

Copyright is owned by the Author of the thesis. Permission is given for a copy to be downloaded by an individual for the purpose of research and private study only. The thesis may not be reproduced elsewhere without the permission of the Author.

# **Analysis and Modelling Quality of Experience for Web based Services**

A thesis presented in partial fulfilment of the requirements for the degree of  
Doctor of Philosophy in Network Engineering at Massey University,  
Palmerston North, New Zealand.

**NGUYEN, LE THU**

[2013]

## Abstract

Web technologies are advancing at a rapid pace and they provide users with better quality and advanced features for their web browsing experience. The research reported in this thesis is focused on the “Analysis and Modelling Quality of Experience for Web based services”. We present a reliable model for measuring and analysing QoE in Web Surfing scenarios based on network traces. We demonstrate that, by integration of further human perception factors involving content and state of mind (SOM) into the assessment, the overall model performance has been improved. The models enable users to feedback their opinion of their Web interactions by revealing the way in which they are interacting with the system rather than by explicitly providing their opinion via a subjective approach such as through a questionnaire. A comprehensive set of metrics was developed to evaluate the overall customer QoE. The results certainly help to get a better understanding on how web surfing is experienced by users.

We have identified existing shortcomings in our knowledge for establishment of QoE and we have proceeded to discuss the current state-of-the-art in QoE modelling for Web based services. We have utilized orthogonal arrays using the Taguchi approach to construct our experiments in order to characterize the application, as well as establishing network performance metrics in our QoE assessment model to ease the experimental load and to reduce time and cost of conducting such experiments. We propose a further experiment with our proposed session control to reduce the boredom effects from users which may impact on their subjective assessments.

We propose two further metrics that are related to content and SOM into our QoE assessment. The objective metric of content has been investigated and evaluated showing its effects on QoE. The subjective metric based on SOM has been captured subjectively by users’ feedback showing random change via customer browsing.

We have also investigated the applicability of a mixed effects model in predicting QoE in World Wide Web based multi-media services. An analysis is presented of both objective factors and a human factor that may impact on the outcomes of observations. The third model is developed to account for these factors and other potential covariates to QoE assessment during the course of our experiments.

## Acknowledgements

I would like to express my sincere appreciation and gratitude to Professor Richard J. Harris for his constant and continuous support, encouragement, understanding and patience. His invaluable guidance, advice, constructive criticisms and opinions allowed me to complete this research.

I would like to thank Dr. Gardiyawasam Amal Darshana Punchihewa for his support, sharing his ideas, and understanding during my research.

My sincere thanks go to Dr. Jusak Jusak for his supervision during my study, especially for contributing his kind efforts from overseas in the later stages of my research.

My thanks also go to the Human Ethics Committee and Research Ethics of Massey University for ethics approval; and the School of Engineering and Advanced Technology (SEAT) for their financial support for overseas conference attendance.

I am especially grateful to the Vietnamese Government for their financial support and assistance to me during this PhD Study in New Zealand. I am thankful to the IEEE Communication Society for its travel grant to the 2012 CCNC conference.

Special appreciation goes to my friends at Massey University and in New Zealand for their motivation and friendship.

Last, but not least, many thanks to my family for all their love, encouragement, support and care during my study. With my respect and full of love, my special thanks to my father, Prof. Dr. Nguyen Van Thang, who gave me much support and encouragement during my PhD study.

## Table of Contents

Abstract.....	1
Acknowledgements.....	2
Table of Contents.....	3
List of Figures.....	6
List of Tables.....	8
Chapter 1 Introduction.....	9
1.1 A General Concept of Quality of Experience .....	9
1.1.1 Concept of Quality.....	9
1.1.2 Concept of Experience.....	10
1.1.3 Concept of Quality of Experience in telecommunications .....	10
1.2 Motivation.....	12
1.3 The research aim and Scope.....	15
1.3.1 Research aim.....	15
1.3.2 Scope .....	16
1.4 Approach and Proposed Framework.....	17
1.4.1 Approach .....	17
1.4.2 Proposed Framework.....	20
1.5 Summary of Contributions .....	21
1.6 Thesis Outline.....	22
Chapter 2 Review of QoE Modelling and Assessment .....	25
2.1 Subjective Quality Evaluation for measuring QoE.....	25
2.1.1 The concept of subjective quality assessment .....	25
2.1.1.1 Survey based on user-score .....	25
2.1.1.2 Subjective Quality Evaluation.....	26
2.1.2 Subjective quality evaluation-Current status .....	28
2.2 Objective Quality Evaluation for measuring QoE.....	30
2.2.1 The concept of objective quality assessment.....	30
2.2.2 Current status .....	31
2.3 Objective-Subjective correlation for measuring QoE .....	34
2.3.1 QoS-QoE mapping .....	34
2.3.2 QoE-as a function of QoS, and human factors .....	38
2.3.3 Summary .....	40
2.4 QoE of Web based services .....	42
2.4.1 ITU-T G.1030.....	42
2.4.2 Objective-Subjective link in QoE of Web based services .....	44
2.4.2.1 QoE-QoS mapping.....	44
2.4.2.2 Characterizing the meaning of QoE related to Web based services application.....	45
2.4.3 Analysis of psychological perspectives related to QoE.....	47
2.5 Summary and Discussion.....	47
Chapter 3 Experimental Design.....	51
3.1 Experimental Design .....	51
3.1.1 Taguchi quality method approach.....	51
3.1.2 Factor analysis .....	54
3.1.2.1 Noise factor .....	54

3.1.2.2 Controlling factors.....	55
3.1.2.3 Human factors.....	55
3.2 Test-bed and Controlling Performance QoE Web Assessment.....	56
3.2.1 Test-bed experiment setup.....	56
3.2.2 Controlling Session.....	57
3.2.3 Networking performance configuration.....	60
3.2.3.1 Controlling Network performance.....	60
3.2.3.2 Controlling Application performance.....	61
3.2.3.3 Controlled values.....	62
3.3 QoE Web content design.....	64
3.4 Data Collection.....	66
3.4.1 MySQL Database setup.....	66
3.4.2 Human-Subjects.....	67
3.4.3 User score-MOS.....	67
3.4.4 User tasks.....	68
3.5 Summary.....	70
Chapter 4 Modelling Approach for the QoE of Web based services.....	71
4.1 Conceptual modelling.....	71
4.1.1 Introduction.....	71
4.1.2 Design of a suitable QoE model.....	73
4.2 Evaluation of Metrics for QoE of Web based services.....	74
4.2.1 Networking metric for QoE of Web based services.....	74
4.2.1.1 Pilot Study - Evaluation of delay metric from users' scores.....	74
4.2.1.2 Pilot Study - Evaluation of delay and packet loss metrics from user's scores.....	76
4.2.2 Application metric for QoE of Web based services.....	78
4.2.2.1 Pilot Study - Evaluation of requests per second metric from user's scores.....	78
4.2.3 Evaluation of metrics derived from both networking and application layers to users' scores.....	80
4.2.3.1 Experiment under the second controlling session using the Taguchi Approach.....	80
4.2.3.2 Experiment under the third controlling session by Taguchi Approach.....	82
4.2.4 Subjective metric-Boredom and state of mind for Web QoE assessment.....	85
4.2.4.1 Boredom.....	85
4.2.4.2 State of mind.....	86
4.2.5 Content metric for QoE of Web based services.....	88
4.2.6 Summary.....	92
4.3 Analytical modelling framework.....	93
4.3.1 Proposed framework.....	93
4.3.1.1 Application characteristics-Derived metric.....	93
4.3.1.2 Network –Derived metric.....	94
4.3.1.3 Subjective human-Derived metric.....	94
4.3.1.4 CT-Derived Metric.....	94
4.3.2 Summary.....	95
Chapter 5 Results and Evaluation of QoE assessment.....	96
5.1 A Mixed Effects Model for Predicting Web-based Services QoE.....	96
5.1.1 Introduction.....	96

5.1.2 Model parameters estimation.....	100
5.2 Building the non-linear Mixed Effects Model .....	102
5.3 Building the Linear Mixed Effects Model .....	104
5.3.1 Model response for QoE of Web based services .....	104
5.3.2 Fixed effects component-Performance .....	105
5.3.3 Cooperating random effects method for QoE of Web based services- Performance.....	105
5.3.4 Performance of the model for QoE of Web based services.....	107
5.3.4.1 Introduction.....	107
5.3.4.2 Model 1.....	109
5.3.5 Fixed effects for QoE of Web based services-Evaluation .....	114
5.3.5.1 Log-likelihood .....	114
5.3.5.2 95% confidence intervals.....	115
5.3.6 Random effects method for QoE of Web based services-Evaluation .....	116
5.4 Model of QoE of Web based services-Evaluation.....	117
5.4.1 Residual distribution.....	117
5.4.2 Information criteria .....	118
5.4.3 Correlation .....	120
5.5 Discussion and Conclusions.....	122
Chapter 6 An Enhanced Analytical Framework Involving Psychological Support .....	123
6.1 Model 2 and model 3.....	123
6.2 Enhancements using Content .....	124
6.2.1 Model development.....	124
6.2.2 Comparison of model 2 with model 1.....	125
6.2.3 Estimation of model 2 parameters.....	126
6.3 Enhancements Using State of Mind.....	127
6.4 Discussion and Conclusions.....	132
Chapter 7 Summary, Conclusions and Future Work .....	134
7.1 Summary.....	134
7.2 Conclusions.....	136
7.3 Future Work.....	139
List of Abbreviations .....	140
Bibliography .....	142
Appendices.....	147

## List of Figures

Figure 1-1: QoE-the next issue for QoS research [30] .....	13
Figure 1-3: Three-method combination in QoE Model of Web based services .....	19
Figure 1-4: Proposed Framework and Approach of QoE of Web based services assessment .....	21
Figure 2-1: A five level scale Mean Opinion Score (MOS).....	27
Figure 2-2: Five grade Absolute Rate Scale (Right) [12, 18]and Impairment Scale (Left) [19].....	28
Figure 2-3: ITU's standards on QoE using subjective evaluation methods .....	29
Figure 2-4: ITU's standards on QoE using objective quality evaluation methods .....	31
Figure 2-6: The relationship between different QoS and QoE approaches (extracted from Figure 2 in [11]) .....	38
Figure 2-7: Proposed QoE model in [20] .....	39
Figure 2-9: Average connection duration, Number of packets lost and User Interruptions vs. Loss Rates ([27]).....	46
Figure 2-10: Global Consumer Internet Traffic in period 2011-2016 (Source: Cisco)...	47
Figure 3-1: Network Topology for the Test Bed .....	57
Figure 3-2: Network Session of the Pilot Study .....	58
Figure 3-4: The third Network Session by Taguchi Approach .....	59
Figure 3-5: The fourth Network Session by Taguchi Approach.....	59
Figure 3-6: Linux Queueing Discipline.....	60
Figure 3-7: Queueing discipline for Ethernet 0 [81] .....	61
Figure 3-8: Example of setting in mod_qos.....	62
Figure 3-10: Type 1 (left) and Type 2 (right) of Content .....	64
Figure 3-11: Type 3 (left) and Type 4 (right) of Content .....	64
Figure 3-11: MySQL database .....	66
Figure 3-12: Setup of users and website sources .....	67
Figure 3-13: MOS and its score meaning used in the project.....	68
Figure 3-14: Feedback page .....	69
Figure 4-1: A closed relationship between QoS, QoE, end user action, and user satisfaction.....	72

Figure 4-2: User's Experience for different Delay Scenario .....	75
Figure 4-3: User's Experience for different Delay and Packet Loss Scenarios .....	77
Figure 4-4: Concept of "Rule of Nature" .....	78
Figure 4-5: Request packet in the whole packet for different delay .....	79
Figure 4-6: Mean Opinion Score vs. Delay and Requests per second .....	83
Figure 4-7: Concept of "Rule of Nature" in a complex test.....	84
Figure 4-8: MOS vs. Delay and Requests per second.....	87
Figure 4-9: MOS of {1, 2, 3} matching type of Content .....	91
Figure 4-10: MOS of {4, 5} un-matching type of Content .....	91
Figure 5-1: Non-linear predicted values for MOS vs. Delay and Requests per second.	103
Figure 5-2: Caterpillar plot of normal probability of random variables of time in non-linear model.....	103
Figure 5-3: Nesting label of ( <i>id: time</i> ).....	108
Figure 5-4: Model 1 Approach.....	109
Figure 5-5: Matrix <i>X</i> of Model 1.....	111
Figure 5-6: Matrix <i>Z</i> of Model 1 .....	111
Figure 5-7: Matrix of variance-covariance of random effects associated grouping factor of Model 1 .....	112
Figure 5-8: Caterpillar plot of normal probability of the conditional modes of the random effects for grouping ( <i>id:time</i> ) factor from Model 1.....	112
Figure 6-1: Enhancement approach using Content .....	125
Figure 6-2: Estimation of fixed factor of <i>D</i> , <i>RPS</i> and <i>Content</i> in Model 2.....	126
Figure 6-3: Residuals distribution of Model 3.....	128
Figure 6-4: Conditional modes of the random effects for <i>SOM</i> factor .....	129
Figure 6-6: Estimation of fixed effects of <i>D</i> , <i>RPS</i> and <i>Content</i> and 95% confidence intervals of Model 3 .....	132

## List of Tables

Table 2-1: Guidelines showing the relationship between $R$ and user satisfaction (Table B.1 of G.107 [21]).....	32
Table 2-3: QoS parameter and range related with QoE (extracted from Table 1 in ref. [23]).....	37
Table 3-1: Orthogonal array $L_{18} (2^1 \times 3^7)$ extracted in ref. [72].....	54
Table 3-2: Noise factor and level.....	55
Table 3-3: Level and Metric correspondences.....	55
Table 3-4: Psychological effects during the course of experiments.....	56
Table 3-5: Values of delay and packet loss for the first experiment.....	62
Table 3-6: Values of control factor for the second experiment.....	63
Table 3-7: Values of control factors for the third experiment.....	63
Table 4-1: Factors and Levels.....	81
Table 4-2 Analysis of Variance.....	81
Table 4-3: Analysis of Variance.....	82
Table 4-4: Statistic test for each level of factors.....	85
Table 4-5: MOS vs. Content.....	89
Table 5-1: Calculation of the number of observations.....	104
Table 5-2: Null model and Model 1.....	114
Table 5-4: Model (2.45) vs. Model 1.....	117
Table 5-5: Fixed effects of Model 1.....	120
Table 5-6: Correlation test.....	121
Table 5-7: Test of Significance of Model 1.....	122
Table 6-1: Result of model structural assessment based on likelihood test approach ...	125
Table 6-3: Results of model structural assessment based on likelihood test approach (cont.).....	128
Table 6-4: Estimated values of random effect of $SOM$ .....	129

## Chapter 1 Introduction

*In this chapter, we provide a brief introduction to the concepts of Quality of Experience. Some current research directions and existing problems are identified in the area of Quality of Experience, which lead to the motivation for carrying out the proposed research. The chapter ends with a summary of the research contributions and the thesis outline.*

### 1.1 A General Concept of Quality of Experience

#### 1.1.1 Concept of Quality

The concept of Quality has been defined in many ways by various articles in the literature and other appropriate contexts. The widely accepted definition is “*Quality is the degree to which performance meets expectations*” [1]. Another definition adopted by the American Society for Quality (ASQ) is, “*Quality denotes an excellence in goods and services, especially to the degree they conform to requirements and satisfy customers*” [2].

It can be said that the history of Quality is as old as civilization, and has been evident since 3000 BC with the need for high precision in measurement of length, mass, and time. Quality requirements appeared in factories during the 1970s but were essential as early as the 19<sup>th</sup> century.

Following that time, the concept of Quality Assurance (QA) has been focussed on from a number of different perspectives. Before the 1980s, the concept of quality involved inspecting for quality after production. For example, Kuehn, A.A and Day, R.L said “*in the final analysis of the market place, the quality of product depends on how well it fits the pattern of consumer preferences*”[3]. Since the 1980s, the concept of quality has been driven by customers, so that the concept has been driven to build quality into the process, to identify and correct causes of quality problems. For example, Juran, J.M in 1998 defined “*Quality is fitness for use*” [4]. Oakland, J.S in 1989 defined quality as “*the core of total quality approach is to identify and meet the requirement of both internal and external customer*”[5].

The generally accepted concept of Quality used nowadays involves objective methods of measuring and ensuring dimensional consistency with some specific principles, for example for a product, a system or a business.

### **1.1.2 Concept of Experience**

Experience is the general concept of knowledge or skills of something or some event gained through involvement. There are many types of experience such as physical, mental, emotional, spiritual, religious, or social experience. One may also differentiate between different types of experience, but in general, experience is an encounter of a human being with a system having been defined from beginning to end [6]. Watson, S.J in 1991 described “*experience is the exposure of people to situations and the development of their skills and knowledge as a result of this expose*” [7].

Customer experience is the sum of all experiences that a customer has with a supplier of goods or services over the duration of their relationship with that supplier. Alternatively, customer experience can involve subjective responses from customers to suppliers via direct or indirect means [8].

### **1.1.3 Concept of Quality of Experience in telecommunications**

Over recent decades, the term Quality of Service (QoS) has been used as the principal descriptor for specifying the performance quality of both circuit switched and packet switched networks and, in particular, Internet Protocol (IP) based networks. A multitude of QoS characterisations have been studied in various contexts. Recently, a new study area has been proposed with the aim of interpreting end-to-end quality in the proper sense of including a human user as being at the start and the end of a communication chain. Thus, the notion of Quality of Experience (QoE) was born.

There are many different sources for a general definition of QoE, but all definitions express the fact that QoE is subjective in nature and based on human opinions. According to Kalevi Kilkki [9], “*QoE, sometimes also known as “Quality of Experience” is a subjective measure of a customer’s experiences with a vendor. It is related to, but differs from, QoS, which attempts to objectively measure the service delivered by the vendor.*” According to the International Telecommunication Union (ITU) [10], “*QoE is defined as the user’s perception of the acceptability of an application or service*”. Thus, an assessment of QoE may be influenced by a user’s expectations and pre-conceived concepts. Recent

studies involve measurements both objectively and subjectively of a user's perception. For example, a view from the European Telecommunications Standards Institute (ETSI) [11], "*QoE is a measure of user performance based on both objective and subjective psychological measures of using an ICT service or product.*" However, recent studies led to a more complicated QoE definition, in which the definition is related to more areas such as content, network, device, individual personality, etc. For example, a view from the Qualinet Group[6] is, "*QoE is the degree of delight or annoyance of the user of an application or service. It results from the fulfilment of his or her expectations with respect to the utility and / or enjoyment of the application or service in the light of the user's personality and current state. In the context of communication services, QoE is influenced by service, content, network, device, application, and context of use*".

The notion of a Mean Opinion Score (MOS) was created in 1996 [12], and it represented the first subjective approach to the measurement of QoE. MOS is a method in which users are questioned and they provide an assessment of the QoS that they have experienced via a user survey and a score. This methodology is applied to voice, video, and multimedia services such as video conferencing, storage, and retrieval application [12-20] for example, in which users are listeners, viewers, and audiences respectively. To assess the quality of a voice service, each listener is required to rate each observation using a five point scale: 5-Excellent, 4-Good, 3-Fair, 2-Poor, 1-Bad which are referred to "Imperceptible," "Perceptible but not annoying," "Slightly Annoying," "Annoying", and "Very Annoying" respectively [12]. To assess quality of video, each viewer is asked to vote on some frames of impaired video using an impairment scale [18]. To assess video conferencing, after each presentation, audiences are asked to evaluate each of the presented sequences' quality [19].

Since subjective quality evaluation, such as the MOS approach, is time consuming and expensive, objective methods are preferred. A paradigm for this type of work is a model known as the E-Model, a computational model for voice quality that was issued by ITU-T in the early 2000s [21]. Another model is known as Perceptual Speech Quality Measure (PSQM) and this is for evaluating speech coding [22] whereas the model known as Perceptual Evaluation of Speech Quality (PESQ) is a method for measuring one-way voice quality [23].

Studies carried out since 2010 show that QoE is influenced by service, content, network, device, application or context of use in the context of communications. The correlation between objective and subjective parameters is used to measure QoE. This approach has been widely used to avoid subjective limitations, and to increase the user's perception accuracy [24].

Recently, a group of researchers have worked on a comprehensive concept of QoE and related concepts in Dagstuhl, 2012 [6], the concept of QoE was updated to a more complex idea in which QoE can also be influenced by network conditions, devices used to receive the service, applications used to view and/ or hear the service, user's past experience and their psychological status at the time of using the service.

## **1.2 Motivation**

Surfing the Web is a practical task that is being used in activities ranging from bank and business/commerce transactions to entertainment. Web technology is rapidly advancing in order to provide its users with better quality and more sophisticated and advanced features. Based on networking parameters, Internet Service Providers (ISPs) set out to ensure that they provide a better QoS for their customers.

However, a question arises concerning the assessment of this QoS, since; typically it depends on an ISP's perspective, as to whether this is sufficient to infer a fair assessment of a user's experience of the service when they are the direct users of the service. Therefore it may be meaningless if an ISP guarantees that the service is good based on only QoS measurements, but their users are not satisfied with the actual service delivered [25]. Moreover, a user's opinion, which is subjective, may not be a totally correct assessment either due to personal bias, past experiences and other factors [26].

The measurement model that has been attempted in the current work is related to the emerging technology of web services [27-29]. It is clear that in this area, objective performance metrics that correlate well with subjective scores are still in their early stages of development. Thus, more effort and new standards are still needed for defining the measurement model to represent the perceived quality which is experienced by end users.

As mentioned in Section 1.1.3, QoE contains the concept of end to end QoS within itself. After more than 25 years of QoS research and usage, it has been realized that end to end QoS is a crucial contribution [30], but it is not enough to determine the quality of telecommunication services from a user perspective. QoS is based on technical performance while QoE is based on end-user behaviour. A service provider may conclude that a service with a certain level of QoS used for particular communications intentionally offer users good QoE, other levels of QoS may offer users a poor QoE. Alternatively, end to end QoS schemes are considered to be unable to provide services with a certain quality as expected by users. Therefore, QoE is necessary as a way of expressing the level of quality that users believe they have experienced [11, 30].

QoE is expected to be the next end-to-end QoS research direction [30] as illustrated in Figure 1-1 which shows the three key questions that need to be answered by this research:

- 1) How can we understand what users tell us about service quality?
- 2) How can we measure QoS at a subscriber level?
- 3) What are the key factors involved in the determination of service quality between end-users and operators?

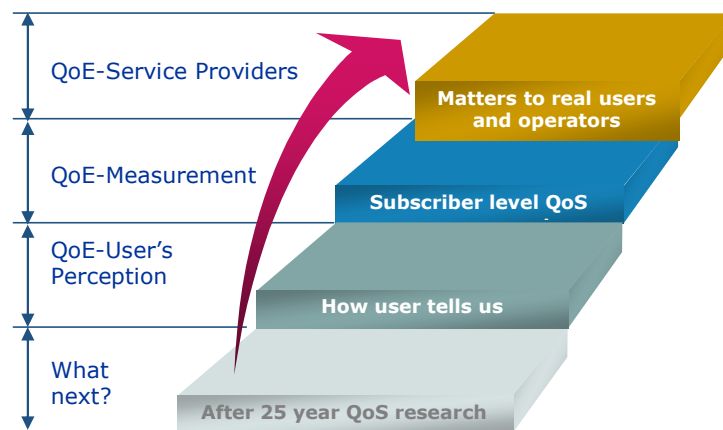


Figure 1-1: QoE-the next issue for QoS research [30]

The current literature discussed in detail in Chapter 2 shows that the concept of QoE is receiving a great deal of attention at the present time and, in particular, it represents initiatives of the European Telecommunications Standards Institute (ETSI) [11], the network of Excellence Euro-NF (Network of Future) [25], and the ITU-T Study Group

12 [31]. It is observed that QoE study, in general, and for QoE of Web browsing in particular, is a newly evolving important issue for three reasons:

- 1) Its concept is still new and not fully defined.
- 2) It needs a reliable, reproducible objective and subjective technical and psychological assessment.
- 3) A recommendation for defining the QoE of Web browsing needs to be updated as it is not really available at this time.

Studies have been published by various recent researchers to identify specific goals that need to be achieved and comprehensive mathematical models that should be developed to support standards recommendations. There are two main important questions that need to be answered and they are:

- 1) How to build a measurement model that represents a good correlation between objectively observed parameters and subjective-based parameters?
- 2) How to build a mathematical model that automates the process of multimedia QoE assessment?

The QoE of Web based services will be reviewed in more detail within Chapter 2, together with some current trends in the QoE of Web based services. From the literature review, it is noted that researchers are working on objective measurement of QoE for Web based services. The questions raised for objective measurement of QoE are:

- 1) What are the objective parameters to be measured?
- 2) How to measure them for the case of Web browsing?

Furthermore, not only the technical parameters measured objectively have an influence on QoE, but also some psychological effects and other effects, hereafter referred to as noises, impact on users for their web browsing assessment.

In overall, there are three current main issues involving the QoE of Web based services. Firstly QoE depended on users' score; however, the users' scores are subjective in nature. That is why they are prone to various errors. Secondly, Due to the first problem, it is necessary to be objective the users' score; however what are the objective measurable parameters and how they can correlate to the real user's scores are question

for the case of web based services, because the objective assessment is in the early development. Thirdly, researchers [32] [11] recognized other psychological effects such as memory and noises such as content have impact on the user's scores. Therefore, how those effects can be minimized in the system or can take in to account in assessment is demanding questions. Figure 1-2 described these main three existing problems.

