., Jakimowicz, J. J., & Dankelman, J. (2010). The effect of augmented feedback on grasp force in laparoscopic grasp control. *Haptics, IEEE Transactions on, 3*(4), 280-291.

Van Moorsel, A. (2001). Metrics for the internet age: Quality of experience and quality of business. *Fifth International Workshop on Performability Modeling of Computer and Communication Systems, Arbeitsberichte Des Instituts Für Informatik, Universität Erlangen-Nürnberg, Germany, , 34*(13) 26—31.

Verdejo, A. J., De Moor, K., Ketyko, I., Nielsen, K. T., Vanattenhoven, J., De Pessemier, T., et al. (2010). QoE estimation of a location-based mobile game using on-body sensors and QoS-related data. *Wireless Days (WD), 2010 IFIP,* 1-5.

Viau, A., Najm, M., Chapman, C. E., & Levin, M. F. (2005). Effect of tactile feedback on movement speed and precision during work-related tasks using a computer mouse. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 47*(4), 816.

Vicentini, M., & Botturi, D. (2009). Human factors in haptic contact of pliable surfaces. *Presence: Teleoperators and Virtual Environments, 18*(6), 478-494.

Wagner, C. R., Stylopoulos, N., Jackson, P. G., & Howe, R. D. (2007). The benefit of force feedback in surgery: Examination of blunt dissection. *Presence: Teleoperators and Virtual Environments, 16*(3), 252-262.

Wann, J., & Mon-Williams, M. (1996). What does virtual reality NEED?: Human factors issues in the design of three-dimensional computer environments. *International Journal of Human-Computer Studies, 44*(6), 829-847.

Watanabe, K., Yamagishi, K., Okamoto, J., & Takahashi, A. (2008). Proposal of new QoE assessment approach for quality management of IPTV services. *Image Processing, 2008. ICIP 2008. 15th IEEE International Conference on,* 2060-2063.

Whalen, T. E., Noel, S., & Stewart, J. (2003). Measuring the human side of virtual reality. *Virtual Environments, Human-Computer Interfaces and Measurement Systems, 2003. VECIMS '03. 2003 IEEE International Symposium on,* pp. 8-12.

Whalley, L. J. (1995). Ethical issues in the application of virtual reality to medicine. *Computers in Biology and Medicine, 25*(2), 107-114.

Witmer, B. G., & Singer, M. J. (1998). Measuring presence in virtual environments: A presence questionnaire. *Presence: Teleoperators and Virtual Environments, 7*(3), 225-240.

Wu, W., Arefin, A., Rivas, R., Nahrstedt, K., Sheppard, R., & Yang, Z. (2009). Quality of experience in distributed interactive multimedia environments: Toward a theoretical framework. *Proceedings of the Seventeen ACM International Conference on Multimedia,* 481-490.

Xiaojun Shen, Jilin Zhou, & Nicolas D. Georganas. (2006). Evaluation patterns of tele-haptics. *Electrical and Computer Engineering, 2006. CCECE '06. Canadian Conference on,* 1542-1545.

Yan Gong, Fangchun Yang, Lin Huang, & Sen Su. (2009). Model-based approach to measuring quality of experience. *Emerging Network Intelligence, 2009 First International Conference on,* 29-32.

Yap, K. & Marshall, A. (2010). Investigating quality of service issues for distributed haptic virtual environments in IP networks. *Intelligent Environments (IE), 2010 Sixth International Conference on,* 237-242.

Yu-Chun Chang, Chi-Jui Chang, Kuan-Ta Chen, & Chin-Laung Lei. (2010). Radar chart: Scanning for high QoE in QoS dimensions. *Communications Quality and Reliability (CQR), 2010 IEEE International Workshop Technical Committee on,* 1-6.

Zadeh, M. H., Wang, D., & Kubica, E. (2007). Human factors for designing a haptic interface for interaction with a virtual environment. *Haptic, Audio and Visual Environments and Games, 2007. HAVE 2007. IEEE International Workshop on,* 21-26.

Zhou, R. (2007). How to quantify user experience: Fuzzy comprehensive evaluation model based on summative usability testing. *Usability and Internationalization.Global and Local User Interfaces,* 564-573.