Figure 1-2 illustrates the problem domain in pictorial form and identifies the key components of the model required for QoE of web based services Figure 1-3.

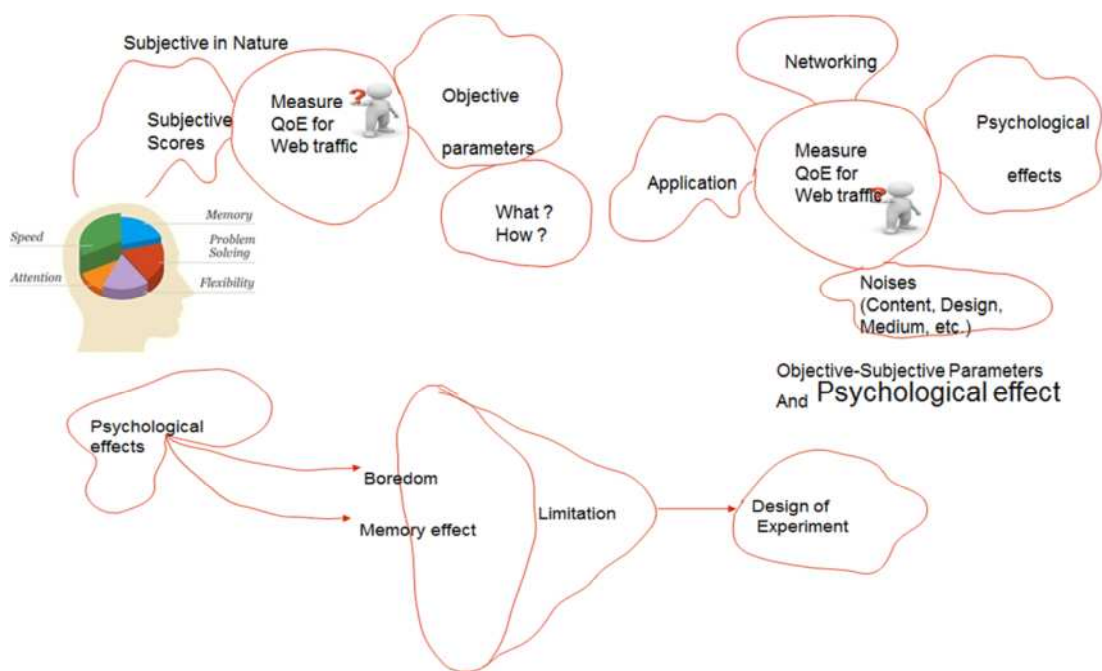


Figure 1-2: Problem Description

From this review, the shortcomings of current methods, and current issues involving the QoE of web based services were identified and we have described the motivation for the study described in this thesis to build up a model to match the aims of the Introduction as discussed in Section 1.3.1

## 1.3 The research aim and Scope

### 1.3.1 Research aim

The research aims to develop a reliable and comprehensive model for determining the QoE of Web based traffic which is not currently available in existing standards.

Furthermore, the proposed model needs to provide good results when compared to traditional MOS methods, but go beyond these to attain greater accuracy; therefore, the model should include the necessary features to account for potential random sources of effects since the data that is collected from users may be a potential source of error.

The objectives of this research are:

- 1) To develop a new QoE model specifically for web based traffic.
- 2) To observe subjective user's state of mind within the system as one input parameter in the development of a new QoE model.
- 3) To observe objective parameters derived from networking, web application performance and content perspectives which impact on a user's score as input parameters in the development of the Web based QoE model.
- 4) To understand and incorporate objective content and subjective state of mind effects within the system as input parameters for the development of the model.
- 5) To evaluate the proposed model by comparing it with real user scores.

The deliverable goals:

- 1) A validated new QoE Model for web traffic: The final QoE web assessment model to be developed using a comprehensive set of the above metrics in order to construct a full and detailed understanding of a user's experience in a web based service environment.
- 2) A comparison and critical review of the approach when compared with competing models.

### **1.3.2 Scope**

In determining QoE, we realize that other factors may impact on the results of the model such as: users who are/are not computer literate, systems affected by viruses, spyware, and the quality and capability of computers that are being used. In this research, it will be necessary to assume that all such access devices are in good condition and all these factors are not having a significant impact on our results. It means that users are skilful in using their computer, equipped with high-speed computers as well as networks, and there are no viruses or spyware in their computers or devices.

Browsing during the course of our experiments is based upon Internet Explorer 8.0 [33] and was not altered during the course of our experiments. In this thesis, it is assumed that there is little impact on users when using different types of browsers such as Chrome, Firefox, etc. which represent the top five browsers as of December 2012 [34].

To determine the behaviour of users, their state of mind is measured together with other effects. However, in the scope of our thesis, state of mind is subjective and measured by users' feedback. To obtain the true objective state of mind, we would need further techniques and further ethics approval from the University that would greatly broaden the scope of this research. (For example, the use of a camera system to detect user behaviour and other methods are needed in capturing and analysing user's mood via video cameras.)

Four-level content measurement carried out objectively is used as a factor which differentiates QoE assessment. In our terminology, *Content* is considered as an objective factor based on downloading time. The definition of *Content* in our experiment has been rather narrowed in comparison with the true meaning of content. Although its definition is narrow, an effect of content is still indicated when the performance of networking and web applications satisfies its users, participants pay more attention to the enjoyment of content. In this way, their assessments are not only dependent on objectively determined technical parameters, but also they depend on the content.

Data for the proposed research is taken from real time traffic via web browsing. It may include text, images, flash objects, audio and video as sources of web pages.

## **1.4 Approach and Proposed Framework**

### **1.4.1 Approach**

Our approach in this research involves combining (by taking the advantages and limiting the disadvantages) those three existing methods which are subjective, objective and psychological methods that will be described in more detail in Chapter 2 and applied to an integrated QoE model for web based traffic. Based on the best of our knowledge there has been no such work documented in the literature to date.

It can be said that, our novel scientific approach to QoE modelling combines three methods of (1), (2) and (3) as illustrated by Figure 1-3. It also improves shortcomings where the three current problems in those three methods need to be solved in order to apply for QoE of Web browsing as outlined below:

- 1) Mitigating the time and cost required in subjective evaluations based on user scores: Firstly, it is necessary to invite users to attend the experiments. Furthermore, the experiments are controlled using a test bed, then a participant needs to attend the tests in the laboratory. As a result, it is difficult, both in time and cost, to carry out a test with a large number of participants. By using the Taguchi robust design method, it is possible to obtain an effective result with fewer experiments and users. It is noted that by applying this robust design, “high quality products can be produced quickly and at low cost” [35], and Taguchi’s proposed design enables experiments to involve fewer runs (data points). For example, if there is 1 control factor with 2 levels and 7 control factors with 3 levels for each control factor then this would normally require 686 individual combinations to consider. This means that there will be 686 ( $=7^3 \times 2$ ) experiments required; however, using the Taguchi approach, only 18 experiments are required by using Taguchi’s  $L_{18}$  matrix. The  $L_{18}$  matrix allows us to study one control factor having two levels, usually in the first row, and the main effects of up to seven control factors where each of them has 3 levels (the orthogonal array is described in more detail in Section 3.1). As a result of following Taguchi’s approach 21 participants require only 882 observations need to be made during the course of our experiments as described in more detail in Section 5.3.1. By way of contrast, if explicit enumeration of all possible combinations were performed, then it would have required  $686 \times 21 = 14406$  observations. Clearly this represents a significant experimental load. Hence the Taguchi approach is much more efficient and there is no loss in overall accuracy of the experiment.
- 2) Defining objectively measurable technical inputs for QoE assessment: This results in not only mapping QoE-QoS but also characterizing the meaning of QoE related to web based service-characteristics.
- 3) Understanding psychological effects and taking into account those effects in the model: Web based services in our project involved viewing web content using a

standard browser where the content, delivery of that content and users state of mind have been subjected to parameter modifications that are expected to influence user perception of web services.

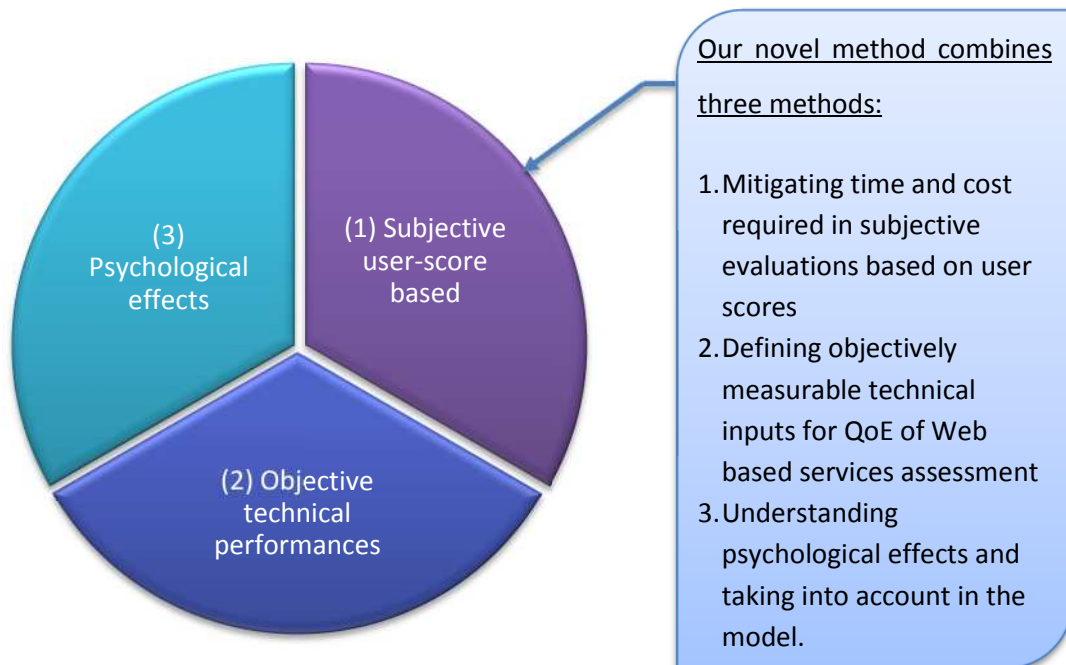


Figure 1-3: Three-method combination in QoE Model of Web based services

This study proposes a new approach to the measurement and analysis of QoE based on a modelling approach. It combines the three factors identified in Figure 1-3, viz: (1) Subjective user scores based; (2) Objective technical performance; (3) Objective and subjective psychological effects.

The approach is aimed at combining and improving upon the three methods for QoE of web based services mentioned above. Firstly, we considered a subjective method that asked questions from users. We realized that using this traditional method would involve a lot of resources and time in carrying out user surveys and determining user scores; therefore, we decided to objectively observe users' behaviour as they interacted with websites and infer their experience from this interaction. However, we need firstly to identify the users' scores and then interpret in terms of resource needs, thus the

Taguchi robust method was applied to ease the experiment load and to mitigate both time and cost required in this subjective method.

Our new model enables people to feedback their opinions by revealing the way in which they are performing with the system and how a system can be capable of meeting their demand; rather than obtaining an opinion via a totally subjective approach. Therefore, the users' opinions can be predicted by both objective technical performance and psychological information. By doing it this way, we have eliminated biased answers that comprise subjective opinions from users.

Objective technical information is measured by both information from the application layer and networking layer to analyse and assess information which is captured from the first step. Our different approach in this step is that we measured and assessed the information from a variety of objective technical information correlations. This ensured that the assessment was not biased which could easily occur when only one parameter is assessed. This leads to the building of a measurement model that can be shown to represent a good correlation between objectively observed parameters and subjectively based parameters.

In addition, psychological effects are investigated and taken into account in the model. Those effects are performed in both subjective and objective analysis. To understand those effects has led to a comprehensive model of Web based services. Thus, the results of proposed models certainly produces a better understanding on how web surfing is experienced by users.

#### **1.4.2 Proposed Framework**

The proposed framework develops measurement methods for individual components of QoE for Web based services. Each component is tested before integrating it into the framework where required. The framework includes four different metrics derived from different categories of networking derived metrics, web application derived metrics, human derived metrics and content-time (*CT*) derived metrics as suggested by Figure 1-4. Figure 1-4 describes the proposed framework that combines three different methods in which delay (*D*) is derived from an objective networking measurement, requests per second (*RPS*) is derived from an objective application performance

measurement, the user's state of mind (SOM), content (C), time (T) and subject (id) are derived from the psychological side.

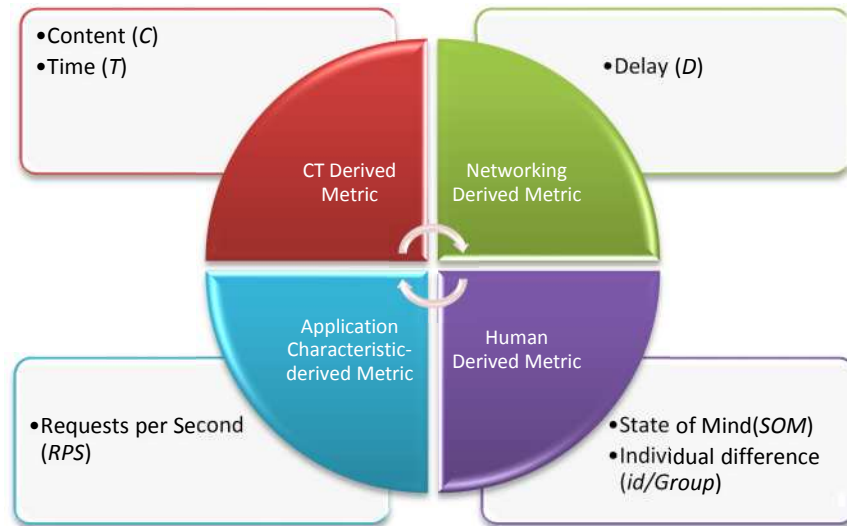


Figure 1-4: Proposed Framework and Approach of QoE of Web based services assessment

## 1.5 Summary of Contributions

The main contributions of this thesis are as follows

- (1) Reviewing the problems arising from earlier and recently proposed QoE models in general and for QoE of Web browsing in particular.
- (2) Reliable modelling of the relationship between MOS scores, web application performance and network performance for the purpose of QoE of Web based services assessment.
- (3) Novel application of the Taguchi method in the context of web performance evaluation.
- (4) A proposal for new metrics to be incorporated into a novel QoE model for web based services using parameters that include delay, requests per second, content and the users' state of mind.
- (5) A novel application of the mixed effect modelling method in the context of analysis and modelling of QoE for Web based services involving objective cases of networking and application assessment.

- (6) Development of an understanding of psychological effects such as state of mind, and content in enhancing the standards for QoE of Web based services.
- (7) Analysis and Modelling of QoE for Web based services in a fully comprehensive assessment situation involving derived metrics for networking, applications, content and the users' state of mind. Such QoE estimates are aimed at providing service and planning, as well as understanding user opinions.

## **1.6 Thesis Outline**

This thesis describes an analysis and modelling of quality of experience for web based services.

### Chapter 1: Introduction

This chapter introduces the Quality of Experience concept, and presents some background knowledge needed for understanding web-based services and user's scores in the thesis. A brief summary of the thesis contributions and an outline of the thesis are presented. Furthermore, this chapter presents background for current problems involving QoE for Web browsing which has led to the combination of the three categories of QoE-QoS mapping, Web application characteristics and psychological effects that are involved in QoE of Web browsing assessment.

### Chapter 2: Review of QoE Modelling and Assessment

This chapter gives an overview of existing QoE measurement studies in general, and existing QoE for Web browsing in particular. QoE measurement studies are divided into subjective quality evaluation, objective quality evaluation, and existing QoS-QoE correlation evaluation. Although QoE for telecommunications services such as video, audio, voice have been updated and recommended, the concept for web browsing remains open in the literature and in the standards community. Current status and trends of QoE of Web based services are presented in the chapter. Recent studies show that the factors influencing QoE for web browsing are not only limited to quality of service, but also related to web application characteristics and other psychological effects are involved.

### Chapter 3: Input Generation for QoE assessment

This chapter describes a method and procedure for assessing and conducting QoE measurements. Firstly, the design of an experiment using the Taguchi Quality Method Approach is undertaken to ensure accurate determination of quantitative user scores with a limited number of users, and makes sure that the test runs are naturally random. Factors are defined and chosen within the two categories of noise and controllable factors. Secondly, constructing a test-bed is described using the networking infrastructure of Massey University. Four controlled sessions are defined in order to collect user's scores.

#### Chapter 4: Analytical Framework for the QoE of Web based services

An analytical framework is presented for assessing QoE web based services. The framework has been developed using a comprehensive set of metrics derived from networking and application layers, and human factors in order to construct a full and detailed understanding of a users' experience in this environment.

#### Chapter 5: Result and Evaluation of QoE assessment

This chapter investigates an application of a mixed effects model in predicting QoE of Web based services. The mixed effect model described in this chapter takes into account effects such as objective parameters for delay and requests per second and the uniqueness of individuals themselves during the course of experiments with other potential covariates. The model demonstrates acceptable correlation between values of the fitted analytical QoE model and observed user scores.

#### Chapter 6: An Enhanced Analytical Framework Involving Psychological Support

This chapter presents an integration of the human factor referred to as the state of mind, and content as well as the uniqueness of individuals themselves during the course of experiments. The integration of state of mind and content does improve the correlation between fitted values and observed values, and also improves the model performance based on information criteria.

#### Chapter 7: Summary, Conclusion and Future Work

This chapter summarizes the main conclusions from the research work and proposes future work to extend the analysis of QoE of Web based services. Future work includes

an objective assessment of state of mind and an extension to a concept involving taking account of content in the assessment.

## Chapter 2 Review of QoE Modelling and Assessment

*In this chapter, an overview of existing QoE measurement studies is presented. These can be divided into three major groups: (1) Existing studies involving subjective quality evaluation; (2) Existing studies involving objective quality evaluation; (3) Existing studies involving QoE-QoS correlation evaluation. In addition, an overview of existing research for web traffic related to the field of QoE is also described. Typical features of these studies are provided through our discussions that help to identify their respective advantages and disadvantages, as well as current problems, which build a foundation for our proposed new model and methodology.*

### 2.1 Subjective Quality Evaluation for measuring QoE

#### 2.1.1 The concept of subjective quality assessment

##### 2.1.1.1 Survey based on user-score

A user score is one of the rating scale techniques used in surveys. A survey is a systematic method of data collection for gathering information about individuals for the purposes of constructing descriptors of the attributes of the larger population for which the individuals are members [36]. Its aim is to collect the opinions of participants in the survey via a question or a set of questionnaires.

All rating scales are levels of measurement and can be divided into three categories according to conceptualization and measurement in [37] as:

- 1) Ordinal scale
- 2) Interval scale
- 3) Ratio scale

**Ordinal scale** uses a number to indicate the relative position of the item, but not the relative degree of any difference of items which are measured. One example of using an ordinal scale is the Likert scale. A Likert scale is often used in questionnaires [38]. It is the most widely used system to scale responses in the user-score survey.

**Interval scale** uses numbers to present fixed measuring units, and indicate the magnitude of the difference between items, but there is no absolute zero point as the zero point on an interval scale is arbitrary, and negative values can be used. It tells us about the order of data points, and the size of intervals between data points. The data can be ranked, and the difference between two values of the level can be calculated. An example of data using an interval scale is temperature (degree in  $^{\circ}C$  or  $^{\circ}F$  ) at sea level.

**Ratio scale** is an interval scale with a fixed and a true zero point. For example, a ratio level is used to measure the number of people in a group.

To increase the success rate of user-score surveys, there are some best practices that are guided by [39] and should be applied to question creation and survey construction.

The question creation guidelines include the following:

- 1) Relevant questions
- 2) Short and Straightforward questions
- 3) Simple and Precise language use for the question

The survey construction guidelines include the following:

- 1) Get to the point
- 2) Keep it personal
- 3) Take action
- 4) Motivate customers to respond
- 5) Limit Survey frequencies

#### ***2.1.1.2 Subjective Quality Evaluation***

Subjective Quality Evaluation is a method that uses laboratory equipment, data sets, a testing methodology, and score determination to measure QoE. The foundation of a subjective evaluation of QoE is known as the Mean Opinion Score (MOS), which used user surveys and user scores to evaluate QoE. It involves collection of opinions from test participants who are asked to rate the QoS with respect to the considered medium.



Figure 2-1: A five level scale Mean Opinion Score (MOS)

MOS is based on a five-point subjective scale of {1,2,3,4,5} which correspond to the following qualitative opinions {Not Recommended, Dissatisfied, Fair, Satisfied, Very Satisfied} respectively as shown in Figure 2-1 [12]. It is used to evaluate the subjective quality of speech /video /audio /multimedia such as subjective video quality measurement in [12, 18, 19] and subjective audio measurement in [12, 18, 19].

The outcome of any subjective experiment is mapped into an MOS, and there are a number of different scales based on generic MOS scores. Figure 2-2 shows two popular scales using subjective assessment with a five-grade absolute quality rate scale and a five grade impairment scale. The five grade absolute quality rate scale is divided into five levels {5,4,3,2,1} mapped to {Excellent, Good, Fair, Poor, Bad} respectively. The five-grade impairment scale divided into five levels of {5,4,3,2,1} are mapped to {Imperceptible, Perceptible but not annoying, Slightly annoying, Annoying, Very annoying}.

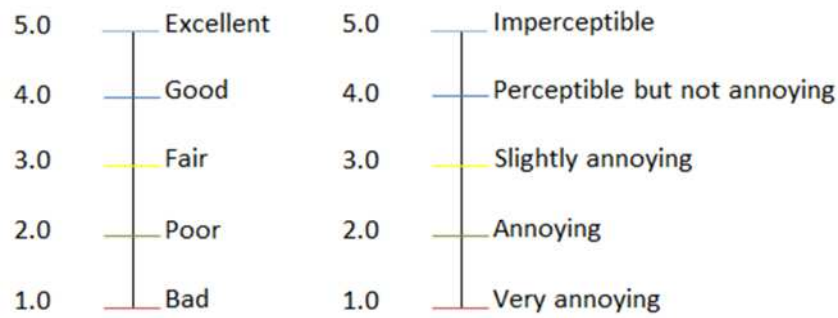


Figure 2-2: Five grade Absolute Rate Scale (Right) [12, 18] and Impairment Scale (Left) [19]

### 2.1.2 Subjective quality evaluation-Current status

Several methodologies using subjective quality evaluation have also been standardized in ITU documents, for voice, video, and multimedia services. Figure 2-3 shows more details regarding these recommendations.

For example, in the ITU Standard P.800 for voice [12], a number of listeners rate the perceived voice quality in a test following the above five-point subjective absolute scale obtained from each user at the end of each conversation.

Formal subjective listening tests are based on recommendation ITU-R BS. 1116-1 [13]. It is considered to be a reliable test of judgement for audio quality. However, these subjective results may not fully reflect actual user perceptions.

A methodology for subjective assessment of visual quality assessment was formalized in ITU-R BT. 500 [16]. Subjective tests used in the experiments conducted under this recommendation involve estimating the performance of a television system by a system viewer. However, this recommendation is no longer used for television, replacing it with ITU-T Rec. 910 [18] for multimedia applications instead.

The five-grade absolute quality scale is used for video as ITU Standard P.910 [18], where a number of viewers voice their opinions on a particular video segment and rate each test individually without comparison. After each presentation of a particular sequence, the viewers are asked to evaluate the quality of the sequence shown. It also recommended some commonly subjective procedures for visual quality assessment such as the double stimulus continuous quality scale, and the double stimulus impairment scale for testing reference video, rating the amount of impairment in test, and rating the instantaneously perceived quality on a slide.

Similarly, the same absolute and impairment scales are used for video visual in multimedia applications ITU Standard P.911 [19] where non-interactive subjective assessment is used to evaluate one-way overall audio-visual quality for bit rates for applications such as video conferencing, storage and retrieval application or tele-medical applications. Without explicit reference, the absolute scale is used when subjects are asked to evaluate the overall quality of the sequence presented under different test conditions after each presentation. With reference, the five-grade level impairment scale is used for the subjects to rate the impairment of a stimulus in relation to the reference.

Several methodologies using subjective evaluation have been standardized in ITU Documents, for speech signals [12, 17, 20, 40, 41] , audio signals [13, 15, 42], and multimedia services [14, 16, 18, 19] as shown in more detail in Figure 2-3.

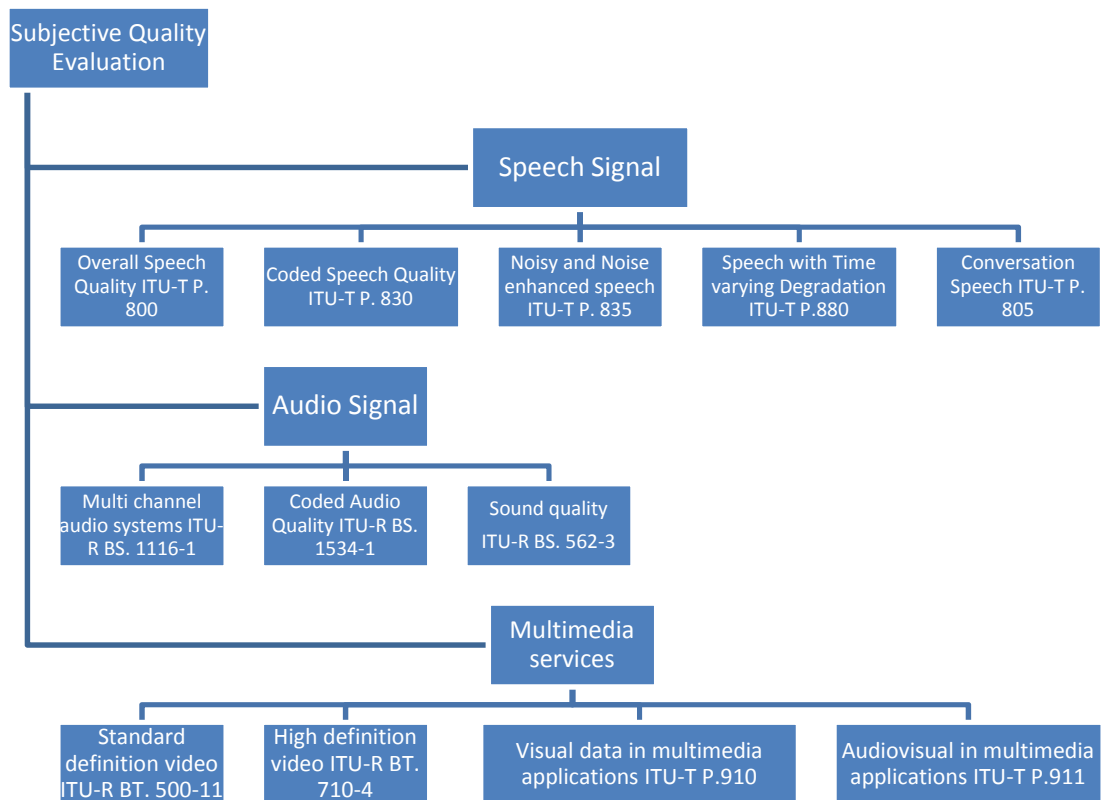


Figure 2-3: ITU's standards on QoE using subjective evaluation methods

Figure 2.3 shows examples using MOS in various recommendations. It is obvious that MOS is used widely in many measurements related to subjective quality of speech signal, audio signal and multimedia services.

Various five-point scales of MOS have been used for different purposes and most frequently used for ITU applications. Of course, this method has some merit such as reflecting more direct user experience and presenting a close match to user behaviour. It can be said that subjective quality evaluation is the most reliable way to measure QoE.

However, since it is a totally subjective model, it is clearly costly and time consuming to obtain user (subjective) scores and to hire experts to estimate users' behaviour and ensure that the results are free of any statistical bias.

## 2.2 Objective Quality Evaluation for measuring QoE

### 2.2.1 The concept of objective quality assessment

Objective Quality Evaluation aims to apply an automatic and reliable way to estimate a user's perception of a service. Its goal is to have a good correlation with subjective quality evaluation methods.

The main purposes of objective quality evaluation for measuring QoE standards are:

- 1) Characterizing the meaning of user opinions related to specific applications;
- 2) Defining a method for reliable user opinions;
- 3) Defining a method for prediction of user opinions.

There are three available methods that are often used for an objective evaluation:

- 1) **Full reference** in which both reference and processed data are available for detailed objective-subjective comparison;
- 2) **No reference** in which only processed data is used for objective-subjective comparison;
- 3) **Reduced reference** in which some features are extracted from reference and processed data are available to derive and compare objective and subjective correlation.

The **full reference** method represents the highest accuracy, but it increases the non-data load. The **no reference** method may give low accuracy because network conditions may affect its quality estimation; however, it has no effect on networking load. The **reduced reference** promises a benefit over the first and second methods as it represents the combination of advantages from the first two methods such as higher accuracy but less non-data load.

### 2.2.2 Current status

Several methodologies using objective evaluation have been standardized in ITU Documents, for voice [21-23, 43], audio [44], and multimedia services [45, 46] as shown in Figure 2-4.

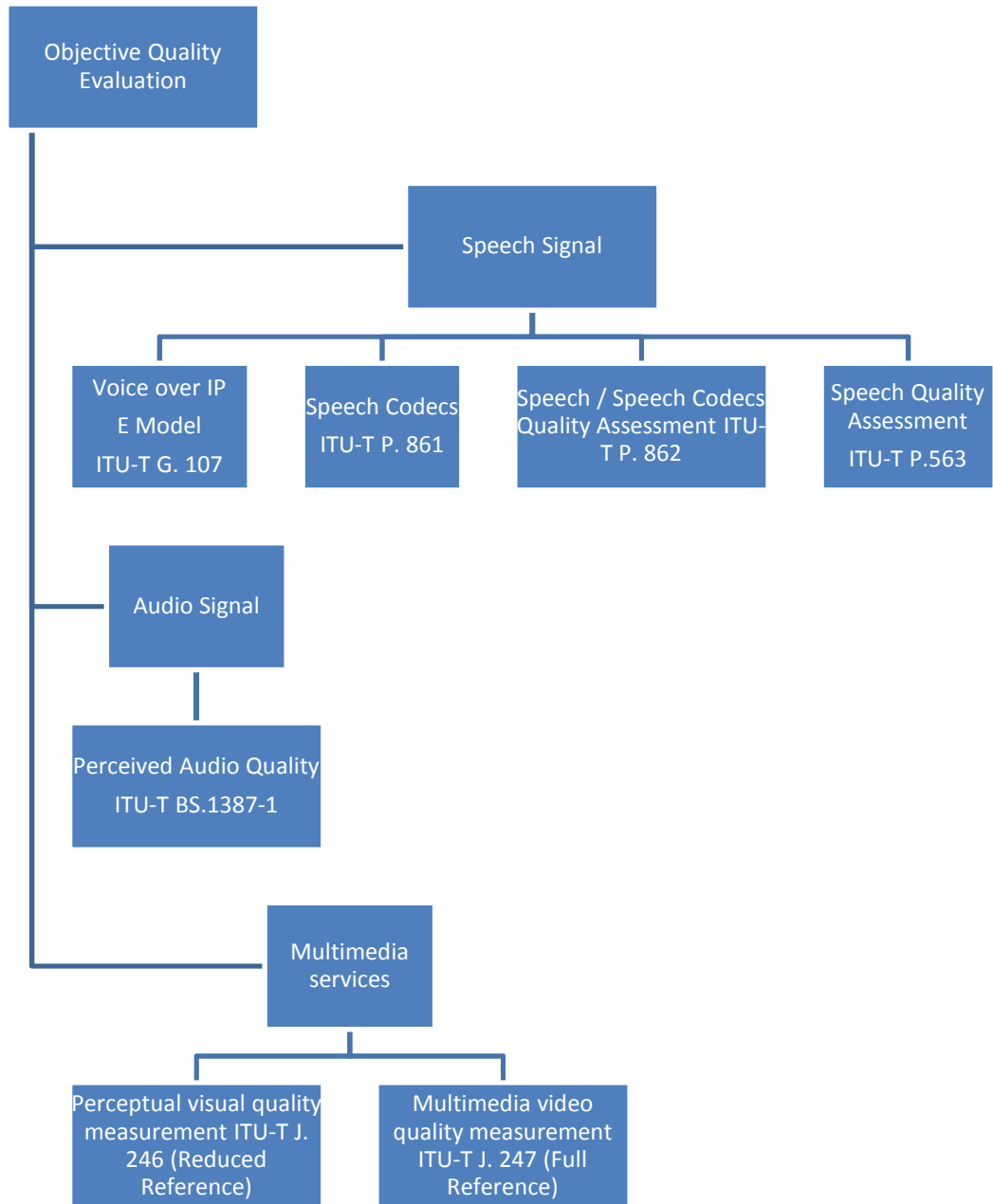


Figure 2-4: ITU's standards on QoE using objective quality evaluation methods

An important transmission rating model for measuring QoE for voice that is widely-known and adopted is the E-Model (ITU-G.107)[21]. This model estimates two way

conversational qualities as perceived by users in terms of listener and talker. It requires extensive knowledge of the system components for estimating user satisfaction and enables a determination as to whether a user will be satisfied with end to end transmission performance. In ITU-T G.107, a parameter known as the  $R$ -factor is used as a measure of quality and is defined by Equation(2.1).

$$R = R_0 - I_s - I_d - I_e + A \quad (2.1)$$

where:

- $R$  : Transmission rating factor
- $R_0$  : A basic ratio of signal to noise
- $I_s$  : Simultaneous impairment factor
- $I_d$  : Delay impairment factor
- $I_e$  : Equipment impairment factor
- $A$  : Advantage factor for expectation.

According to G.107 [21], the  $R$  factor is always a numerical value between 0 and 100 and an acceptable value of  $R$  is greater than 60 while an unattainable value of  $R$  is over 94.5 for current VoIP services. Table 2-1 is an extract of Table B.1 from the G.107 recommendation that shows how the  $R$  value and user satisfaction are related.

Table 2-1: Guidelines showing the relationship between  $R$  and user satisfaction (Table B.1 of G.107 [21])

$R$ value	MOS	Good or Better (%)	Poor or Worse (%)	User Satisfaction
$\leq$	$\leq$	$\leq$	$\geq$	
90	4.34	97	Nearly 0	Very satisfied
80	4.03	89	Nearly 0	Satisfied
70	3.60	73	6	Some users dissatisfied
60	3.10	50	17	Many users dissatisfied
50	2.58	27	38	Nearly all users dissatisfied

The results of the MOS scores are presented in the second column of Table 2-1, and the values for a VoIP conversational situation are computed using the  $R$ -factor which is then scaled to a range of values from 1 to 5.

The relationship between  $R$  values and MOS is displayed in more detail in Figure 2-5. A five scale MOS is applied. For example, if the  $R$  value is less than or equal to 50, nearly all users are dissatisfied, the voice quality and the MOS lies in the range from 2 to 3.

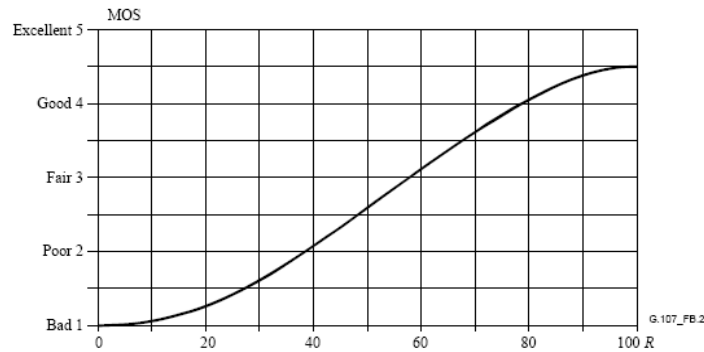


Figure 2-5: MOS and  $R$  values (Figure B.2 of G.107 [21])

It could be said that this model gives an approach to modelling QoE by deriving application layer performance metrics based on network related performance parameters. However, because it is based on impairment values, therefore, it is too complex and needs more extensive knowledge of human perception; moreover, it is not supported for web traffic.

An alternative objective model is referred to as the Perceptual Speech Quality Measure (PSQM-P.861) [22]. This method is used specifically for speech coding and is linked to human auditory perception. Unfortunately, the PSQM model is suitable only for speech codecs and not for networked situations. Moreover, it gives poorer results when correlated with subjective opinions in some normal situations where there is background noise or packet loss. This recommendation has been recognized as having certain limitations in the specific area of application, and thus it was replaced by P.862 [23].

By injecting a signal into the system under test, degraded output is compared by Perceptual Evaluation of Speech Quality (PESQ-P.862) [23] using a reference input signal; PESQ can measure one-way voice quality, it demands no knowledge of the system under test, but it does require extensive knowledge of human perception.

The Video Quality Experts Group (VQEG) was formed in 1997 to address video quality issues. Their current goals are to advance quality assessment of the field of video,

and investigate new subjective assessment methods in which subjective rates are recorded, then used to predict objective quality metrics [47, 48]. The method for objective measurements of perceived audio quality is recommended in ITU-R BS. 1387.1 [44]. Objective perceptual video quality measurement with an available full reference is recommended in ITU-T J.247 [46] which defines four full reference models. ITU-T J.247 recommended a selection of appropriate objective perceptual video quality measurement methods such as testing a codec or testing a transmission chain. ITU-T J. 246 [45] defines three reduced reference models to measure perceptual visual quality for multimedia services over digital cable television network. For example, the edge peak signal to noise ratio (EPSN) reduced reference model calculated the mean squared error from a degradation of edge pixels.

The above models described in Sections 2.1 and 2.2 are examples of models that are both subjective and objective for QoE assessment. However, these measurement models are suitable for specific kinds of traffic involving audio signals. Thus, they cannot be applied in situations which involve web traffic in its many application forms. Moreover, they rely too much on scores and user opinions that are expensive in time and money to obtain.

## **2.3 Objective-Subjective correlation for measuring QoE**

### **2.3.1 QoS-QoE mapping**

Since QoE relates to a user's experience it partly involves a form of psychological measurement (subjective). However, it is important to telecommunication service providers to express QoE objectively in relation to their networks and equipment.

Until now, there has been only limited investigation into this relationship [27, 49-51]. By combining both the experience of users (subjective) and measurements (objective) QoE may be more reliably measured or estimated. Some recent models [27, 49-51] have shown methods for correlating QoE-QoS with QoE measurement models. Four different research papers have developed formulae to compute QoE in the correlation category. They are identified below as QoE<sub>1</sub>, QoE<sub>2</sub>, QoE<sub>3</sub>, and QoE<sub>4</sub> respectively.

QoE<sub>1</sub> is calculated via the QoS quality parameters by Fielder, M. et al. and it is described as the IQX hypothesis [31]. This methodology provides a bridge between subjective and objective measurement. Fielder, M. et al., mentioned a generic quantitative relationship

between QoS and QoE called the IQX hypothesis [27]. The IQX hypothesis can be said to be straightforward to use in a QoE-QoS relationship whereby measured QoS values are inserted into a corresponding exponential QoE formula to manage and control QoE.

$QoE_1 = \phi(I_1, I_2, \dots, I_n)$  is a function of  $n$  influence factors  $I_j$ . A singular impact factor  $I = QoS$  is used to derive the fundamental relationship  $QoE_1 = f(QoS)$ .

Table 2-2 [52], for example, listed some QoS parameter attributes that included technical parameters such as delay, packet loss or jitter and their values can be related to users' experience perspectives as user satisfaction and user enjoyment.

Table 2-2: QoE Attributes and their main parameters[52]

Attribute	Parameter	Examples
Communication situation	User task	Give instruction, negotiate an outcome
	User group	Business people, elderly people
	User environment	Conference room, in a parked car
Service prescription	Service type	Video call, audio call, video on demand, IPTV
	Terminal type	Laptop computer, mobile handset
Technical parameters	Bit rate	1 Mb/s, 64 kb/s
	Media protocol	H.264, MPEG2, AAC
	Network protocol	TCP-IP, UDP-IP, RTSP
	Delay	50 ms, 500 ms, 1 s
	Audio-video asynchrony	0, 50 ms, -100 ms
	Jitter	50 ms, 100 ms, 1 s
	Packet loss	0.5%, 1%, 5%
	Video frame rate	7 frames/s, 25 frames/s, 30 frames/s
	Video resolution	CIF, 1920 × 1080, XGA
	Task effectiveness	Task accuracy, value of negotiated agreement
	Task efficiency	Task time, number of speech interruptions
User experience	User satisfaction	Acceptability of the service, satisfaction with communication
	User enjoyment	Level of engagement, level of fun

$QoE_2$  is suggested via a non-linear function of QoS in [49]. By measuring the QoS information at a network level, a QoS-QoE correlation model for objective assessment of QoE has been demonstrated in theory by the authors of [49]. Using this theoretical

approach, an objective QoE measurement is obtained via QoS parameters. The fundamental assumption behind traditional network provisioning is that the measured QoS is closely related to the QoE for the end-user. In order to describe the QoS and QoE correlation model, a subjective quality evaluation method using a QoS parameter in a converged network environment is studied. By using the QoS information measured at a network-level, a QoS-QoE correlation model for objective QoE<sub>2</sub> was proposed using Equations (2.2) and (2.2).

$$QoS = F(\text{Delay, Jitter, Error Rate, Bandwidth, Signal Success Rate}) \quad (2.2)$$

$$QoE_2 = QoE(QoS) = K \left\{ \frac{(e^{QoS-\alpha} + e^{-QoS+\alpha})}{(e^{QoS-\alpha} + e^{-QoS+\alpha} + \beta)} + 1 \right\} \quad (2.2)$$

Where:

$\alpha$  = QoS class required in the network and is mapped to the 5-point MOS scale.

$\beta$  = Class of Service with an associated constant grade of service probability.

$K$  = User satisfaction scale factor

By using quality indexes at a network level, the QoS scores can be obtained. The QoE class measured by the QoS quality parameter at a network-level is mapped as for the existing MOS grade scheme that consists of five classes. Depending on QoS class agreement, the service provider will make a decision whether or not packets are transferred at the specific networking capacity.

The determination of  $\alpha$ ,  $\beta$  and  $K$  values are now part of on-going research. On the other hand, the QoS parameter can be taken according to a common standard as shown in Table 2-3 that has been extracted from their paper.

In [49], Hyun Jong, K. et al., confirmed that QoE<sub>2</sub> assessment can be obtained from QoS parameter values. The paper presented a QoE-QoS correlation model, and developed an end-user measurement based QoS. The QoE<sub>2</sub> class measured by the QoS quality parameters at the network level is mapped onto the existing MOS grades with 5 classes. From this model, it can be seen that the relationship between QoE and QoS is clear or, alternatively, QoE<sub>2</sub> can be reflected in the QoS values which are mapped to

existing MOS. However, choice of the values for  $\alpha$ ,  $\beta$  and  $K$  in a real time multimedia service still has no answer; and web traffic is not studied in their model.

Table 2-3: QoS parameter and range related with QoE (extracted from Table 1 in ref. [23])

QoS parameters and range			Note
Parameter	Range	Score	
Delay (D)	70ms-100ms	10	Y.1541 G. 1010
	100ms-150ms	8	
	150ms-200ms	6	
Jitter (J)	30ms-50ms	10	Y 1541
	50ms-60ms	5	
	60ms-70ms	3	
Packet loss rate (L)	$\sim 10^{-5}$	10	Y. 1541 DSL-Forum
	$10^{-5} \sim 10^{-4}$	7	
	$10^{-4} \sim 10^{-3}$	3	
Call success rate (S)	100%-99.9%	10	Undefined
	99.9%~99.0%	9	
	99.0%~98.0%	8	
Handover (HO)	HO within the cell	N/A	Undefined
	HO between cells	N/A	
	HO between the frequencies	N/A	

QoE<sub>3</sub> is accessed via sets of goals for learning (QoL), interaction (QoF), and quality of service (QoS), and is described by Equation (2.2) [50]. This formula involves learning and interaction from users that is described in more detail in Section 2.3.2 where QoE is as a function of both QoS and human factors.

QoE<sub>4</sub> is defined via an exponential function of QoS by Khirman, S. [51]. This formula involves developing a relationship for session time in order to calculate user's experience for web browsing. Further details are presented in Section 2.4.2.1.

The proposed work in [27, 49-51] leads to new perspectives in measuring QoE when QoS parameters are linked to QoE perspectives in specifically end-user multimedia QoS categories of IP-based services as recommended in Y. 1541 [53] and G. 1010 [54].

### 2.3.2 QoE-as a function of QoS, and human factors

It is obvious that QoS is an influencing factor behind the measurement of QoE. However, some existing studies lead researchers to further related concepts in measuring that user perception. According to an updated report of the European Telecommunications Standards Institute (ETSI) [11], “QoE is user-centered, expressed in technical QoS measure and based on both subjective and objective psychological measures”, this is illustrated by Figure 2 of their report that we have reproduced as Figure 2-6. Figure 2-6 describes a QoE approach by ETSI, in which QoE is defined via a technical-centered QoS and usage outcome. The usage outcome is sourced from user side related to user perception.

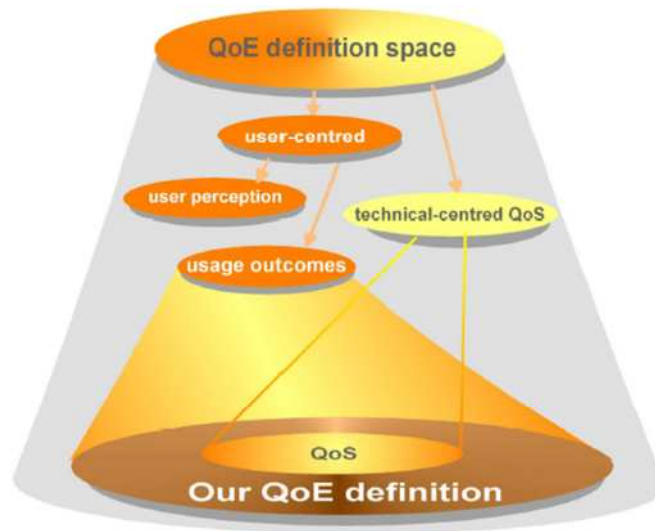


Figure 2-6: The relationship between different QoS and QoE approaches (extracted from Figure 2 in [11])

Other approaches are based on a pervasive context computing environment in order to evaluate QoE [55] in which a rough-set based algorithm is proposed to reduce context attributes and determine the weight of each attribute. Its advantages are that it delivers QoE evaluation results that closely match the real feelings of users and it produces a mass of evidence information related to the experience of users which can be gathered through context awareness computing. However, it needs to be further enhanced with pervasive computing developments.

Another QoE model, which is QoS based, is known as Experience-aware Adaption [50]. This is a new approach to QoS that involves learning from a user and their experience which is shown in Figure 2-7. Moebis, S.A. [50] balances the constraints imposed by

QoS restrictions with the requirements of flow and learning to produce the highest possible QoE for the learner using an adaptive multimedia system. Factors which most influence QoE in an e-learning setting are identified. He uses these to develop an adaptive hypermedia e-learning system that best improves the user's QoE. Flow related as well as learning-related aspects are the main components of the QoE<sub>3</sub>.

$$QoE_3 = f(QoL(QoS, QoF(QoS))) \quad (2.3)$$

Where:

*QoL* = A clear set of goals for learning

*QoF* = A clear set of goals for interaction, skills, and challenges

*QoS* = A clear set of goals for quality of service

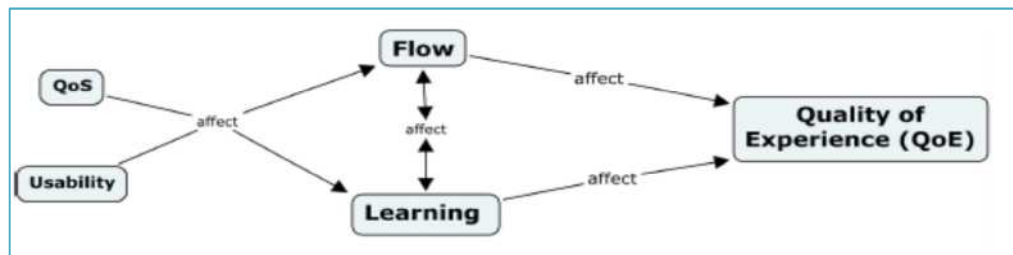


Figure 2-7: Proposed QoE model in [20]

Accordingly, QoE<sub>3</sub> is affected by four main factors: flow, QoS, learning, and usability as specified by Function(2.3). Evaluating QoE<sub>3</sub> it is noted that it is not only based on QoS, but also from learning and flow. The interaction of QoS, learning, usability, and flow is a clearer way to have a more satisfactory QoE<sub>3</sub> model. The demonstrated model is comprehensive with respect to inputs for QoE<sub>3</sub> assessments. Moebs, S.A. follows the idea of Adaptive multimedia e-learning systems to gain better information from users. A learning mechanism is important to filter any unnecessary information from users. However, how to do this filtering for each internal element and what are the affects between them are not provided in their paper, as it is only outlined by the author for future work.

Furthermore, according to the Qualinet white paper on QoE [6], factors influencing QoE are any characteristics of a user, system, service, application, or context grouped

into three categories namely: Human, System, and Context. Human is a characteristic of human users described by their mental constitution, socio-economic background, etc. System determines the quality of an application or service. Context is related to user environment as physical, temporal, social, economic, task, and technical characteristics.

### **2.3.3 Summary**

The research described in Sections 2.1, 2.2 and 2.3 identifies a new approach to measurement of QoE that provides a way of correlating QoS with QoE. Table 2-4 summarizes the advantages and disadvantages of selected studies in the three QoE assessment methods discussed in Sections 2.1, 2.2, and 2.3 above.

QoE is assessed via a function of QoS in theory for many different applications mentioned in Section 2.3.1. This shows that QoE-QoS mapping is possible for the case of Web based services or Web browsing.

The IQX hypothesis used in [27] has described the QoE-QoS relationship for web browsing. The author mentioned relationships for QoE as a function of one singular factor, for example. The correlation between MOS and a QoS singular impact factor is acceptable. As a result, this method is possible for the case of web browsing. However, the authors have not considered the practicality of combining the two QoS factors therefore the relationship is a singular impact factor separately and is not concurrently combined impact factors. Basing QoE on only a single parameter could lead to a biased view of a user's quality of experience or be difficult to assess their scores in more complicated service performance scenarios.

The work in [11, 50, 55] highlights the fact that QoE is not just a simple QoS function as QoE actually involves human-subjects and their perceptions, otherwise only technical QoS assessment is not comprehensive for users' perceptions. Therefore, for QoE of Web browsing, in particular, human psychological effects should be taken into account. This leads to our work where we claim that we should have human psychological measurements that are combined with both subjective MOS tests and objective technical measurements.

Furthermore, this method needs to be more quantitative [52] as a slight change in QoS may produce a big effect on human perception and as each person's assessment may

depend on many different psychological human factors and the work needed to avoid the MOS bias.

Table 2-4: Summary of main research in three methods of assessment

QoE Model approach		Advantages			Disadvantage		
Method	Foundation						
Subjective assessment	P.800 [12-15, 17-20, 40, 42, 56]	High accuracy			Time consuming	Cost consuming	Subjective assessment
Objective assessment	E-model [21]	Objective assessment			Need extensive knowledge of human perception as based on impairment values		Only for VoIP
	QoE-QoS mapping [51]	Objective assessment	Capable	High mapping	Only two main factor bandwidth and latency are crucial role in end user satisfaction		
	IQX hypothesis [24]	Objective assessment	Capable; Automatic	High mapping	Concurrently combined impact factors are not mentioned.		
Method	Current	Advantages			Disadvanges		
Objective assessment	[52]	Automatic	Comprehensive parameters		In theory, has not applied in practical situation		
	[49]	Capable	Assigned different classes	Automatic	In Progress	Not included web browsing	
Psychological assessment	[50]	Human factor involved	Understanding customer QoE involving human factors		State of Art		
	Memory effect[32]	Capable	Automatic	Further understanding of user's side			

## 2.4 QoE of Web based services

QoE for traditional telecommunications services such as voice, audio or video were investigated before 2000s. However, the concept of QoE for different IP based applications including web-browsing was first mentioned in 2002 [57], and recommended in 2005 [29] and with a re-evaluation of QoE for Web-browsing more recently in 2009 [28] and recommended for further study in 2010s [27, 32].

Defining a particular QoE for web services is a challenging area covered by existing and on-going research. Figure 2-8 describes some existing recommendations [29, 54] and recent web traffic studies [57-59] which are related to human perception of specific web services or web browsing experiences.

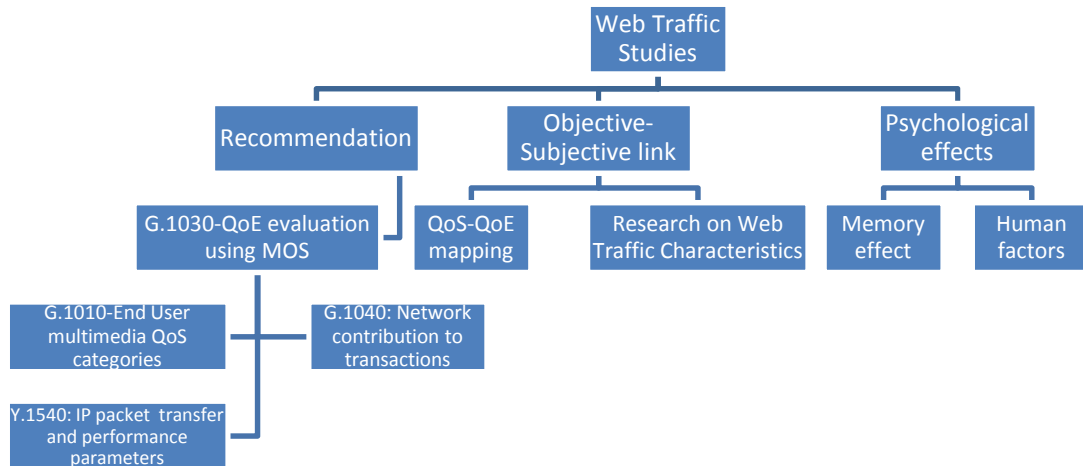


Figure 2-8: Some ITU Recommendation related to Web Traffic and current studies

### 2.4.1 ITU-T G.1030

End-to-end performance was considered as a concept for QoE for web-browsing and recommended in G.1030 [29] in which session time is measured in terms of MOS. The G.1030 is based on recommendations from the end user multimedia QoS categories in G.1010 [54], network contribution to transactions time in G.1040 [60] and available performance parameters in Y.1540 [61]. G.1010 considers quality of service performance considerations for different IP based applications including web-browsing. The quality of service of delay considered for web-browsing is not higher than 10ms.

ITU-T Recommendation G.1030 [29] published in 2005 provides a way to estimate end to end performance in IP based networks for services including web browsing. G.1030 involves estimation of performance from end to end data applications in Internet Protocol Networks. G.1030 is an opinion model for web-browsing applications, which used:

- 1) A subjective web browsing quality experiment and results
- 2) Estimated network performance with limited information

However, this model has some drawbacks which are shown in [28], where they present an evaluation of QoE on an IP network by changing conditions and estimating different services and applications in order to update ITU-T Recommendation G.1030 which has not been amended to match current networking architectures and user behaviour.

Accordingly, estimating QoE seems not to be a simple task. ITU-T G.1030 provides guidelines to achieve this goal. However, this recommendation needs to be updated with the current network, user requirements, and technical improvements. For example, G.1030 used scales around 6, 15, and 60 seconds that represent fast, moderate, and slow networks respectively but this has not been updated.

It can be said that ITU-T. G.1030 is the only ITU recommendation involving QoE for Web browsing in which the QoS for Web browsing is recommended by the ITU-T. G. 1010 [54]. According to Ibarrola, E., et al. [28], G.1030 is a reference point for QoE evaluation; however, it needs to be updated by current service characteristics in order to bring the QoE model up-to-date. From the guidelines stated in ITU-Recommendation G.1030, Ibarrola, E. et al., provides a framework for evaluating QoE based on:

- (1) Having relevant measurements or network simulation results
- (2) Governing protocols with specific options or using network performance and customer performance as a key metric of application performance
- (3) Interpreting the application performance as an estimate of QoE by typical users

Furthermore, as mentioned earlier, G.1030 recommendation published in 2005 [29] aims to estimate end to end performance in IP based networks for services including web browsing. This recommendation is a starting point for early QoE studies of web based services. As mentioned earlier, with limited network information, a user's

behaviour is not matched fully in G.1030 and this references its shortcoming. Psychological effects have not been taken into account in this standard. In recent literature from ETSI [11] they concluded that there should be other effects on QoE assessment from human perspectives that should be further studied.

Therefore, in order to take account of G.1030's shortcomings, the directions for ITU-T for QoE of web based services should be:

- (1) Current network performance for the case of web browsing should be updated in the existing recommendation.
- (2) Customer's perception should be accessed via key metrics of application and networking for web browsing.
- (3) Human psychological effects such as *Content* and *SOM* should be integrated and combined together with (1) and (2) for a comprehensive assessment of customer QoE for web browsing, fully understanding and further matching with customer behaviour when they are web browsing.

## **2.4.2 Objective-Subjective link in QoE of Web based services**

### ***2.4.2.1 QoE-QoS mapping***

In 2002, Khirman, S. and Henriksen, P. [51] discussed a relationship for objective networking service conditions and an objective user perception to measure QoE. Human satisfaction of HTTP service or web browsing is affected by two main network QoS parameters those are bandwidth and latency. The results of their analysis concluded that those factors represent a crucial role in end-user satisfaction.

QoE Web is one of Specific Joint Research Projects of The Network of Excellence Euro-NF (Network of Future). The original idea to assess QoE for Web is based on the general shape of QoE-QoS mapping in [27], in which the  $QoE_4$  of a transmitting website is a function of a QoS parameter. For example, MOS is matched with an exponential function of weight session time as described in Equation(2.4).

$$QoE_4 = f_{\text{exp}}(x) = 4.298 * \exp(-0.347 * x) + 1.390 \quad (2.4)$$

Where  $x$  is session time, and  $f_{\text{exp}}(x)$  is QoE of a transmitting website.

#### ***2.4.2.2 Characterizing the meaning of QoE related to Web based services application***

Obviously, any requirement for a new QoE model should incorporate the merits of previous methods in order to make the model suitable for packet traffic, and especially for web traffic. The QoE for a user of a new multimedia service depends not only on network features but also on higher layer application characteristics. Based on this observation, a method for objectively assessing QoS is needed.

Din, I., et. al. presented the effects of packet loss on web traffic and correlated them to a user's behaviour [59]. They estimated the user's perception and measured packet loss and delay on network performance. Then, packet loss and delay on web traffic are monitored to detect a user's QoE. This paper utilized passive models for detecting packet loss. The authors have demonstrated different parameters based on traffic characteristics to relate them to user actions. Packet loss can be detected by observing retransmission at the measurement point or out-of-sequence packets. However, methods for detecting retransmissions lead to negative effects as a result of malfunctioning devices, routing loops, or network duplicates. On the other hand, methods of out-of-sequence detection were also considered inaccurate as they might be produced as a result of balancing load, parallel processing on a router, or route oscillation in the network. To overcome this problem, the authors described a method for detecting loss at the other end's host. This method not only filters out the different sorts of retransmission, but also provides a lower bound on the number of packet losses.

The results from the effects on connection duration and size are partially analysed in this paper [59] as: 1) decreasing the average connection duration produces a lower loss rate and 2) increasing both the time and size produces a decreasing loss rate. The first effect is caused by different factors such as the patience of the user so that they do not terminate their connection instead of a high loss or the user may open many connections and wait for each response. However, surprisingly the second effect does not show any “impatience” indication from the users. It is concluded that a connection duration and size do not represent the user’s behaviour. The author proposed a method for measuring inter-arrival times and increments the number of destination IP addresses accessed by a host as shown in Figure 2-9 extracted from the paper. Figure 2-9 plots the number of destination IPs accessed, user interruptions in different average inter-arrival times and packet losses. The left y axis shows the number of destination IPs accessed and the number of user interruptions. The right y axis describes the average inter-arrival time of connections in seconds. The x axis is the number of packet lost.

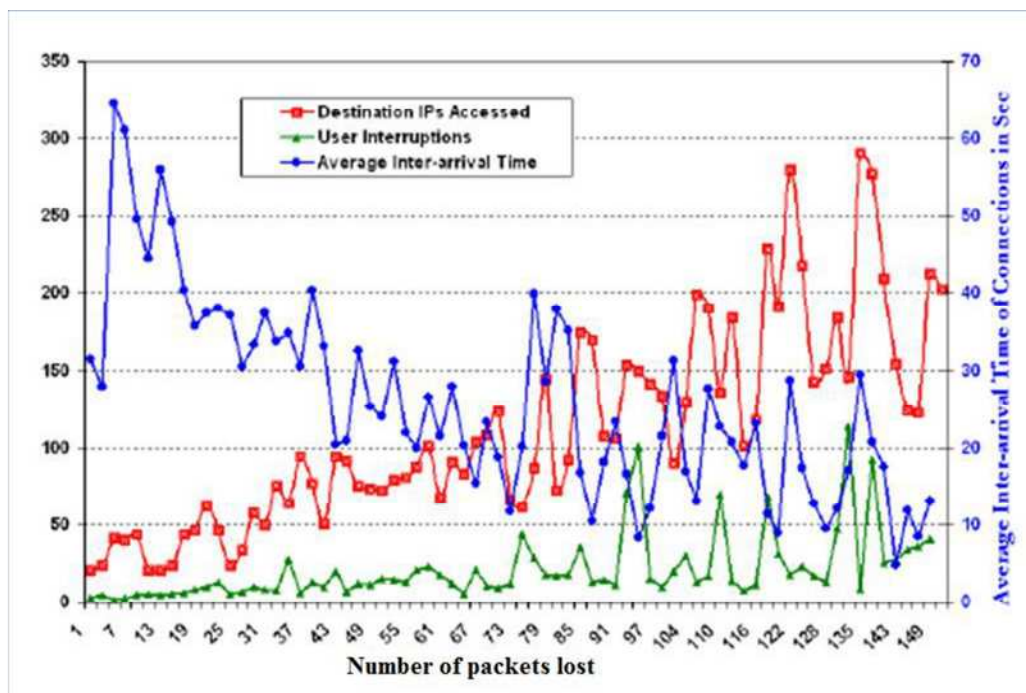


Figure 2-9: Average connection duration, Number of packets lost and User Interruptions vs. Loss Rates ([27])

It can be seen that the current results based on traffic characteristics may give a new and clear approach for a web-QoE model, although the analysed web traffic characteristics need to be updated because they are not quantitative according to the authors.

Therefore, correlating a user's behaviour with the effects on traffic characteristics are part of on-going research.

### 2.4.3 Analysis of psychological perspectives related to QoE

According to the different QoE-QoS approaches in Figure 2-6, both objective and subjective psychological measures are expressed in QoE in general.

In a recent study, Hoßfeld, T. et al., explain the impact on QoE from a psychological perspective in which the memory effect is an implication to measure QoE Web [32]. In this paper, the psychological influence factor of a user's past experience is mentioned and referred to as the *memory effect*. The previous user's score affects the current user's score to assess the download time for web pages.

### 2.5 Summary and Discussion

Modelling QoE for web browsing is a challenging topic as it involves many different factors related to human as well as technical issues. To understand such a large number of human and technical metrics is both difficult and tricky.

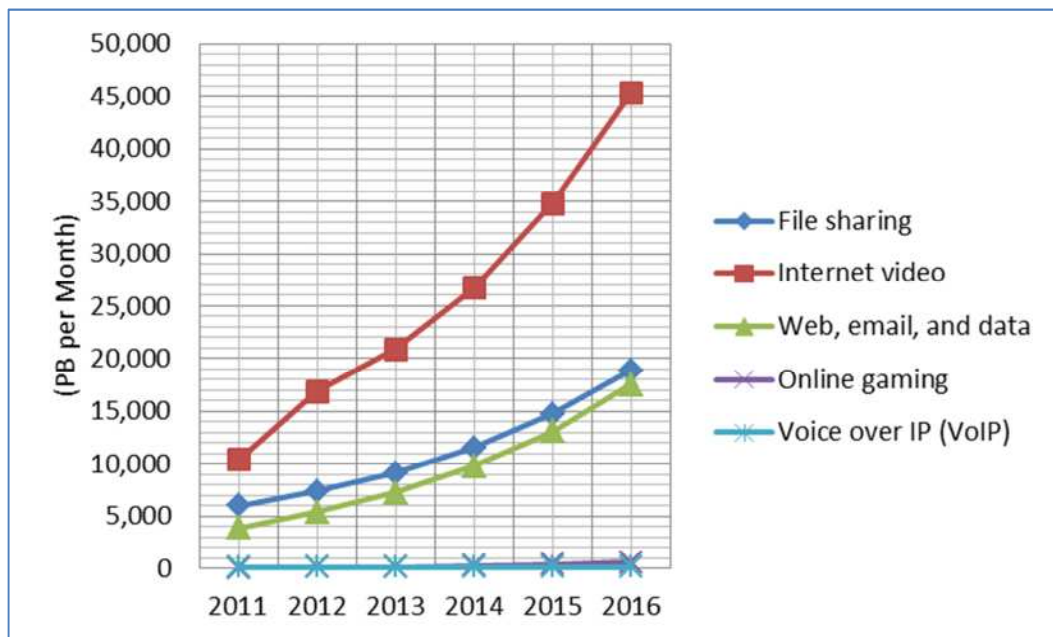


Figure 2-10: Global Consumer Internet Traffic in period 2011-2016 (Source: Cisco)

Furthermore, QoE of Web based services should be paid much attention as HTTP based traffic for web-browsing applications have become the majority traffic of both residential broadband Internet and commercial usage. According to Cisco's predictions,

Internet video and web, email and data which are mainly web traffic will be the most frequent flow in the future network during the period from 2011 to 2016 [62]. Figure 2-10 shows the most frequent global consumer Internet from 2011 to 2016 that are Internet video, and web, email and data shown in the red and green lines respectively.

Additionally, there are three reasons why researchers need to work on QoE- modelling for web based services:

- 1) Firstly, a Web-QoE model needs to be developed as it has not properly existed so far in any ITU-Recommendations. As mentioned above, the only existing recommendation of the ITU-T, G. 1030 needs to be updated.
- 2) Secondly, the model needs to have a reliable measurement regime to improve user satisfaction and QoS: As mentioned in Section 2.4, recent studies mentioned are partially in one of three categories viz: QoS-QoE link, Web browsing characteristics and psychological effects. This shows that a comprehensive model QoE of Web browsing or Web based services is a significant research opportunity.
- 3) Finally, improving the quality of offered services as perceived by users is significant and challenging for any service provider who would like to minimize customer churn and to maintain their competitive edge, otherwise they risk being out of the market competition for providing good service to their users.

Recently, the ITU-T Study Group 12 [31], the lead study group on Quality of Service, highlighted their current work in the development of the modelling of network/terminal configurations and the predicting of the user impact of associated impairments, and revising standards on the planning and deployment of IP-based networks during a period ranging from 2013 to 2016. The ITU-T Study Group 12, again emphasized the need to measure new significant parameters as packet loss, delay or jitter and to know their user impact that are customers expecting the QoS of traditional communication services. Thus, their plan includes a standard revision of IP-based network planning and deployment during that period of time.

From the current main literature review, the results of analysing traffic characteristics [59], suggests that researchers should have a fresh look at a Web-QoE model which does not rely totally on MOS but can still have comparable results with MOS. However, the way that people react to an application is different from person to person. We might

say that these results are only suitable for typical users who know how to use the web service, at least. Another drawback is that it is based on a large amount of data, which is obtained from the network, in order to detect user behaviour effectively.

Studies in [27, 51] relate objective networking service conditions and subjective user perception in order to measure QoE for users, and a suitable approach for establishing the relationship between QoE and QoS is confirmed for telecommunications services in general and can be applied for web-browsing in particular. In addition to this, [58] confirmed that QoS, QoE, user action, and end user satisfaction can be established as a closed relationship loop.

It is difficult to represent the features of the various services from only the bandwidth and latency time in an integrated network environment; however, based on the application characteristics of multi-media services, the analysis of traffic characteristics can be done, but they can potentially take a long time.

Based on the exact results from MOS grades, a new model should have network-related performance criteria that correlate well with MOS grades [63]. As mentioned above, a new model should avoid the pitfalls observed for an MOS that is based on scores and user surveys and the complexities associated with the E-model. “MOS is not enough” to assess QoE [26]. Therefore, action-dependent user perception that is received from application level performance characteristics and from user-learning mechanisms should be considered. In addition to this, detecting packet traffic should include both passive and active measurements in near real time processing. With currently faster computational techniques, using no reference but increases in accuracy in QoE assessment method should be possible.

Acknowledgement of results for Web QoE in The Network of Excellence Euro-NF [27, 32], show that integration of both objective QoS and psychological effects should be mentioned. Objective measurement of QoS should not be the only parameter. The interaction of different factors in the categories mentioned in Section 2.4 should be topics for a future research. Parameters of any new model must be objectively measured, but take into account subjective psychological human effects to understand the observed MOS. The model must be aligned to MOS or similar subjective measurements of service quality for the purposes of consistency. However, a variety of

subjective psychological effects such as user motion, memory, boredom and etc., can impact on QoE for Web applications in addition to objective networking performance. For example, human memory phenomena as reported in [64], the authors explained and verified that such recognition conditions as an old-new recognition task and a remember-known recognition task, and cued recall will lead to a recall error and recognition misses which may impact directly on users' experience and their scores.

To be objective, the subjective psychological effect requires further techniques which depend on special types of psychological effects. For example, a video processing technique could be used to detect user motion [65], or by using video behaviour profiling to detect anomalies [66].

## Chapter 3 Experimental Design

*In this chapter, a method and procedure for observers' tests to assess QoE are described. Procedures for collecting users' scores and controlling the network performance in the test bed are presented in a step by step manner.*

*As it is not possible to find large sample size in order to investigate the impact on parameters thought to be relevant, our experiments were designed around the Taguchi robust method which ensure that the experiments can cover all important combinations of those parameters.*

*As it is difficult to do with respect to the parameters that were investigating in the real network environment, it had been to undertake the study with available subjects - who were the group that I eventually secured at the test-bed of networking lab located at Massey University.*

*As different networking environments related to different parameters were tested, the network emulator was situated in the topology to control the networking performance.*

*The web server had been located in the topology to provide the web services. Each webpage was popped out for web browsers to surf and provide their perception about the quality.*

### 3.1 Experimental Design

#### 3.1.1 Taguchi quality method approach

Taguchi refers to experimental design as “off-line quality control” [67] because it is a method of ensuring good performance in the design quality into products and processes [68]. In the robust design, it is emphasized that high-quality products can be produced quickly and at low cost [69, 70].

As it is not possible to find large sample sizes in order to investigate the impact on parameters thought to be relevant to QoE, our experiments were designed around the Taguchi robust method [70]; which aims to make a process less variable in the face of variation over which we have little or no control and ensure the number of controlling

experiments will be satisfied for the case of limited resources such as a minimum timeframe and number of subjects.

The Taguchi approach is a collection of principles which construct the framework of a continually evolving approach to quality [71]. Taguchi divided the factors that affect a system into two categories: control and noise factors. Control factors are the main effects of a system, and easily set by the experimenter. Noise factors cannot be controlled in real life and their influences are not known.

However, it is very expensive and exhaustive to collect the information from control factors in a real situation. This is due to the likelihood of getting a huge number of combinations from the control factors each with different levels. Therefore, special arrays called orthogonal arrays have been developed by Taguchi to reduce the number of controlled experiments while ensuring that these experiments are able to gather dependable information. By another way, these standard orthogonal arrays can be used to conduct a minimal number of experiments that can give information about all factors that have an effect on the outcome.

The term orthogonal array was used by Dr. Genichi Taguchi [72], in the 1980s when he discovered a method of robust design for quality that is now commonly referred to as Taguchi's Robust Design.

An orthogonal array is used to gather dependable information around control factors with an acceptably small number of experiments. These standard orthogonal arrays constructed by Taguchi are a special set of mutually orthogonal Latin squares. A Latin square is an  $k \times k$  array with  $k$  different characters occurring once in each row and column. A mutually orthogonal Latin square is a set of two or more Latin squares of the same order, all of which are orthogonal to one another. Orthogonal Latin squares were used to create compiler tests by substituting the values of each Latin square with relevant parameters; different necessary combinations of those parameters were then created. Thus, all covered pairwise and combinations are created in a series of tests by setting up an orthogonal array with testing parameters and appropriate values [73].

To define an orthogonal array, three questions need to be answered:

- 1) The number of factors,
- 2) The number of levels for each factor
- 3) The specific two factor interactions to be estimated.

The orthogonal arrays need to be chosen and used based on situations where there is no concern that an interaction with a specific column or columns will lead to confusion [71].

Three orthogonal arrays have been recommended by Taguchi and chosen for use in our experiments depending on the number of controllable factors that have been chosen in each case. The recommended orthogonal arrays used in our experiments are:  ${}^1L_{18}(2^1 \times 3^7)$  that allows us to study one control factor having 2 levels ( $2^1$ ), usually in the first row and the effects of up to 7 control factors where each of them has 3 levels ( $3^7$ ), located from the second to the eighth rows. The columns from 1 to 18 correspond to the number of combinations of values.  ${}^2L_{25}$  with 25 random combinations using two factors and each factor has five levels.  ${}^3L_{36}$  with 36 random combinations using two control factors at 3 levels with 4 noise factors.

For example, 1 control factor with 2 levels and 7 control factors with 3 levels for each control factor normally will require 686 ( $= 7^3 \times 2$ ) combinations of control factors. This means that there will be 686 experiments required. However, using the Taguchi approach, 18 required experiments are acceptable as described in Table 3-1 which is Orthogonal array  $L_{18}(2^1 \times 3^7)$  extracted in ref. [72] As shown in Table 3-1,  $L_{18}(2^1 \times 3^7)$  allows us to study one control factor having two levels ( $2^1$ ), usually in the first row, and the main effects of up to seven control factors where each of them has 3 levels ( $3^7$ ), located from the second to eighth row. The columns from 1 to 18 correspond to the number of combinations of values [71].

Table 3-1: Orthogonal array  $L_{18}(2^1 \times 3^7)$  extracted in ref. [72]

Number of Run	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	
1 two-level factor	1	1	1	1	1	1	1	1	1	2	2	2	2	2	2	2	2	2	
7 three-level factors	2	1	1	1	2	2	2	3	3	3	1	1	1	2	2	2	3	3	3
	3	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3
	4	1	2	3	1	2	3	2	3	1	3	1	2	2	3	1	3	1	2
	5	1	2	3	2	3	1	1	2	3	3	1	2	3	1	2	2	3	1
	6	1	2	3	2	3	1	3	1	2	2	3	1	1	2	3	3	1	2
	7	1	2	3	3	1	2	2	3	1	2	3	1	3	1	2	1	2	3
	8	1	2	3	3	1	2	3	1	2	1	2	3	2	3	1	2	3	1

As recommended by Taguchi [71], using an orthogonal array of  $L_{18}(2^1 \times 3^7)$  allows covering all the important combinations of those relevant parameters while reducing the actual size of 686 into 18 runs.

### 3.1.2 Factor analysis

#### 3.1.2.1 Noise factor

Noise factors are effects on a system that cannot be controlled in real life, and their influences are not known. In the QoE assessment for web based services, *Content* is chosen as a noise factor because it cannot be controlled in the real World Wide Web in terms of users' choice. Users may surf any web site containing any wanted content for their purpose. *Content* has its influence on a QoE assessment at the time of user's perception, but its effect is not clearly known. During the course of our experiments, the noise factor and levels used are shown in Table 3-2.

Table 3-2: Noise factor and level

Level	1	2	3	4
Content	Text	Image	Image/Flash Image	Text/Image/Flash Image/Video

### 3.1.2.2 Controlling factors

Controlling factors are the main effects of a system under our direct management. They are easily set by the experimenter. In the QoE assessment for web based services, we propose a novel approach to use a QoS measurement via a network derived metric while the traffic characteristic-derived metric will be based upon objective assessment of user behaviour in response to the web service that they are using.

In our approach, we formulate a model that brings together two categories of metric, a network derived metric, and an application-derived metric, to form an appropriate QoE metric, thus, the controlling factors are taken from networking and application parameters such as delay, and the requests per second. The network-derived metric estimation (delay) is responsible for network feature estimation. Application-derived metric estimation incorporates the number of requests per second as these are responsible for application level performance.

Table 3-3: Level and Metric correspondences

Level	Delay	Requests per second
1	5	Unlimited/15500
2	10	1000
3	100	500
4	200	100
5	500	5

The performance parameters were changed for different controlled environments. The delay and the requests per second are set at different levels as shown in Table 3-3 and users are not informed about the nature of the changes that are being made.

### 3.1.2.3 Human factors

During the course of running these experiments published in [74] [75], some psychological effects have been observed as shown in Table 3-4. It is clear that QoE

assessment for web based services has been subjectively affected by human factors relating to the state of mind of the subjects when they undertook the experiments.

Table 3-4: Psychological effects during the course of experiments

Name	Characteristics	Sources of Studies
Boredom	Increase level over time	Interview/Feedback
State of Mind	Random change	Feedback

## 3.2 Test-bed and Controlling Performance QoE Web Assessment

### 3.2.1 Test-bed experiment setup

The experiments were carried out using the networking infrastructure using a test-bed located at Massey University in Palmerston North. The network topology of the test-bed consists of seven routers and seven switches as shown in Figure 3-1.

The network topology for the test-bed consisted of seven cisco routers, seven cisco switches, an IBM server, and a Web Server. There are 12 personal computers using the Windows™ XP operating system. The subject uses a browser such as Internet Explorer 8.0. The IBM server runs on a Windows Server operating system. The web server runs on the Ubuntu server operating system, version 12.04.1 LTS [76]. The switches belong to the Catalyst 1950 and 1924 family [77]. The routers are Cisco 1710 [78]. The routers and switches run under the Cisco original Internetwork Operating System (Cisco IOS) software release 12.2 [79]. The IP addresses are version 4 (IPv4) [80] and divided into 6 sub-networks as described in Figure 3-1.

It should be noted that this is a controlled environment in which the network is isolated from the true Internet and network performance is carefully manipulated in order to enable systematic observation of user behaviour without problems caused by miscellaneous external events that might distort the results of our measurements.

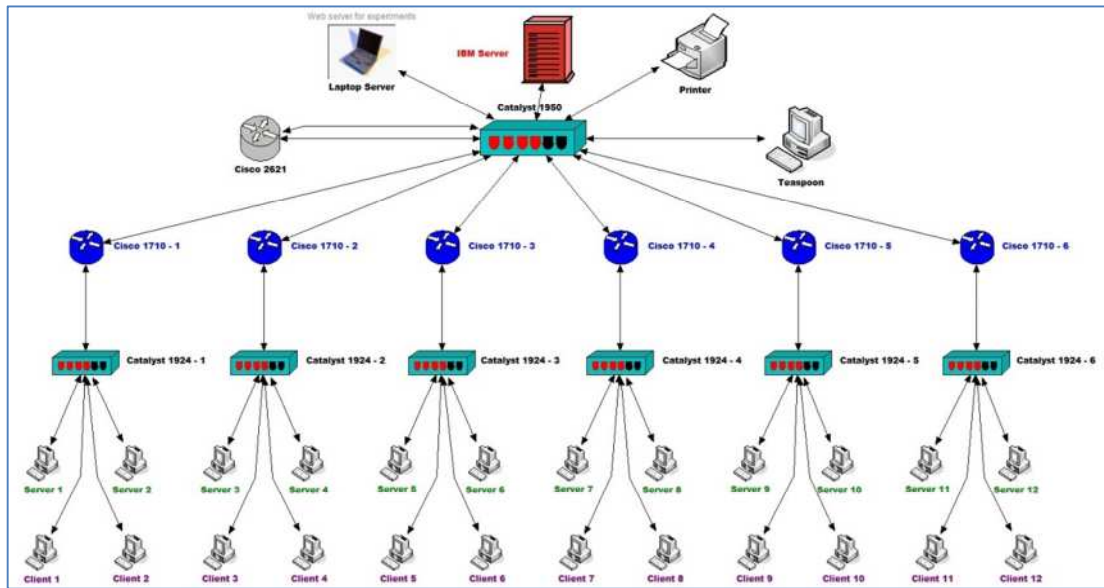


Figure 3-1: Network Topology for the Test Bed

The experiment involved viewing web content on a browser where the content and delivery of that content have been subjected to parameter modifications that are expected to influence user perception of the web service.

### 3.2.2 Controlling Session

The experiments have been performed via different sessions in which the controlling factors and noise factors have been investigated. In each experiment, a session has been defined as consisting of two, three or five minute intervals.

The first controlling session was used for a pilot study, and it actually did not follow the Taguchi method in this case only. The purpose of this pilot study was to study the relationships among network performance, user action, and the actual mean opinion score. The results were targeted at helping us with calibrating the main strong impact factors in our QoE model. The outcome of the pilot study indicated that it is possible to measure QoE for web surfing in combination with MOS scores, network parameters such as delay, and user's actions such as an increase in the number of requests [74].

Its controlling session involved five randomly chosen performance activities conducted over a period of approximately 25 minutes during which time the networking conditions were changed at approximately five minute intervals, producing a total of five separate sessions. The structure of the experiment is depicted in Figure 3-2 below.

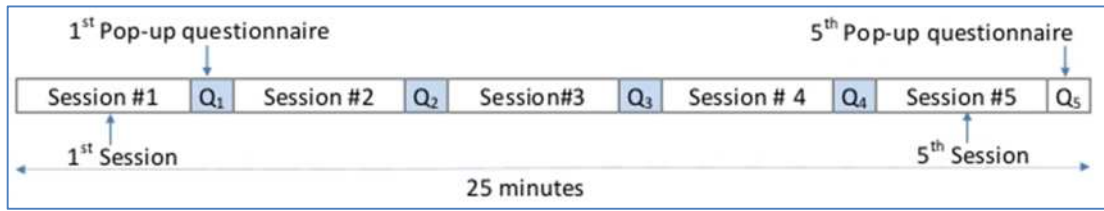


Figure 3-2: Network Session of the Pilot Study

After the first pilot study, all the designs of our controlling sessions followed the Taguchi design methodology. Depending upon the purpose of the subjective test, the levels of controlling factors were increased or decreased, but we kept at least three relevant levels of low, medium and high in the experiment. The design of each controlling session involved a step-by-step process as follows:

- 1) Defined controlling factors and their levels
- 2) Defined noise factors and their levels
- 3) Defined the orthogonal arrays needed
- 4) Defined the number of runs needed
- 5) Finally, defined the relevant controlling session to perform the tests

The second controlling session was carried out with three controlling factors of delay, packet loss and requests per seconds at 3 levels (low, medium and high level), a two-level controlling factor of response time and a four-level noise factor of Content. The recommended orthogonal array by Taguchi [71] was  $L_{18}$  with 18 runs needed for the controlling factors, and performed a full test on the noise factor of content, thus 54 different sessions were needed as described by figure below. The second controlling session ran for a period of approximately 120 minutes with 54 different individual sessions.

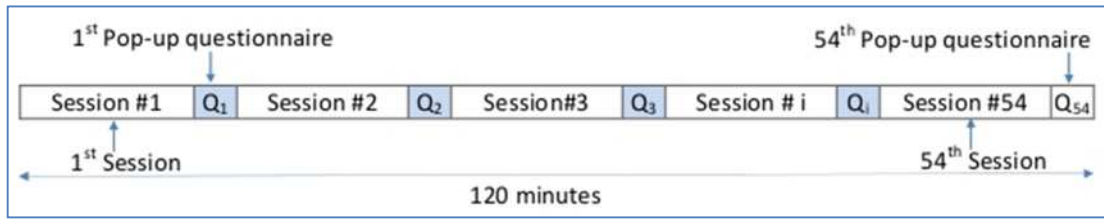


Figure 3-3: The second Network Session by Taguchi approach

The third controlling session was carried out with two five-level controlling factors of delay and requests per second (3 levels (low, medium and high level), and 1 level between low and medium level, and 1 level between medium and high level) and a two-level noise factor of Content. The recommended orthogonal array by Taguchi [71] is  $L_{25}$  with 25 runs being required for the controlling factors, and it performed a full test on the noise factor of content, thus 50 different sessions were needed as described by Figure 3-4. The third experiment ran for a period of approximately 115 minutes with 50 different individual sessions.

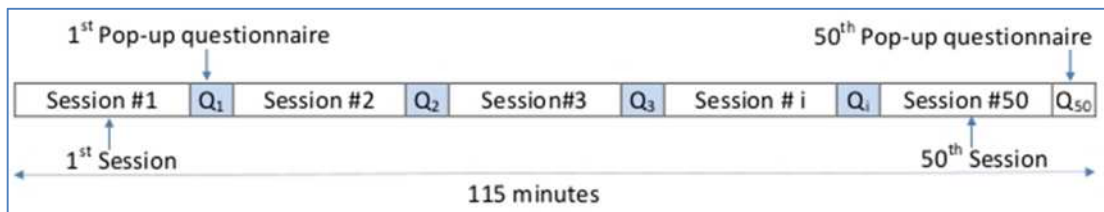


Figure 3-4: The third Network Session by Taguchi Approach

The fourth controlling session ran for a period of approximately 108 minutes with 36 different sessions in which each session was changed at roughly three minute intervals. Its structure is shown in Figure 3-5.

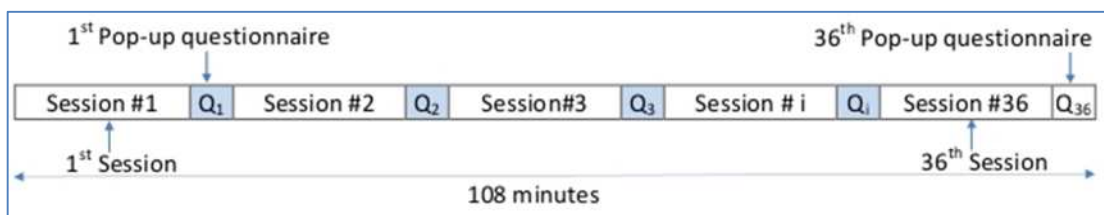


Figure 3-5: The fourth Network Session by Taguchi Approach

The fourth controlling session was carried out with two three-level controlling factors of delay and requests per second (3 levels of low, medium, and high), and a four-level noise factor of Content. The recommended orthogonal array by Taguchi [71] is  $L_{36}$  with 9 runs being needed for the controlling factors, and performed a full test on a four-level

noise factor of content, thus 36 different sessions were needed as described by Figure 3-5.

### 3.2.3 Networking performance configuration

#### 3.2.3.1 Controlling Network performance

The network performance was changed for each of the different controlled environments by utilizing NetEm [81, 82], a package which is available for the Linux Kernel [83].

Figure 3-6 describes a Linux queuing discipline [84, 85] in which the queuing discipline applies between the protocol output and the network device. In a simple way, the packets from the transportation layer are queued to enter the network interfaces based on selected queuing policies. The queuing discipline specifies a policy to decide which packet will be sent based on the current setting.

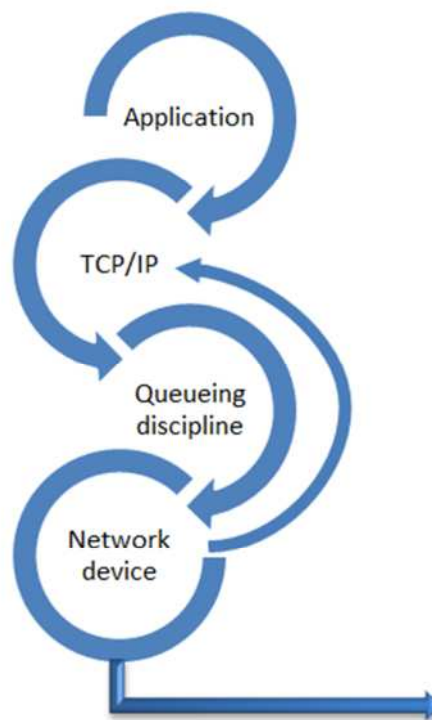


Figure 3-6: Linux Queuing Discipline

NetEm [81, 82] is a form of network emulation developed for testing protocols by emulating the properties of a wide area network. Basically, NetEm is a queuing system as described by Figure 3-7.

The aim of using NetEm is to emulate the network delay behaviour effect. All traffic enters at the network interface, Ethernet 0. When there is no additional parameter, the queuing discipline is a basic First-In-First-Out (FIFO) waiting queue [86].

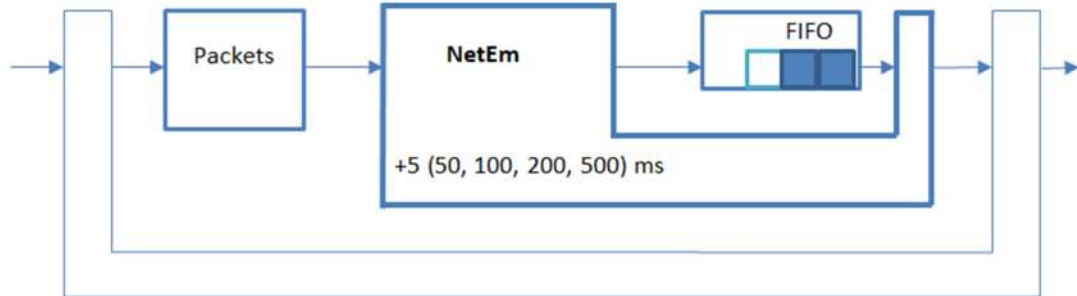


Figure 3-7: Queuing discipline for Ethernet 0 [82]

Overall, our experiment ran over periods of approximately 25 minutes, 120 minutes, 115 minutes and 108 minutes respectively; during which time the networking conditions were changed at approximately five, three, and two minute intervals.

### 3.2.3.2 Controlling Application performance

The number of requests per second and conditional rules were controlled by using `mod_qos` [87] which is the quality of service module for the web apache server [88, 89]. `Mod_qos` is a control mechanism for managing web resources with different policies. For example, it manages how many IP addresses can be accessed at the same time, or how many concurrent requests per second are allowed and so forth.

The current release of `mod_qos` 10.13 mentioned in [87] can perform as follows:

- 1) Maximize the number of concurrent requests to a location/resources
- 2) Limit the bandwidth
- 3) Limit the number of request events per second

Condition rules, events, request level control, client level control, connection level control, request level control, and variables are configured on a per-server basis.

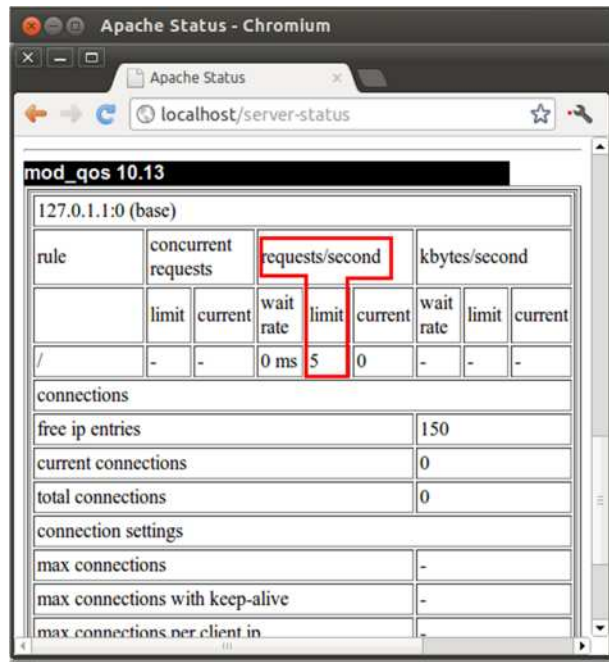


Figure 3-8: Example of setting in mod\_qos

Mod\_qos is an open source software product licensed under the GNU general public licence [90, 91], its download is managed by SourceForge.net. The first version of mod\_qos was released in 2007[87], and subsequently included in the Ubuntu distribution [92]. The number of requests to an http application was limited by the mod\_qos module in order to ensure application availability. One example setting is described in Figure 3-8 when the total number of requests per second allowed to access the web server is 5 requests per second.

### 3.2.3.3 Controlled values

In the pilot study, networking performance, the delay and the packet loss were tested using values as shown in Table 3-5.

Table 3-5: Values of delay and packet loss for the first experiment

Delay (ms)	Packet loss (%)
50	5
100	5
500	10

In the following experiment, more control factors for both networking and application performance were included, such as delay, packet loss, the number of requests per second and response time. Their values are reported in Table 3-6 below.

Table 3-6: Values of control factor for the second experiment

Control factor	Level		
	1	3	5
Delay (ms)	50	100	500
Packet loss (%)	5	10	15
Request per second (req/s)	15500/Unlimited	500	5
Response time (level)	Improve	No	---

In the next experiment, the factors of response time and packet loss have been removed. The chosen factors from networking and application performance, the delay and the number of requests per second are selected from the second experiment but increasing levels are assigned to each factor whose controlling values are presented in Table 3-7.

Table 3-7: Values of control factors for the third experiment

Control factor	Level				
	1	2	3	4	5
Delay (ms)	5	50	100	200	500
Request per second (number)	15500	1000	500	100	5

In the final experiment, the networking and application metric derived factors of delay and requests per second have been kept the same as for the previous experiment. In this case, the runs have been controlled to test human factors that will be described in more detail for human factors as described in Sections 4.2.4 and 4.2.5. Depending upon the purpose of the subjective test, the controlling factor is the same, limited run with delay levels of 1, 3, and 5 and request levels of 1, 4, and 5 (as the same as for Table 3-7 above).

The last experiment is focused more on the test of the content noise factor. Therefore the subjective tests have been tested on low, medium, and high levels of controlled factors, and full test on all levels of noise factors of content.

### 3.3 QoE Web content design

Content is the embedded information on website [93]. Alternatively, content is broadly described as “the stuff in a website, this may include documents, data, applications, e-services, images, audio, video, or achieved email messages”[94].

In our experiments, the contents of any particular web page typically may include text, images, videos or other multimedia content such as flash objects, online games, and Java script objects. The nature and content of the web pages dictates the elements that are presented on the web. The websites consisted of four different configurations which were categorized based on their content as described in Figure 3-9 and Figure 3-10.

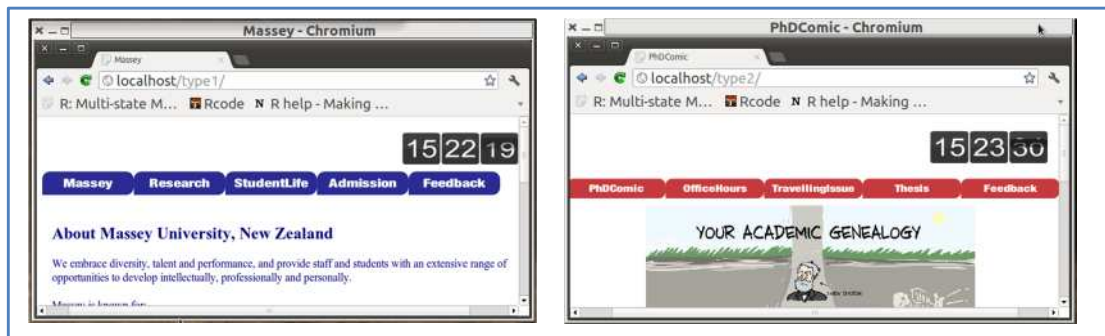


Figure 3-9: Type 1 (left) and Type 2 (right) of Content

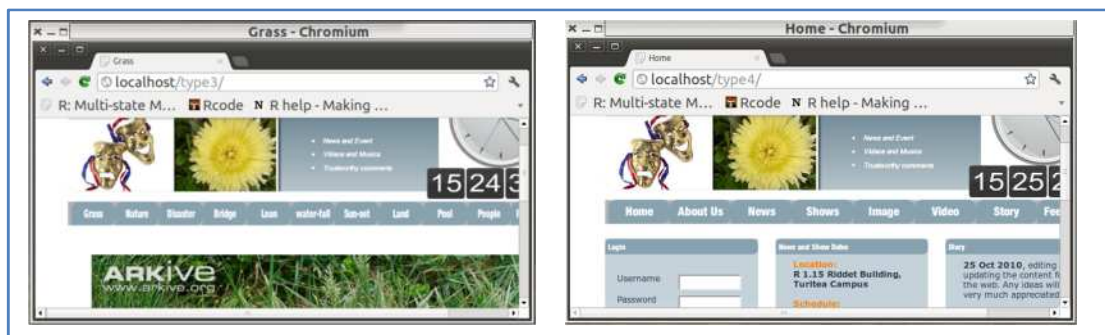


Figure 3-10: Type 3 (left) and Type 4 (right) of Content

The websites are designed to cover both light and heavy payloads. The first website contained text only and was classified as “Noise 1” - based on the Taguchi approach.

The second website contained pictures that were classified as “Noise 2”. The third website contained pictures and flash objects that were classified as “Noise 3”. The fourth website contained text, pictures, flash objects, and videos that were classified as “Noise 4”.

The participants were not informed about the changes in performance quality or content changes that were to be implemented during the experiment.

The webpages were coded using HTML [95, 96] and the PHP language [97, 98]. The webpages themselves included both static and dynamic content [99]. HTML is the basic language for coding static web pages while PHP code was used for dynamic web pages; Dynamic web pages are generated by a web application, for example a dynamic web page of “[feedback.php](#)” for an interaction with table of “[result\\_table](#)” in the MySQL database for saving users’ scores. The web browser in the test bed was Internet Explorer 8.0, the browser has been setup without caching; thus, the web pages will be freshly loaded from the server whenever users accessed a page.

## 3.4 Data Collection

### 3.4.1 MySQL Database setup

A standard MySQL database [100] was chosen and setup to save users' scores, comments, and relevant data for the web. The collectable data from the system were a users' score, state of mind, user comments (optional). Other additional data is sourced from built-in-options such as login and a chat forum on the web pages.

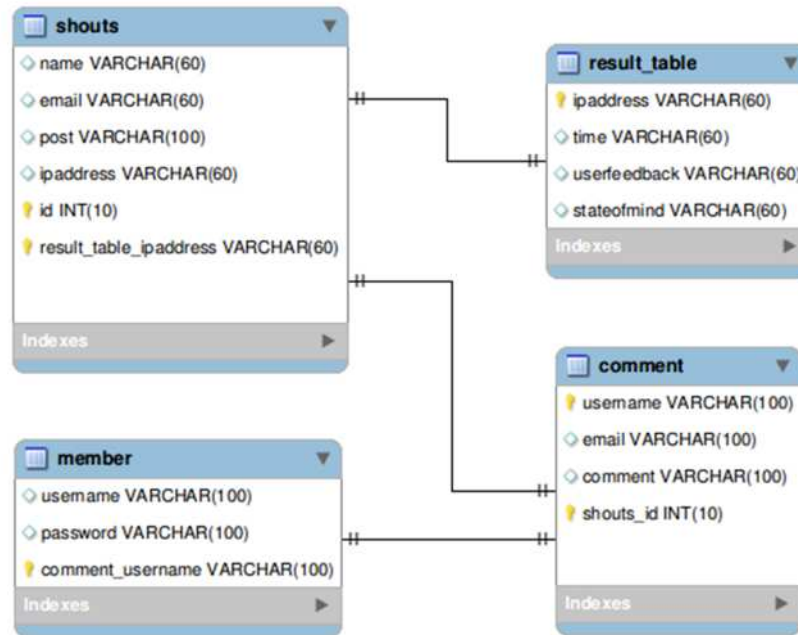


Figure 3-11: MySQL database

The database is built using the MySQL database server which contains four tables described in more detail in Figure 3-11. The database contains four tables of “**result\_table**”, “**comment**”, “**shouts**”, and “**member**”. The table of “**result\_table**” has fields of **ipaddress**, **time**, **userfeedback**, and **stateofmind** in order to save the IP address, time, user’s score and user’s state of mind respectively. The table identified as “**comment**” has fields of **username**, **email**, and **comment** in order to save user comments (optional). With this comment option, users can enter any comments together with their numerical scores. The tables of “**member**” and “**shouts**” are created for additional built-in-options in the web sites such as login and a chat forum.

### 3.4.2 Human-Subjects

There were 12 users involved in our pilot study, 21 users took part in the second experiment with a three-level state of mind and two-level content factors integrated, and 12 users took part in the third experiment with a three-level state of mind and four-level content factors integrated.

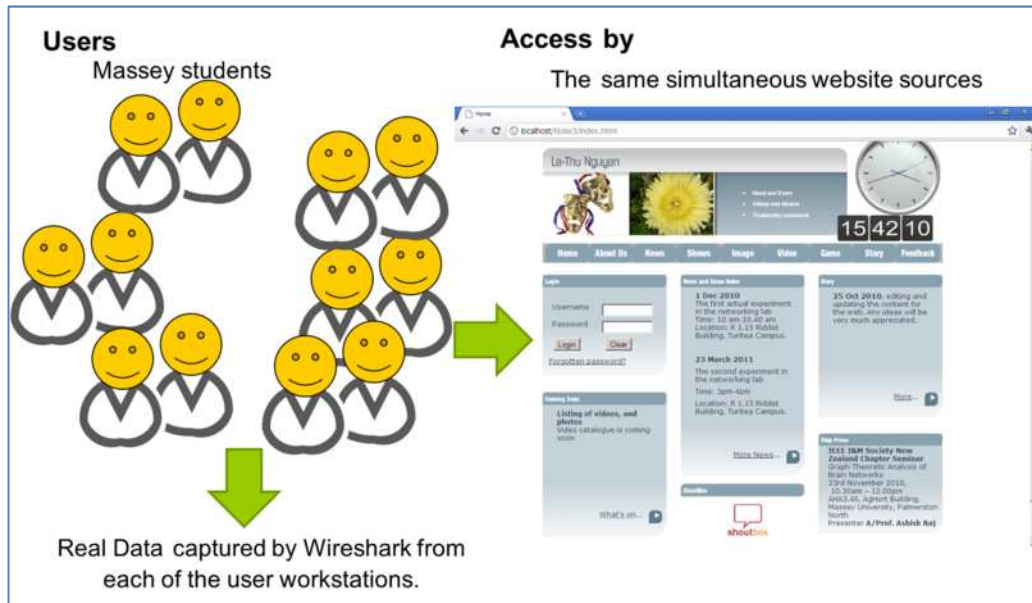


Figure 3-12: Setup of users and website sources

All of them were post-graduate or undergraduate students studying at our university. They consisted of a mixture of individuals who have a computer background (e.g. Computer Science) and those coming from non-computer backgrounds (e.g. Food Science, Agronomy). Of the 45 users, there were 5 taking undergraduate studies and 40 involved in postgraduate study. Based on their sex, there were 14 female and 31 male. They accessed the same website sources simultaneously. In this experiment, the impairments that were applied were the same for all participants and experienced simultaneously.

While users accessed the website, the users' actions were captured to see the behaviour that they exhibited through their selection of hyperlinks on the displayed page, together with the frequency and nature of these selection activities.

### 3.4.3 User score-MOS

The users' scores were collected from them in the form of Mean Opinion Score (MOS) recommendations [12, 101] which are based on the traditional five level scale

{1,2,3,4,5} that links these numerical values to a user’s perception as defined by {Not recommended, Dissatisfied, Fair, Satisfied, and Very Satisfied}.

When users indicate a score of 5 or “Very Satisfied”, it means that performance quality from the web server was excellent as the user expected. Similarly, when users enter a score of “4”, “3”, “2”, or “1”, it means that performance quality from the web server was good, normal, poor, or unacceptable respectively.

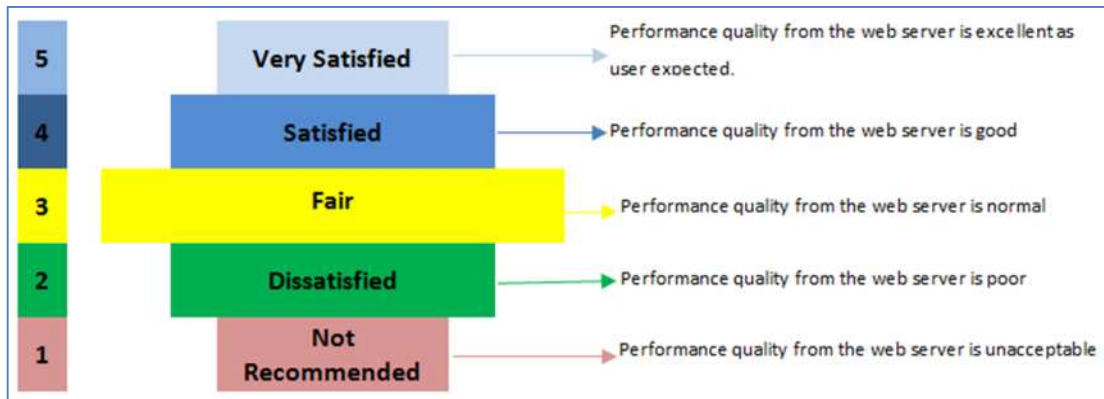


Figure 3-13: MOS and its score meaning used in the project

### 3.4.4 User tasks

Users were asked to give an evaluation of their own experiences at two to five minute intervals through the use of a user scoring system for the particular web pages that they were surfing under different networking situations. Users gave their evaluation by sending their users’ scores via a feedback page accessed via the pre-prepared web pages.

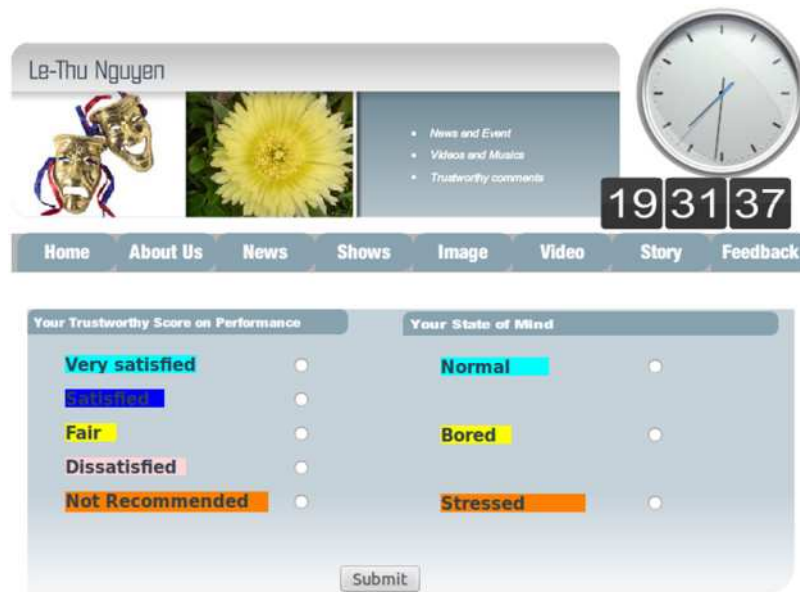


Figure 3-14: Feedback page

The feedback page described by Figure 3-14 was designed as a radio button selection; Users selected their preferred scores: one for an overall performance of system, and one for their own state of mind, using two single clicks on the two options and a single click on the submit button. Therefore, it is expected that this task would not involve much time for the users in sending their user feedback experience.

These users were asked to access the prepared local websites; then to rate:

- 1) Score on performance: This score is to rate the web site performance as described in Figure 3-14 for each of the objects and to record a score based on their perception. The meaning of each score is described in detail in Figure 3-13.
- 2) Score on your state of mind: This scores show users' feelings during the experiment whether or not they feel Normal/Bored/Stressed during the time they have to do with the experiment, rather, it should represent their general mood for the day and how it might be changing as the experiment proceeds. In more details, do they feel "Bored/Normal/Stressed" when they have to repeat various actions during the experiments. Please note that this score is not intended to be related to the networking performance when the users choose an option for their state of mind.

### 3.5 Summary

The experiment has been carried out to collect the users' scores and assess objective and subjective factors which affect their QoE for the case of web based services. The output of this chapter is the subjective users' score related to objective controlling network parameters which will serve as an input for our QoE assessment in the next chapter.

The experiment has been designed using the Taguchi quality method approach which ensures that a quantitative number of observations were carried out. As noted earlier in this thesis, it consumes significant amounts of cost and time when using a substantial number of participants in any research involving human participants. However, by carrying out the experiments using the Taguchi method, we can mitigate both time and cost as well as collecting accurate quantitative observations.

The controlling sessions have been designed in order to limit boredom and memory effects that were mentioned in Section 3.2.2. The control factors and noise factors have been assigned based on Taguchi's guidelines.

The format of the web design has been described and the elements of the web pages outlined. A database based on MySQL server has been built to save data for the purposes of both collecting users' scores, and saving additional built-in data related to the various web functions.

User tasks were designed to be as simple as possible; therefore, participants would not be annoyed when asked to achieve their respective tasks.

## **Chapter 4 Modelling Approach for the QoE of Web based services**

*In this chapter, a framework is presented for assessing QoE web based services. The framework arises from a conceptual model whose metrics are derived from the network and application layers and integrated with human factors such as the user's state of mind. The objective metrics derived from these two layers have been selected and tested using appropriately constructed experiments to see their effects on users' perception. Additional subjective metrics such as boredom, state of mind based upon the effects of the content have been identified and then taken into account during the course of the experiment. The final QoE web assessment framework has been developed using a comprehensive set of these metrics in order to construct a full and detailed understanding of a users' experience in this environment.*

### **4.1 Conceptual modelling**

#### **4.1.1 Introduction**

The measurement models that are proposed and analysed in this chapter are related to the emerging technologies for web services. It is clear that in this area the availability of acceptable objective performance metrics that correlate well with subjective scores are still in their early stages of development. Thus, more effort and standards are needed for defining a measurement model to represent the perceived quality which is experienced by end users commonly referred to as QoE.

Obviously, any requirement for a comprehensive QoE model should incorporate the merits of previous methods in order to have a suitable QoE model for packet traffic in general and for Web based services in particular. The QoE for a user of web-based services not only depends on network features but also on higher layer characteristics. Based on this observation, a method for objectively assessing QoE is needed.

Since QoE relates to a users' experience, it partly involves a form of psychological measurement (subjective); however, it is important to telecommunication service providers to express QoE parameters in relation to network performance measurements

and equipment, which are objective in nature. There have been a number of investigations into this kind of relationship that are reported in the literature, [24, 27, 51] relating objective networking service conditions and user perception. This is an obvious way to provide a measure of QoE for users, and a suitable approach for establishing the relationship between QoE and QoS. In addition, Pais, I. [102] confirmed that QoS, QoE, user actions, and end user satisfaction can be established as a closed relationship loop as shown in Figure 4-1. From this figure, it can be seen that network performance, end user action, end user experience and end user satisfaction are closely related in which user experience is linked to end user action via network performance, and users perception is contributed to further network performance. This loop conveys the idea of the recent concept of quality that should be specifically based on the end user. Pais, I. research conveys the idea that the end user experience is not only related to network performance, but it also reflects their expectations and actions at the time of accessing the system. This shows that an end user's experience is complicated, and needs more integrated information to understand, rather than a single parameter such as a network performance measurement or subjective user satisfaction scores. Basing QoE on only a single parameter could easily lead to a biased view of a user's quality of experience or be difficult to assess their scores in more complicated service performance scenarios.

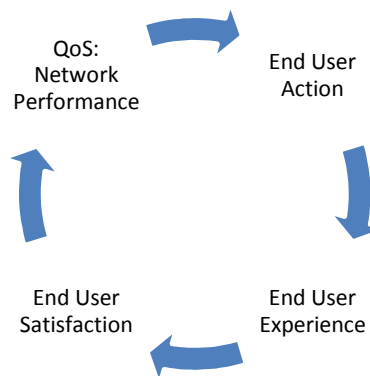


Figure 4-1: A closed relationship between QoS, QoE, end user action, and user satisfaction

Therefore, in a real integrated network environment, it is difficult to represent features of the various services using only the bandwidth and latency time; however, based on the higher layer characteristics of multi-media services, an analysis of service platform performance can be carried out. Any new model should have application-network-related performance criteria that correlate well with known or exact MOS scores. As mentioned above, a new model should incorporate objective effects and any human

factors that could affect outcomes of observations in order to avoid the pitfalls observed for an MOS that were based on scores and user surveys. As mentioned in the review discussed in Section 2.1, subjective assessments are prone to various errors because users' scores are subjective in nature. Therefore, parameters of any new model should be based mainly on objective measurements, but they should avoid the complexities associated with, for example, the E-model that is based on impairment values [21]. The new model also needs to take into account any human factors, to ensure that the model can be aligned with MOS or similar subjective measurements of service quality.

#### **4.1.2 Design of a suitable QoE model**

From the comprehensive review presented in Section 2.4, there are typically three main approaches for measuring QoE. They can be divided into the following categories: firstly, a subjective quality evaluation [12] which is based upon statistical methods, user scores and user surveys; secondly, an objective quality evaluation [21, 22, 46]; and third, correlation between QoS and QoE [49-51].

Our novel model combines these three main approaches as mentioned in Section 1.4 in order to achieve two main goals that are required in the existing literature for a new model:

- The QoE model should correlate well (better than 92.1%) with the Mean Opinion Score approach. The best correlation that has been achieved for QoE of web based services ( $R^2$ ) is 0.921 based on session time [27].
- Observations should be taken in an objective manner to infer the QoE being experienced by users [11, 51].

To build a model that satisfies our requirements, existing methodologies and models for QoE have been analysed. As previously mentioned, the above approaches all have their own drawbacks and merits as discussed in more detail as part of Chapter 2. Our novel approach that combines all three methods draws upon the merits of both of objective and subjective methods, and includes an understanding of the psychological aspects of users in order as described in Section 1.4, to produce some important enhancements. Firstly, using a QoS-QoE correlation approach to directly assess QoE based on current QoS information flows that have been recently captured and analysed; It can be said

that, inputs for the QoS-QoE correlation are real (objective) and based on current networking conditions and not predicted from only subjective inputs. Secondly, that QoS-QoE is accessed concurrently with the information reflected from the application layer in which the web service can be executed. Last but not least, human factors and content/time (*CT*) effects are taken into account in the model to provide a deeper understanding of a users' perception.

## **4.2 Evaluation of Metrics for QoE of Web based services**

To build up an analytical modelling framework, it is necessary to ensure that appropriate inputs can be obtained through measurement/observation to feed into the model framework, and then suitable factors need to be selected for the framework. Thus, separate factors have been tested to see their impact on the mean opinion scores. Our experiments have been devised in order to determine the viability of obtaining such inputs from observing users' behaviour based on objective measurements.

To achieve the aim of our QoE modelling approach, the purpose of these experiments is to consider issues such as:

- 1) The relationship between a user's MOS score and the observed metrics in order to define objective measures of QoE [52] .
- 2) The trends in a user's perception reported or transmitted back to service providers for further QoE enhancement [103].

As our approach combined the three different methods mentioned in Section 1.4, the different metrics have been captured from each of these three-different methods. These metrics obtained from subjective tests of state of mind, objective tests of delay and requests per second and psychological tests of content had been used to assess their effect on web browsing user experiences.

### **4.2.1 Networking metric for QoE of Web based services**

#### **4.2.1.1 Pilot Study - Evaluation of delay metric from users' scores**

In this experiment, network delay is tested as a possible metric to see if users can perceive distorted performance via a change in network delay. A randomly selected group of 12 users participated in the experiment and gave their opinion scores without

any knowledge of networking performance. The purpose of this experiment was to see whether delay can affect a users' perception of quality.

The numerical results from this experiment clearly show that users can recognize changes in networking performance via their scores as illustrated in Figure 4-2. It depicts the percentage of user scores for a range of different delay values. The values of delay had been chosen regarding to the relevant low, medium and high level as recommended in end-user multimedia QoS categories [52, 54] It can be observed that, when the network is operating appropriately (delay is less than 50 milliseconds as controlled in the local networking environment), it can be seen that 75% of users feel "Very Satisfied", "Satisfied" and "Fair" in their evaluation of the system. Correspondingly, the percentage of users who voted for the "Dissatisfied" and "Not Recommended" categories increased when the network performance became worse. It should be noted that, in the poorest networking situation depicted in Figure 4-2, there is a 35% higher proportion of users who voted for "Dissatisfied" and "Not Recommended" than compared to the first case.

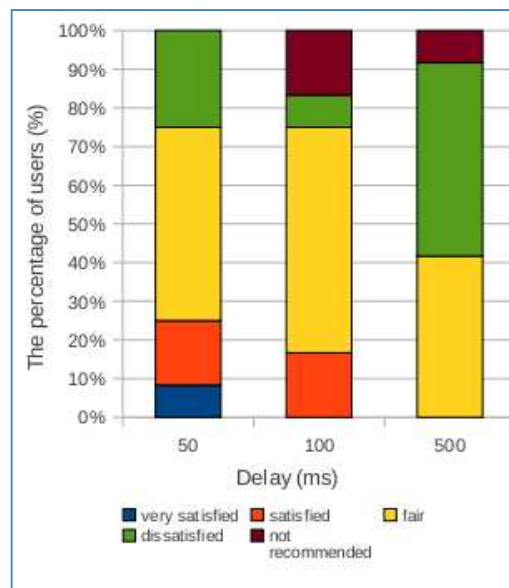


Figure 4-2: User's Experience for different Delay Scenario

In general, it is observed that the users' scores in the experiment follow our intuition or obey a "rule of nature". For example, the users' scores tend to be "Dissatisfied" when we increase the delay. However, some abnormalities are observed within the specific trends of "Fair" and "Not Recommended" levels, showing that users cannot accurately

distinguish between performance delay levels as they deteriorate beyond normally acceptable limits.

#### ***4.2.1.2 Pilot Study - Evaluation of delay and packet loss metrics from user's scores***

In this experiment, we have introduced the additional networking parameter of packet loss. Controlled values were chosen for a limited number of combined runs only. This is a pilot study, the values of delay and packet loss were chosen as suggested in [52]. The combinations of those values were selected randomly. The purpose of this experiment was to see the impact of two parameters from networking derived metrics on the users' perceptions.

Loss and delay combinations that were chosen are {packet loss (%), delay (ms)} of {5%, 0ms}, {5%, 50ms} and {10%, 100ms} respectively.

Figure 4-3 shows users' scores for both increasing values of delay and packet loss simultaneously. In the case of both increasing values of delay and packet loss, it can be seen that the users' scores directly reflect the controlled network parameter changes. An increase in the proportion of "Dissatisfied" and "Not Recommended" categories is noticed when both delay and packet loss values are increased. On the other hand, under the same network conditions, the figure also shows the corresponding reduced proportions for the "Very Satisfied", "Satisfied", and "Fair" categories.

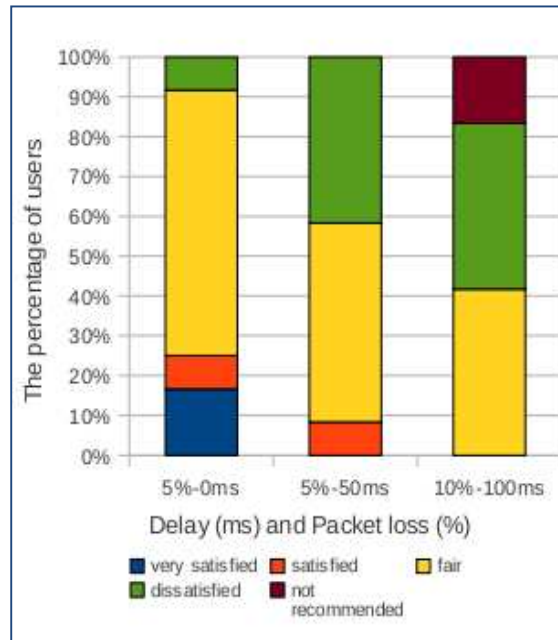


Figure 4-3: User's Experience for different Delay and Packet Loss Scenarios

From this figure, it can be concluded that there is an agreement between the users' perception and increased values for both delay and packet loss which we identified as "the rule of nature" mentioned in Figure 4-4.

Our concept of the "Rule of Nature" is that users' scores in the experiment follow our intuition. The intuition is that the users' scores directly reflect the controlled network parameter changes. For example, the users' scores tend to be "Dissatisfied" when we increase both values of delay and packet loss simultaneously. On the other hand, an increase in the proportion of "Dissatisfied" and "Not Recommended" categories is noticed when both delay and packet loss values are increased.

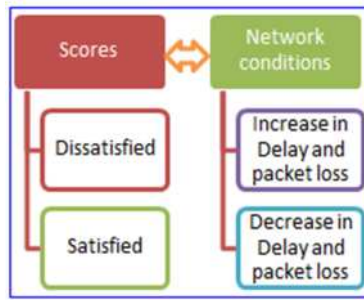


Figure 4-4: Concept of “Rule of Nature”

However, the limited combinations of both delay and packet loss in this experiment may lead users to guess or recognize a controlled change of parameters, and this should not be observed in the more complicated networking environmental conditions.

## 4.2.2 Application metric for QoE of Web based services

### 4.2.2.1 Pilot Study - Evaluation of requests per second metric from user's scores

It is known that web services are located at the application layer of the 7 layer open systems interconnection (OSI) Reference model [104]. Quality of web service also depends how the web server can deal with its workload of the numerous numbers of user requests, and returned responded http requests [105]. The quality depends upon web server performance [106].

Our research does not aim at improving a web server's performance. However, we study whether application performance (actually web server performance) has any impact on users' scores in the case of high and low numbers of requests from the users. To achieve this goal, firstly this experiment is aimed at seeing how the requests per second from users are changed based on the current controlled environment. Once there is a relationship between users' perception via their requests and controlled network, web server performance have to meet that demand from users' requests in terms of quality [107]. Therefore, it is deduced that both network and application layers (on the web server side) may contribute to QoE.

In this experiment, the percentage of request packets with respect to the total number of packets was measured within a controlled network environment. The request packets are communicated between a user and the web server. The server request packets emanating from the web server have been filtered out in this measurement. We used

Wireshark [108] to capture all http traffic in this experiment, then filtered out the http requests that were sourced from only hosts in order to obtain the numbers of user requests.

The purpose of this measurement is to see a users' perception of the effects that are due to application performance.

Figure 4-5 shows the percentage of request packets from users in response to distorting the network's performance. When we set the delay to less than or equal to 50 milliseconds, the percentage of user request packets is more than 61%, while, on the other hand, it is just around 56% when the delay is increased to 500 milliseconds.

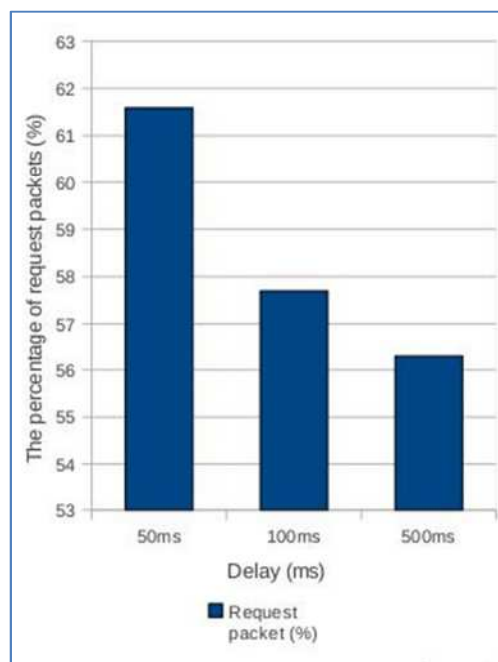


Figure 4-5: Request packet in the whole packet for different delay

Figure 4-5 shows the percentage of user request packets within the network during the delay-controlled networking environment. It can be seen that the percentage of request packets is decreasing as we increase the delay across the network.

It can be concluded from these observations that service platform performance contributes to a user's QoE perception. Thus, both networking and application performance needs to be taken into account in our modelling.

### **4.2.3 Evaluation of metrics derived from both networking and application layers to users' scores**

The first experiment mentioned in Section 4.2.1.1 was to test the usefulness of network delay as a metric; the results obtained show that subjects were indeed sensitive to delay though not in a consistent manner at delays for which the performance was unacceptably bad. The third experiment mentioned in Section 4.2.2.1 was to test the usefulness of the number of requests per second as a metric; the results obtained show that there is a relationship between users' scores and requests per second. From these results, it is deduced that both delay and requests per second may contribute to QoE. Therefore, a fourth experiment has been developed to include both metrics from the networking and application layers. For example, both delay and requests per second are controlled concurrently.

To avoid any unbiased controlling environment to users, the experiment was carried out under Taguchi approach in which the second controlling session mentioned in Chapter 3 has been used. As noted in Figure 4-3, if combinations of controlled parameters are not random, users may easily perceive the distorted performance quality. Therefore, the Taguchi approach is used to ensure randomly combined runs of the controlled parameters. This also ensures that users cannot guess or predict which controls have been applied and must not know about changes of the control parameters. The only thing they can do is to assess their perceptions or expectation of a service via their experience from previous networking and platform performances. Using the Taguchi approach, our experimental design enables us to combine the relevant but a reasonable number of runs with controlled parameters to cover all necessary combinations of factor, but ensuring that they cannot be predicted or guessed. Thus, the changes of controlled values are seen to be random.

#### ***4.2.3.1 Experiment under the second controlling session using the Taguchi Approach***

The aim of this experiment is to recognize the impact of these various metrics and their two-way interactions both from a networking and an application performance perspective in assessing QoE.

The factors and their levels of this experiment were followed using the Taguchi guidelines which were described in detail in Chapter 3. They are summarized in the table

below, using four controlling factors and a noise factor, they are represented as A for delay, B for packet loss, C for requests per second, D for response time, and E for Content with regard to their levels and values as noted in the table.

Table 4-1: Factors and Levels

Factor		Level		
		1	2	3
Control	A:Delay(ms)	50	100	500
	B:Packet Loss (%)	5	10	15
	C:Requests Per Second	1	500	15500
	D:Response time	Improved	Not Improved	—
Noise	E:Content	Picture	Pictures/Text	Pictures/Text/Videos

Statistical tests have been applied for all of the factors that we derived from networking and application layers of the experiment under the second controlling session using the Taguchi approach. The purpose of test was to see whether these factors have any significant impact in terms of statistical tests in this experiment.

The analysis of variance (ANOVA) in Table 4-2 that has been published in [109] shows that all of the four control factors and their interactions viz: delay, packet loss, requests per second and response time improvement have effects on the users' scores. However, delay and requests per second have the most significant affect on the response, since their p-value is less than 0.05 and this is considered to be significant with respect to this statistical test.

Table 4-2 Analysis of Variance

ANOVA;R-sqr=.658;Adj:.624 MS Residual=.477					
	SS	df	MS	F	p
(1)DELAY L+Q	47.746	2.000	23.873	50.046	0.000
(2)PACKETLO L+Q	1.164	2.000	0.582	1.220	0.298
(3)REQPERSE L+Q	31.276	2.000	15.638	32.783	0.000
(4)RESIMP L	0.247	1.000	0.247	0.518	0.473
1*2	47.228	4.000	11.807	24.751	0.000
1*3	29.221	4.000	7.305	15.314	0.000
1*4	20.864	2.000	10.432	21.869	0.000
Error	81.571	171.000	0.477		
Total SS	238.582	188.000			

The linear regression model can predict 65.8% of MOS. This correlation of 65.8% ( $R^2 = 0.658$ ) might be regarded as strong where the measures are based on a five-point Likert scale.

#### 4.2.3.2 Experiment under the third controlling session by Taguchi Approach

The purpose of this experiment was the same as the previous one discussed in Section 4.2.3.1, however, the significant controlling factors that identified from the previous experiment such as delay and requests per second were used. However, these factors were increased in their level for further testing. Therefore, the third controlling session mentioned in Chapter 3 was used.

As for the previous experiment in Section 4.3.2.1, the analysis of variance was performed to see whether delay and requests per second have had a significant impact on the MOS response and this is shown in Table 4-3. The p values show that the factors of delay and requests per second are significantly contributing to the outcome.

Table 4-3: Analysis of Variance

Response: MOS	Df	Sum of squares (SS)	Mean of Squares (MS)	F value	Pr(>F)
(D)	4	172.96	43.241	59.947	< 2.2e-16
(RPS)	4	214.28	53.569	70.549	< 2.2e-16

Figure 4-6 shows the screening data collected related to delay and requests per second from this experiment using the experimental design based on the Taguchi approach. It shows a plot of the different values of delay and requests per second with their associated mean opinion scores which users recorded, based on the settings for delay and requests per second. The x-axis represents the delay; the y-axis represents requests per second. The density of *MOS* distribution is shown using different colours and shades; where scores of “5-Very Satisfied” are in blue, and scores of “1-Not recommended” are in purple.

The density of *MOS* using delay and requests per second indicates that the *MOS* of users depends on controllably fixed factors of delay and requests per second  $\{D, RPS\}$ . For example, when both delay and request per second  $\{D, RPS\}$  are set at  $\{500, 5\}$ ,

the highest value of delay, and lowest level of number requests per second, the density of *MOS* shown in the purple colour, means that it is at *MOS* equal to 1 or {Not Recommended}.

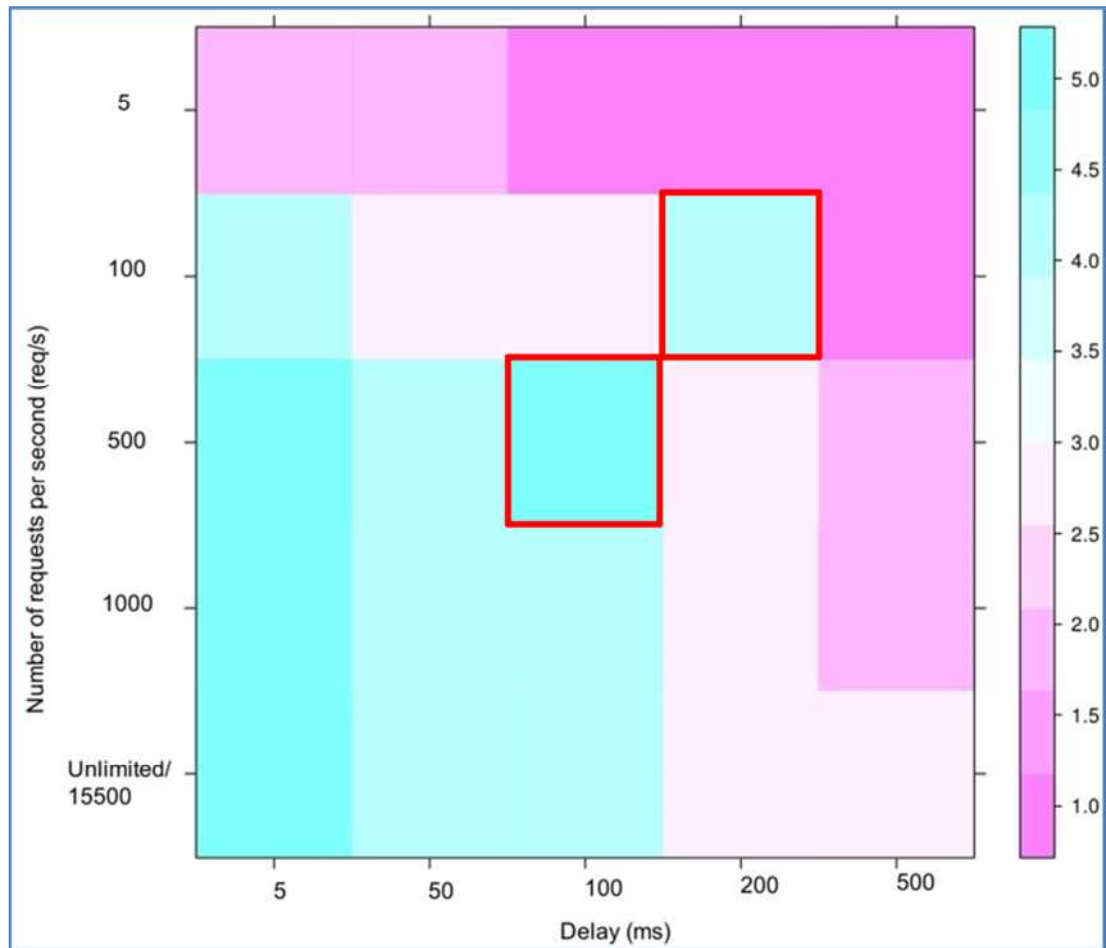


Figure 4-6: Mean Opinion Score vs. Delay and Requests per second

From the screening data, there is a general agreement between users' perception and increased values for both objective metrics of network and application performance and we identified as "the rule of nature" or our intuition in a more complex test as described in Figure 4-7.

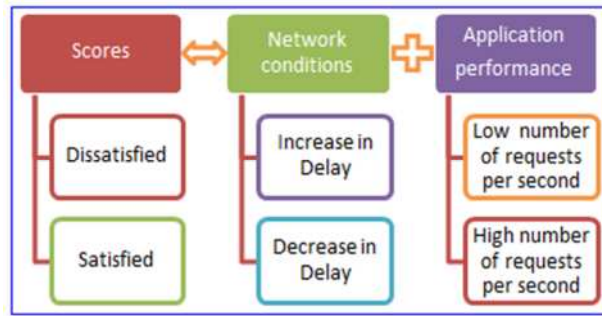


Figure 4-7: Concept of “Rule of Nature” in a complex test

However, since both delay and requests per second are taken in account, there are two remaining abnormalities which are shown highlighted by red-line rectangles of Figure 4-6 in detail as:

- 1) The light blue block between 100 requests per second and 200 milliseconds in delay which should be in an extremely light purple referred to “Fair” of users’ perception instead of a light blue block referred to “Satisfied or 4”.
- 2) The dark blue block between 500 requests per second and 100 milliseconds in delay which should be in a light blue referred to “Satisfied or 4” of users’ perception instead of a dark blue block referred to “Very satisfied or 5”

As in our tuition, “satisfied” or “very satisfied” should be in the lowest level of delay and highest level of requests. These above results showed that there are some abnormalities in the scores of “Satisfied or 4” and “Very satisfied or 5” and that “the rule of nature” (in our terminology) or our tuition is not followed in this situation.

The statistical test of all levels for the factors of delay and requests per second (in terms of a simple linear regression model) reported in Table 4-4. The results from this table show that the delay at lower levels is less significant than for higher levels, therefore, further investigation is needed to answer whether there are other impacts besides those identified from the application and networking metrics.

Table 4-4: Statistic test for each level of factors

Factors: Delay (D) and requests/second (RPS)	Estimate	Std. Error	t value	Pr(>  t )
(Intercept)	1.9311	0.1232	15.670	< 2e-16 ***
(D)50	-0.2889	0.1299	-2.224	0.0267 *
(D)100	-0.1444	0.1299	-3.313	0.2668
(D)200	-0.6667	0.1299	-5.383	< 2e-16 ***
(D)500	-1.7222	0.1299	-16.286	< 2e-16 ***
(RPS)100	1.3333	0.1299	18.369	< 2e-16 ***
(RPS)500	1.7778	0.1299	13.859	< 2e-16 ***
(RPS)1000	1.7667	0.1299	13.157	< 2e-16 ***
(RPS)15500	1.8000	0.1299	23.124	< 2e-16 ***

Note: **p-value ≤ 0.05 is significant**

This suggests that we should develop further experiments to test further factors such as State-of-Mind (*SOM*) and *Content* that may or may not have affected users' scores in the current experiment.

#### 4.2.4 Subjective metric-Boredom and state of mind for Web QoE assessment

##### 4.2.4.1 Boredom

During the course of our experiments, psychological effects such as boredom [64] and different states of mind were noted.

As mentioned in Chapter 3, the Taguchi approach has been applied to the design of our experiments to ensure that we can obtain users' scores for a quantitative limited number of runs and users. The method ensures high quality of experiment can be produced for a low cost and minimum time.

However, in our experiment, the subjects of our experiments are users, and we experienced that by using the Taguchi approach that involved human subjects has led to the possibility of boredom effects for the users that could influence their MOS score. Thus, we need to develop approaches to the study that eliminate or reduce these effects as they may lead to bias in the results from our model.

Boredom is defined in terms of a psychological process as an unpleasant or lack of interest in concentrating on a current activity [110]. In our pilot studies, the boredom effect may have been occurred because users were concentrating less on surfing web pages when they had to repeat that action numerous times during the course of a long experiment. A randomly experimental repetition of combined parameters is necessary to ensure random inputs in the experiment in general, and for the Taguchi method in particular.

Our aim is to avoid boredom that may affect users' scores. However, the motivation for this avoidance is to reduce any boredom effect that may be caused by our course of experiments. Natural boredom generated by the users themselves is out of our scope as we cannot influence this effect in any way.

Therefore, to reduce this boredom effect, we introduced a session known as a controlling session. A controlling session is divided into several segments. The segments in our experimental design are set out to reduce the impacts of the boredom effect by limiting the time in carrying out our approach. Since we acknowledge the boredom impact on users' experiences from our experiments, the design has been divided into several segments, which try to eliminate (or minimise) this effect. By shortening the experiment period, the level of our users' boredom is reduced. This resulted from our pilot study as it shows that if time is limited to around 120 minutes, the users may not feel bored with these activities. Therefore, a time-frame limitation is tightly managed in our experiments.

#### ***4.2.4.2 State of mind***

State of Mind is a mental state or mood state. It is a mental phenomenon in which the qualities of a state are relatively constant even though the state itself may be dynamic. According to [111] mind is "all mental phenomena to these imaginative, or as some would say today, conscious and occasionally unconscious experiences". The users' states of mind can change over time and may depend on many specific circumstances that are almost impossible to quantify. However, such degree of boredom as normal, boredom and stressed can be symbolized states of mind in general [112]. Therefore, in our scope of the research, users' states of mind are recorded subjectively with a 3-level scale based on their current feelings as {"Normal", "Bored", "Stress"} which aims to see how a

user's state of mind has changed during the course of the experiment. The current feeling levels based on the users' feedback via their comments during our pilot study.

Some studies have been described in the literature [66, 113-115] where the state of mind of subjects can be captured and processed using techniques such as image processing, audio-visual processing, or profile based mood extraction and so on. Using image analysis, image features such as low level of contrast, brightness or edge contribute to mood [113]. Using video analysis is to detect abnormal user behaviour [71]. Using a haptic system supported sensing devices is to detect mood [114]. Using a computer vision system to characterize facial motion and muscle activation for mood detection [115].

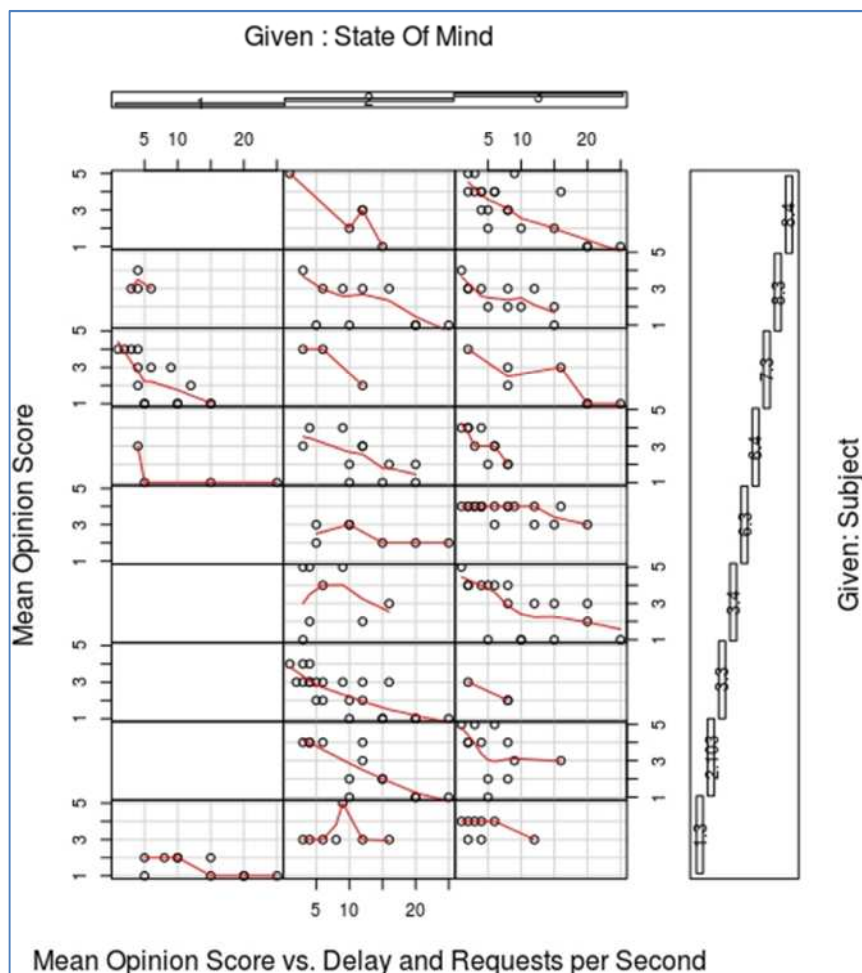


Figure 4-8: MOS vs. Delay and Requests per second

Figure 4-8 gives users' scores for different users representing different delay and requests per second performance with three different conditions relating to the users'

states of mind during the experiment. As indicated in the Figure 4-8, each subject has provided a different assessment under the same conditions of networking and application performance. Their variable assessment can be attributed to their states of mind at the time and the fact that it is very individual assessment. Their scores are seen to be very different depending upon their state of mind as classified into “3-Normal”, “2-Bored”, or “1-Stressed”.

According to Loftus, E.F in [116], “*misinformation effect refers to the impairment in memory for the past that arises after exposure to misleading information, and various inhibited states of mind can increase misinformation effects*”. This hypothesis is true with our participants included in our situation. For example, when users are in a “Stressed” mood as shown in Figure 4-8, their scores do not indicate any trends fitting with our assumption of networking and application performance.

It has become clear that SOM is not easy to capture in a realistic situation although it is needed to support complex analysis from both the psychological and machine learning perspectives. It is a fact that there is really no fool-proof way of assessing these SOM parameters, even though it is clear that it has an impact on users’ scores. As shown in Figure 4-8, SOM has changed over time for each individual. This leads to the observation that shows changing MOS scores even though the controllable effects of delay and requests per second remain fixed.

#### **4.2.5 Content metric for QoE of Web based services**

In this experiment, *Content* has been tested on a group of twelve users. The *Content* has been divided into four categories of {type1, type2, type 3, and type 4} as mentioned in Chapter 3.

In our scope of the research, the complexity of the content is objectively measured by increasing the download time of the Web page. This is because, more resources, and functionalities have been integrated into the higher level of content type to reduce boredom in terms of content itself of the web site.

According to Callan, .D in “Content is King” [93], it is concluded that the more useful and interesting the content is then the more successful a website will be in practice. However, based on Odlyzko, .A in [117], “Content is not King”, he argues that “connectivity is more important than content”.

Therefore, the aim of this experiment is to see whether *Content* influences users' scores under different controlled network and application performance conditions.

Under our assumptions, an increase in content complexity on a website leads to an increase in the downloading time for that website. Therefore, this makes another assumption that, if users are not affected by any content, then they can differentiate between an increase/decrease of downloading time for different types of content.

Table 4-5: MOS vs. Content

MOS	Content			
	Type1	Type 2	Type 3	Type 4
1	4	10	19	21
2	25	26	29	32
3	51	48	36	26
4	18	17	15	23
5	10	7	9	6

Table 4-5 list a number of different MOS scores from web browsing users for different types of content under the same setting of networking delay and available requests per second responded from the web server for each type of content. As shown in the table, it suggests that, the number of users' scores of 1, and 2 is increasing and that for scores of 3 it is decreasing when increasing the complexity of the content. Those for scores of 4, and 5 are indicated no trend when increasing the complexity of the content.

For example, if it is assumed that downloading of web pages is performed using an ADSL 512 K connection infrastructure, then the estimated download time is assigned for different types of content {type1, type2, type3, type4} as shown by the red lines of Figure 4-9 and Figure 4-10. Other lines are MOS at {1,2,3,4,5} for different types of content.

As depicted in Figure 4-9, the line-trends of MOS at {1,2,3} fit with our assumption. When the download time is increased, more complicated content respectively, the number of users who chose either {1,2} illustrated by the brown and green lines is increased, while the number of "fair" users (yellow line) decreases. Thus, we see no effect of *Content* on users at MOS scores of {1,2,3}.

However, the above table suggests that the number of users' scores at levels 4, and 5 are not matching with our assumption as mentioned above. As depicted in Figure 4-10, the

estimated download time for different types of content (shown by the red line) does not match with the line trends of MOS at {4 or 5} illustrated by the blue and light blue lines.

The expected trend should be the same as for the yellow line trend of MOS of 3 (“Fair”) as indicated in Figure 4-9. Thus, it is concluded *Content* has its own effect on users’ scores at levels of {4,5}.

In conclusion, the results obtained show that Content has an impact on users’ scores in several different aspects. Under good conditions in the network, users were interested in Content rather than connectivity, and then their scores were biased more towards Content more than connectivity quality. With a bad network environment, Content did not play any important role as users paid more attention to their connectivity, their scores were biased towards the network quality rather than “real content”

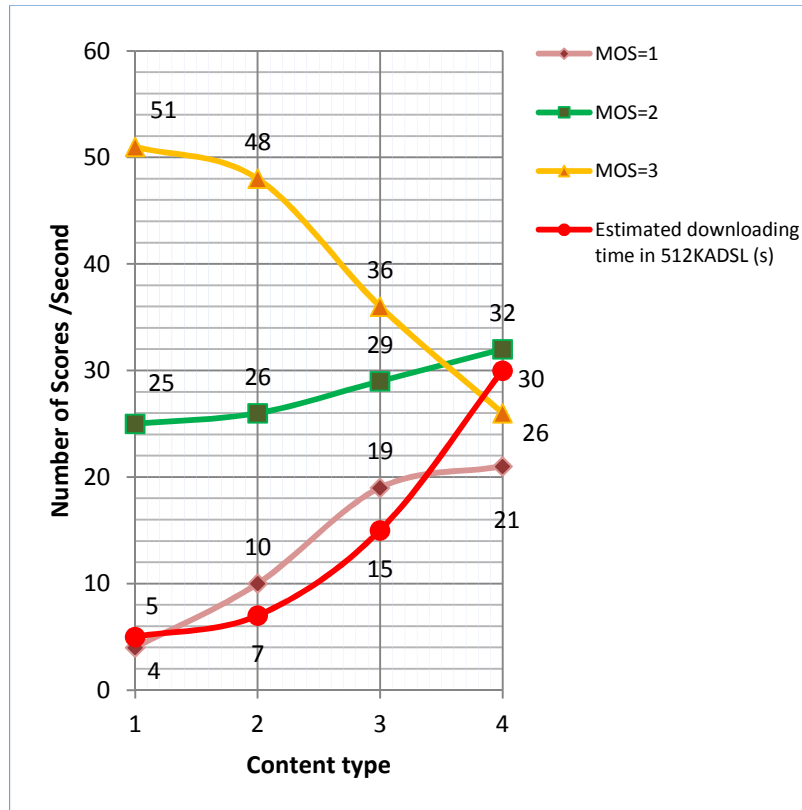


Figure 4-9: MOS of {1, 2, 3} matching type of Content

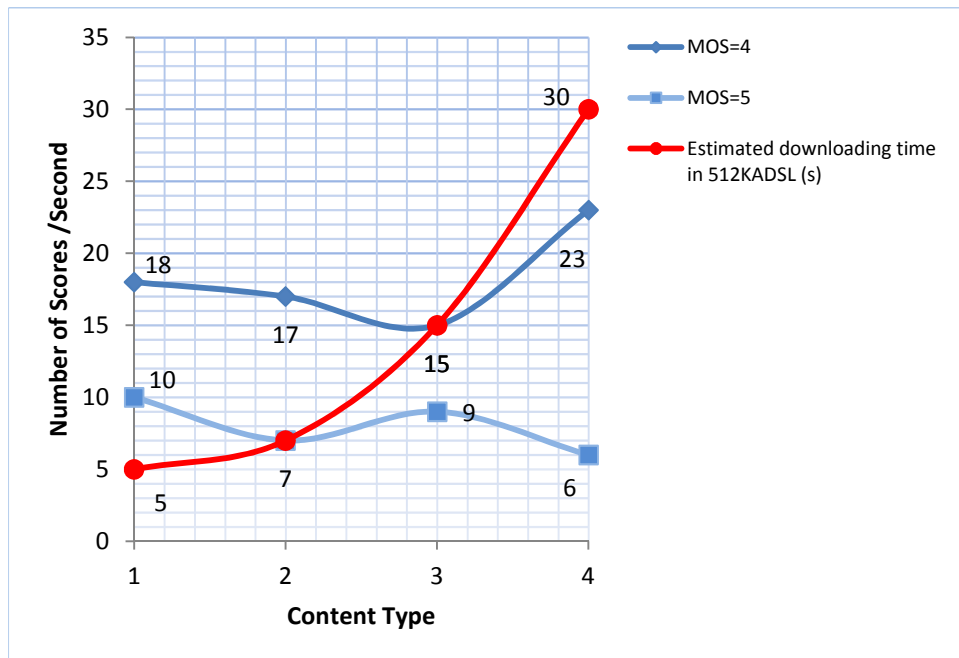


Figure 4-10: MOS of {4, 5} un-matching type of Content

#### 4.2.6 Summary

Generally, we see that the users' scores obey the rule of nature or our intuition for the case of a distorted networking environment. However, there are some abnormalities when the users' scores are at the levels of "Very satisfied-5" and "Satisfied-4" for the networking and application performance with the levels of "Fair" and "Not Recommended" in Section 4.2.3.

Moreover, users' expectations about the system response are higher than that of the nearly perfect networking conditions. For example, it is shown in Figure 4-2 that less than 10% of users are "Very Satisfied" where the Delay is less than or equal to 50 milliseconds. This means that some users' expectations may be higher than the "perfect" performance that can be set in the test bed.

To get a more concise view of a users' perception, the experiments in Section 4.2.2 showed a relationship between the percentage of request packets and the networking controlled performance. Therefore, higher layer application performance should be used for the assessment. The experiments show significant effects of metrics and their interactions both from networking and application performance perspectives. Thus, these metrics and their interaction should not be ignored in any assessment of QoE for the case of Web based services.

Since the abnormalities described in Figure 4-6, it is clear that we need to understand why these abnormalities have occurred for the cases of "Very satisfied-5" and "Satisfied-4" in the users' scores. Thus, to get a better and clear understanding of *MOS*, we carried out further experiments which related to psychological effects such as state of mind (*SOM*) and content (*Content*).

We found that the users' state of mind changed randomly during the performance of the experiment; however, this misinformation effect happened whenever users were in a "stressed" mood.

Last but not least, we investigated that the proposition that *Content* influences users' scores under different controlled networking and application performance conditions. Although *Content* is not the most important factor to be taken into account in the model; however it is necessary to include it in order to have a full assessment of QoE

Web based services and to clearly understand customer QoE. We found that *MOS* scores at levels of {4,5} were not matching with our assumptions related to download time. It is concluded that *Content* influences users' scores at the levels of {4,5}. This also can be explained by abnormalities that occurred in our previous evaluation of the networking and application performance mentioned in Section 4.2.3 where two blocks of user's scores at {4,5} did not follow the setting of networking and application performance. It is concluded that *Content* has affected users' scores in this case.

## **4.3 Analytical modelling framework**

### **4.3.1 Proposed framework**

In the framework, QoE for Web based services is assessed via their four grouped affected factors as described in Figure 1-4 that were extracted from three different methods that we mentioned in the discussion for Chapter 2. Each effect has been evaluated for its effect via experiments described in Section 4.2. From the conceptual modelling presented in Section 4.1, it is clear that users can perceive a change of networking performance. Therefore it can be established that there is a relationship between QoS and users' mean opinion scores. Thus, a networking-derived metric must be the first input factor into our analytical framework.

All of the networking and application based factors need to be integrated with human state of mind (*SOM*), content and time (*CT*) factors in order to create an appropriate framework for a QoE model involving web traffic. The *CT* derived metric from our perspective is a metric related to the specific content present at the time that the user surfs a website. Web based services in our project are involved viewing web content on a browser where the content and delivery of that content have been subjected to parameter modifications that are expected to influence user perception of the web services.

#### **4.3.1.1 Application characteristics-Derived metric**

An application derived metric reflects performance from the perspective of the application layer. The application metric consists of requests per second (*RPS*) and/or response times (*RT*).

The influence of application-derived metric on QoE estimation is represented by the function:  $App = f(RPS, RT)$  (2.4)

#### 4.3.1.2 Network -Derived metric

Network-derived metric is responsible for network feature estimation. The parameters from this metric were delay ( $D$ ), and packet loss ( $PL$ ) which have been controlled in our experiment.

The influence of network-derived metrics on estimation is represented by the function:  $Net = g(D, PL)$  (2.5)

#### 4.3.1.3 Subjective human-Derived metric

During the course of our experiments, it has been undertaken to determine the link between objectively observed and controlled parameters and subjective reporting of QoE by the human subjects. It is known that no two brains are the same, so it is expected that individual scores from human subjects will not be identical in such a study.

Users themselves are all subjective in nature. They represent a random factor whose levels in the experiment are none-exhaustively sampled from a large population. Their states of mind are changing via time and may depend on many specific circumstances. Users are treated as random variables, thus the main effects of objectively observed and controlled parameters represent human users in general, not a local group of users.

The subjective random factors are user identity ( $id$ ), time ( $time$ ), states of mind ( $SOM$ ):

$$b = \{id, time, SOM\} \quad (2.6)$$

#### 4.3.1.4 CT-Derived Metric

The  $CT$  derived metric is in form of  $c = \{Content, time\}$  (2.7)

$CT$  -derived metric takes into account the factors related to content and time. In a real situation, users can browse any wanted of embedded content at an allowed time on the provided web pages.

### 4.3.2 Summary

In this chapter, firstly, we identified and discussed the state of the art for QoE modelling. New metrics were proposed for the model of QoE that uses a comprehensive set of metrics to evaluate an overall customer QoE.

The appropriated metrics have been obtained and evaluated to feed into the model framework through experiments such as a pilot study to reduce boredom, four different experiments to networking and application controlled performance, and two different experiments to *SOM* and *Content* .

An analytical QoE Web modelling framework has been described. Based on our experiments, metrics as inputs for the model have been proposed in order to assess and understand the links between them and users' mean opinion scores. In particular, it presents a comprehensive framework for measuring and analysing QoE in Web surfing scenarios.

## Chapter 5 Results and Evaluation of QoE assessment

*This chapter investigates the applicability of a mixed effects model for predicting QoE in World Wide Web based multi-media services. An analysis of objective factors and human factors that may impact on outcomes of our QoE observations is presented with an emphasis on fixed and random effects. The fixed effects take into account population mean and they are interested in explaining the response itself. The random effects take into account group specific effects, and account for a correlation structure of variations amongst users, and they are interested in explaining the variance of the response. The objective metric for QoE evaluation has been based on networking perspectives from the World Wide Web and metrics based on the application layer and networking layer. The method that has been developed to assess QoE of Web based services takes into account the variance of individual subjects amongst a group and their response time. The model shows an acceptable correlation between values of the fitted analytical model and observed user scores. It is demonstrated that the integration of further human perception factors into the assessment, which represent subjective metrics related to a state of mind and content factors. The model is further developed to account for these factors and other potential covariates to QoE assessment during the course of the experiments.*

### 5.1 A Mixed Effects Model for Predicting Web-based Services QoE

#### 5.1.1 Introduction

The mixed effects model was first proposed in 1980's by Laird and Ware [118], and its improvement and modifications were published in [119]. This model has been applied in many psychological research areas, such as linguistic experiments [120]. However, according to our knowledge, it has not been applied in the QoE area where such psychological or human and subjective factors are being considered for the delivery of services with a certain QoE [6].

The mixed effect model was designed for group/clustered data where there are correlations between subunits within subjects. For QoE of Web based services, observations are repeated on each user and on a group of users, and furthermore, observations on the same users are not considered independent.

Thus the mixed effect model has been chosen to model our data for the following reasons:

- (1) The model incorporates outcomes of an observation as being fixed (population) effects, and non-observable grouped specific random effects accounting for the correlation structure of variation amongst groups.
- (2) We see that individual subjects and items may have intercepts and slopes that diverge considerably from the population means, no two brains are the same; therefore, with different individuals and items, grouped data should be used in which correlation between subunits within subjects are being analysed in the model.
- (3) Mixed effect model is robust against missing data, provided that data are missing at random. Hence, we do not need to impute missing data with debatable imputation methods.
- (4) In addition, a mixed effects model does repeated measure regression. Subject variance is taken into account using the concept of a random effect. This method can eliminate subject-variant variance in each user to the degree in which items produce an effect that is useful, if having several different stimuli (e.g.: time or individuals) which superficially look different but have an underlying form and agree to our assessment hypothesis.

Thus the mixed effect model decomposes the outcomes of an observation into fixed effects (population mean), and random effects (group specific) and accounts for the correlation structure of variations amongst the groups.

The linear mixed effects model estimates the population average rate of change and the subjective specific regression lines. Each subject is permitted to have their own rate and base line level. It can be said that a linear mixed effects model is an extension of the general linear model that can specify additional random effect terms [121].

The linear model has been defined as:

$$y = X\beta + Zb + \varepsilon \quad (2.8)$$

The assumptions underpinning this model include:

$$\varepsilon \sim N(0, \sigma^2 I)$$

$$b \sim N(0, \sigma^2 S)$$

Where:

$\varepsilon$  is a vector of normally distributed effect residuals,  $\varepsilon_1, \dots, \varepsilon_N$

$b$  is a  $q$  dimensional random effects vector,  $b_1, \dots, b_q$

$b_1, \dots, b_q$ , and  $\varepsilon_1, \dots, \varepsilon_N$  are independent.

$y$  is the model response

$\beta$  is the vector of  $p$  dimensional fixed-effect coefficients and is the vector of regression coefficients

$y \perp \beta$  shows independence of random variables;

$N$  is the multivariate normal (Gaussian) distribution;

$X$  is an  $n \times p$  dimensional fixed-effects matrix where the structure of data and values of any covariate are used ;

$Z$  is a  $n \times q$  dimensional random-effects design model matrix;

$0$  is  $n \times 1$  vector of zeroes;

$I$  is the identity matrix of size  $n$

$S$  is the relative variance-covariance matrix for random effects with dimension  $q \times q$  ;

$\sigma$  is a common scale parameter.

The assumption that  $\varepsilon$  and  $b$  are normally distributed with values of mean (E) and variance (V) as defined below:

$$E \begin{bmatrix} b \\ \varepsilon \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$Var \begin{bmatrix} b \\ \varepsilon \end{bmatrix} = \begin{bmatrix} \sigma^2 S & 0 \\ 0 & \sigma^2 I \end{bmatrix}$$

The non-linear mixed effect model extends the linearity assumption to allow for more flexible function forms [122]. It can estimate more physically interpretable parameters.

The general nonlinear mixed effects model can be defined as below

$$y_{ij} = f(\varphi_i, x_{ij}) + e_{ij} \quad (2.9)$$

Where :

$y_{ij}$  is the  $j^{\text{th}}$  response of the  $i^{\text{th}}$  individual

$x_{ij}$  is a predictor vector  $j^{\text{th}}$  response on the  $i^{\text{th}}$  individual

$f$  represents the nonlinear function for the predictor vector and parameter vector  $\varphi_i$

$e_{ij}$  is a normally distributed noise term

At the second level,  $\varphi_i$  is modelled as:

$$\varphi_i = A_i\beta + B_ib_i \quad (2.10)$$

Where:  $b_i \sim N(0, \sigma^2 D)$  and  $\beta$  is a p-vector of fixed population parameters,  $b_i$  is a q-vector of random effects associated with the  $i^{\text{th}}$  individual.  $A_i$  is an  $n \times p$  design matrix for fixed effects.  $B_i$  is  $n \times q$  design matrix for random effects.  $\sigma^2 D$  is the covariance matrix.

### 5.1.2 Model parameters estimation

The underlying assumptions of the model represented by Equation(2.8) imply:

$$\varepsilon \sim N(0, \sigma^2 I) \quad (2.11)$$

$$b \sim N(0, \sigma^2 S) \quad (2.12)$$

$\{b_i\}_{i=1}^q$  is assumed to be independently identically distributed with mean zero, and covariance  $S$ .

All elements  $\varepsilon = \{\varepsilon_1 \dots \varepsilon_N\}$  are independent variables with mean zero, and variance  $\sigma^2$

It is assumed that  $\varepsilon$  and  $b$  are normally distributed with:

$$E \begin{bmatrix} b \\ \varepsilon \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \quad (2.13)$$

$$Var \begin{bmatrix} b \\ \varepsilon \end{bmatrix} = \begin{bmatrix} \sigma^2 S & 0 \\ 0 & \sigma^2 I \end{bmatrix} \quad (2.14)$$

Assume now that the expression for the covariance matrix of  $Y$  is normally distributed with mean  $\mu = X \widehat{\beta}$  and variance  $V$ . Then, the covariance matrix of  $Y$  is a conditional distribution and is multivariate Gaussian:

$$Y = N(X \widehat{\beta}, V) = N(\mu, V) \quad (2.15)$$

$$V = \text{var}(\varepsilon) + Z' \text{var}(b) Z = \sigma^2 I + Z' \sigma^2 S Z \quad (2.16)$$

Where  $V$  is a matrix whose elements are variance estimates of  $Y$ .

As is known, estimation of  $S$  and  $\sigma^2$  are the most important for the model and for improvement of  $\beta$ .

In order to estimate  $S$  and  $\sigma^2$ , the Restricted Maximum Likelihood [123, 124] has been used for the estimates of these variance components.

Basically, REML estimates  $\sigma^2$  in a fixed-effects regression using the residual sum of squares (SSR)/degree of freedom  $(n-p)$  where  $p$  is the number of coefficients. Thus, with a standard linear regression,  $\widehat{\sigma}^2 = \sum (y_i - \bar{y})^2 / (n-p)$ .

In the mixed effects model, REML procedures maximize the part of the likelihood function where the fixed effects parameters are invariant. REML takes the likelihood and partitions it into two components. The first is a likelihood component involving all the fixed effects ( $\mu$ ). The latter component is a residual likelihood ( $\sigma_\varepsilon^2$ ) with variance parameters of random effects ( $\sigma_b^2$ ). Then we maximize each component separately using Equation (2.17) to obtain  $\sigma_b^2$  and  $\sigma_\varepsilon^2$ .

$$-2\log Lik_{REML} = (n - p)\log(2\pi) + \log(|V|) + R^T V^{-1} R + \log(|X^T V^{-1} X|) \quad (2.17)$$

Where  $|V|$  is the determinant of the matrix  $V(\sigma_\varepsilon^2, \sigma_b^2)$ ; Residuals  $R = Y - X\hat{\beta} = Y - \mu$ ;  $X$  are the predictors with fixed effects  $\beta$

The relative variance-covariance matrix  $S$  is a block diagonal matrix in two blocks that are nonnegative multiples of identity matrices as  $S = TS_D S_D T^T$  where  $T$  is lower triangular matrix in which all elements are zeros.  $S_D$  is a diagonal matrix in which all non-negative elements are on diagonal.

The component of fitted model ( $L$ ) is Cholesky factorization [125] of matrix  $A^*(\theta)$  described by Equations(2.18) and (2.19).

$$A^*(\theta) = \begin{bmatrix} Z^{*T} Z^* + I & Z^{*T} X & -Z^{*T} y \\ X^T Z^* & X^T X & -X^T y \\ -y^T Z^* & -y^T X & -y^T y \end{bmatrix} \quad (2.18)$$

$$A^*(\theta) = \begin{bmatrix} T^T S_D & 0 & 0 \\ 0 & I & 0 \\ 0 & 0 & 1 \end{bmatrix} A \begin{bmatrix} S_D T & 0 & 0 \\ 0 & I & 0 \\ 0 & 0 & 1 \end{bmatrix} + \begin{bmatrix} I & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \quad (2.19)$$

where  $\theta$  is a parameter vector to determine  $S(\theta) = S = TS_D S_D T^T$

The fixed effects parameters are computed by using Generalized Least Squares (GLS) which estimate parameters in a linear regression [126]. To estimate the fixed effects parameter  $\beta$  we use the GLS as shown below:

$$X^T V^{-1} Y = X^T V^{-1} X \hat{\beta} \quad (2.20)$$

$$\text{Thus: } \hat{\beta}_{(GLS)} = (X^T V^{-1} X)^{-1} X^T V^{-1} Y \quad (2.21)$$

The random effects parameters are computed by using the Best Linear Unbiased Predictor (BLUP) [127]. BLUP function to estimate random effect parameters is:

$$BLUP(b) = SZ^T V^{-1} (Y - X \hat{\beta}) \quad (2.22)$$

Apply  $\hat{\beta}$  from Equation (2.21), the final calculation for  $b$  using Equation (2.23) is as follows:

$$\begin{aligned} b &= SZ^T V^{-1} (I - X \hat{\beta} Y^{-1}) Y \\ b &= SZ^T V^{-1} (I - (X^T V^{-1} X)^{-1} X^T V^{-1}) Y^{-1} Y \\ b &= SZ^T V^{-1} (I - (X^T V^{-1} X)^{-1} X^T V^{-1}) Y \end{aligned} \quad (2.23)$$

## 5.2 Building the non-linear Mixed Effects Model

The nonlinear mixed effects model had been applied and published in [128]. The nonlinear mixed-effect model has been fitted using the maximum likelihood with a functional exponential form by Equation(2.23).

$$MOS \sim Ae^{B^*D^*RPS} | time \quad (2.23)$$

Where  $MOS$  is mean opinion scores,  $D$  is delay factor,  $RPS$  is requests per second factor, and  $A$  and  $B$  are fitted modelling values. The non-linear transformable form is exponential ( $e$ ).

Mean opinion score was represented as an exponential function of Delay ( $D$ ) and Requests per second ( $RPS$ ) as the fixed effects parameters. The random effect was time.

The non-linear predicted values for the mean opinion score via delay and the number of requests per second with a random effect of time is shown in Figure 5-1.

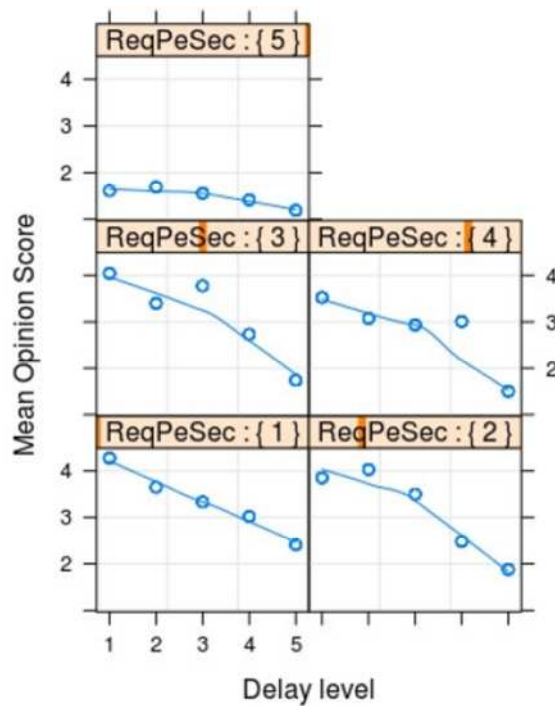


Figure 5-1: Non-linear predicted values for MOS vs. Delay and Requests per second

A trellis dot plot of random effects is described in Figure 5-2 and describes values of  $A$  and  $B$  for the random effects comprising the model described by Equation (2.23). We also investigated a nonlinear mixed-effects model that incorporates the random variable of subject (id) in the simple form of an exponential function, and it results in approximately 84 % correlation (with  $R^2$  of about 0.84.) between modelling fitted values and observed values.

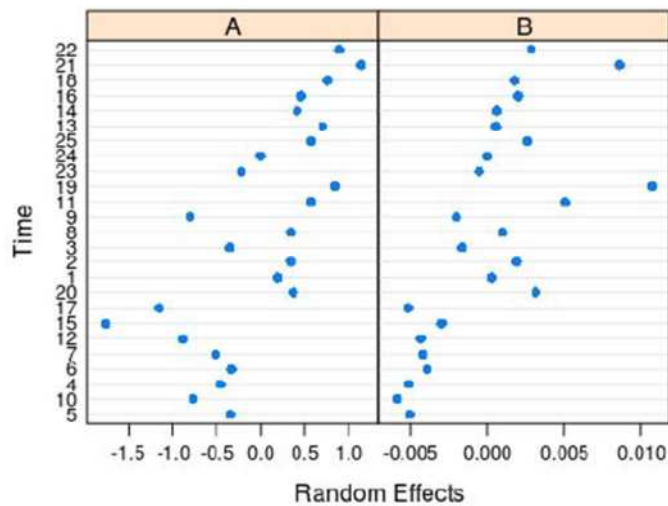


Figure 5-2: Caterpillar plot of normal probability of random variables of time in non-linear model

However, it can be seen that the linear mixed effects model considered here provides the most acceptable model for the experimental data used in this analysis. Consequently, the linear mixed effects models have been further investigated in our study.

It is also noted that not all possible non-linear transforms have been investigated in this work on the mixed effects models as they are very numerous and there is a significant time limitation for this research. However, based on our completed work, the proposed non-linear mixed effects model did not produce better results in terms of modelling of our experimental data than the linear models. However, in the future it might be possible to further investigate other non-linear mixed effects models.

### 5.3 Building the Linear Mixed Effects Model

#### 5.3.1 Model response for QoE of Web based services

The model response:

$$y = \{MOS\} \tag{2.24}$$

The values of *MOS* were observed and collected from the users in our course of experiments as described in Table 5-1. We had two groups of 9 and 12 people who attended during total 50 sessions of the experiments. The first group of 9 users attended 50 sessions, obtained 450 observations. The second group of 12 users attended 36 sessions, obtained 432 observations. The total number of observations that were recorded was 882. The second group remains the same 4 users of the first group. In total, 17 different users contribute to 882 observations.

Table 5-1: Calculation of the number of observations

Courses of Experiments				Observation
Group	Number of User	Run	Number of Content Type	
1	9	25	2	450
2	12	9	4	432
<b>Total</b>				<b>882</b>

### 5.3.2 Fixed effects component-Performance

A fixed effect is one that it can be controlled and manipulated to determine its impact on the predictability of QoE in our QoE modelling of Web based services. The principal fixed factors that we have identified are delay ( $D$ ), requests per second ( $RPS$ ), and content ( $Content$ ).

There are the three fixed factors that we mentioned earlier in Sections 4.3.1.1, 4.3.1.2, and 4.3.1.4. In Model 1, we have selected only two fixed factors viz:

$$\beta = \{D, RPS\} \quad (2.25)$$

In our enhanced models, Model 2 and Model 3, three fixed factors have been chosen as

$$\beta = \{D, RPS, Content\} \quad (2.26)$$

The fixed part of the first model, model 1, is represented as

$$\beta_0 + \beta_D D + \beta_{RPS} RPS \quad (2.27)$$

The fixed part of the enhanced models, model 2 and model 3, can be represented as:

$$\beta_0 + \beta_D D + \beta_{RPS} RPS + \beta_{Content} Content \quad (2.28)$$

In matrix notation, the fixed part of the model is:

$$\beta = \begin{bmatrix} \beta_0 \\ \beta_D \\ \beta_{RPS} \\ \beta_{Content} \end{bmatrix} \quad (2.29)$$

### 5.3.3 Cooperating random effects method for QoE of Web based services-Performance

There are three random factors as identified in Sections 4.3.1.3, and 4.3.1.4 which are the user identity ( $id$ ), time ( $time$ ), and state of mind ( $SOM$ ). They are considered as random effects as their levels in the data are samples of the whole population. Due to repeated measurements, we take into account user identity and time in the model. Time is when users accessed the web service. It was the time when network parameters

remain fixed while users interacted with the web pages, and then they finished their interaction by giving a feedback.

In the first model, model 1, two random effect factors have been chosen:

$$b = \{id, time\} \quad (2.30)$$

And in the enhanced model, model 3, three random effect factors have been chosen:

$$b = \{id, time, SOM\} \quad (2.31)$$

The random effect for the subject is

$$id_i = N(0, \sigma_{id}^2) \quad (2.32)$$

The random effect for the time is

$$time_i = N(0, \sigma_{time}^2) \quad (2.33)$$

The random effect for *SOM* is

$$SOM_i = N(0, \sigma_{SOM}^2) \quad (2.34)$$

The estimation processes assume:

$$COV(id_i, time_j) = 0 \quad (2.35)$$

$$COV(id_i, SOM_m) = 0 \quad (2.36)$$

$$COV(SOM_m, time_j) = 0 \quad (2.37)$$

Observing the covariate of the estimated random effects:  $b_{id_l}$  is the subject (user identity) effect at  $l$ , and  $b_{id_l} = N(0, \sigma_{id}^2)$ ;  $b_{time_n}$  is the time effect at  $n$  and  $b_{time_n} = N(0, \sigma_{time}^2)$ ;  $b_{SOM_m}$  is the *SOM* effect at  $m$ , and  $b_{SOM_m} = N(0, \sigma_{SOM}^2)$ ;

The matrix notation for the random part of the model is:

$$b_{id_l} = \begin{bmatrix} b_{id_1} \\ b_{id_2} \\ \vdots \\ b_{id_{17}} \end{bmatrix} \quad (2.38)$$

$$b_{time_n} = \begin{bmatrix} b_{time_1} \\ b_{time_2} \\ b_{time_{s0}} \end{bmatrix} \quad (2.39)$$

$$b_{SOM_m} = \begin{bmatrix} b_{SOM_1} \\ b_{SOM_2} \\ b_{SOM_3} \end{bmatrix} \quad (2.40)$$

### 5.3.4 Performance of the model for QoE of Web based services

#### 5.3.4.1 Introduction

The purpose of this analysis is to provide an estimate of parameters and determine measures of the reliability for their estimation.

In the Section on evaluation of metrics for QoE Web assessment mentioned in Sections 4.2.1 and 4.2.2 evaluated the metrics from application and networking perspectives to contribute to the response *MOS*, the main fixed effects are delay (*D*), and requests per second (*RPS*), the main random effects are user identity (*id*) and accessed time (*time*).

State of Mind (*SOM*) and content (*Content*) evaluated their potential effects on QoE outcome as mentioned in Sections 4.2.4 and 4.2.5. Therefore, the potential fixed effect of *Content* and potential random effect of *SOM* are integrated into the analytical modelling framework.

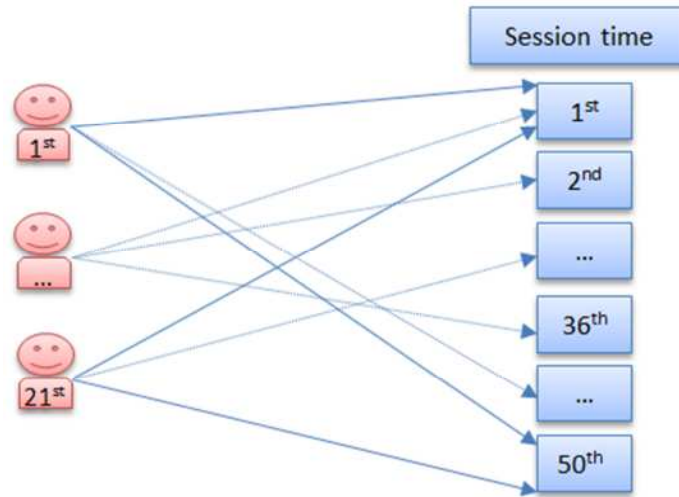


Figure 5-3: Nesting label of (*id: time*)

In our design of the model, the interaction with *id* and *time* has been used as described in Figure 5-3. In our thesis, we use the notation of (*id: time*), called the nesting label (*id: time*), showing two possible relationships that may hold between levels of *id* and the levels of *time*. The reason why we used nesting label (*id: time*) is because different error terms have been found for different levels.

To express the models more generally, it is convenient to switch to a matrix/vector representation. Fixed effect parameters  $\mu$  forms 1 dimensional column vector  $\beta$  with 9 rows and random effects vector  $b$ .

All effects are combined to yield the *MOS* ( $\mu_{ijln}$  for Model 1 and  $\mu_{ijkln}$  an enhanced model, Model 2 and  $\mu_{ijkhmn}$  for Model 3):

Model 1:

$$\mu_{ijln} = \beta_0 + \beta_D D_i + \beta_{RPS} RPS_j + b_{id_i} + b_{time_n} \quad (2.41)$$

The first model, model 1, described by Equation (2.41) takes into account the fixed effects of  $\{D, RPS\}$  and random effects of  $\{id, time\}$ . When the experiment is repeated, the fixed effects of  $(D, RPS)$  remain fixed, but the random effect of (*id: time*) is changed during the experiment. The model 1 allows unstructured covariance matrices  $(D + RPS)$  or  $(1 + D + RPS)$  in which all correlations are estimated. It allows an

additive shift for each combination level of subject  $id$  and time ( $id : time$ ) as described by Equation(2.42).

$$MOS \sim D + RPS + (D + RPS | id : time) \quad (2.42)$$

### 5.3.4.2 Model 1

Model 1 is described by Equation(2.42) as  $MOS \sim D + RPS + (D + RPS | id : time)$ . This model does not integrate such psychological effects as  $SOM$ , and  $Content$  effects. The approach of model 1 is described by Figure 5-4.

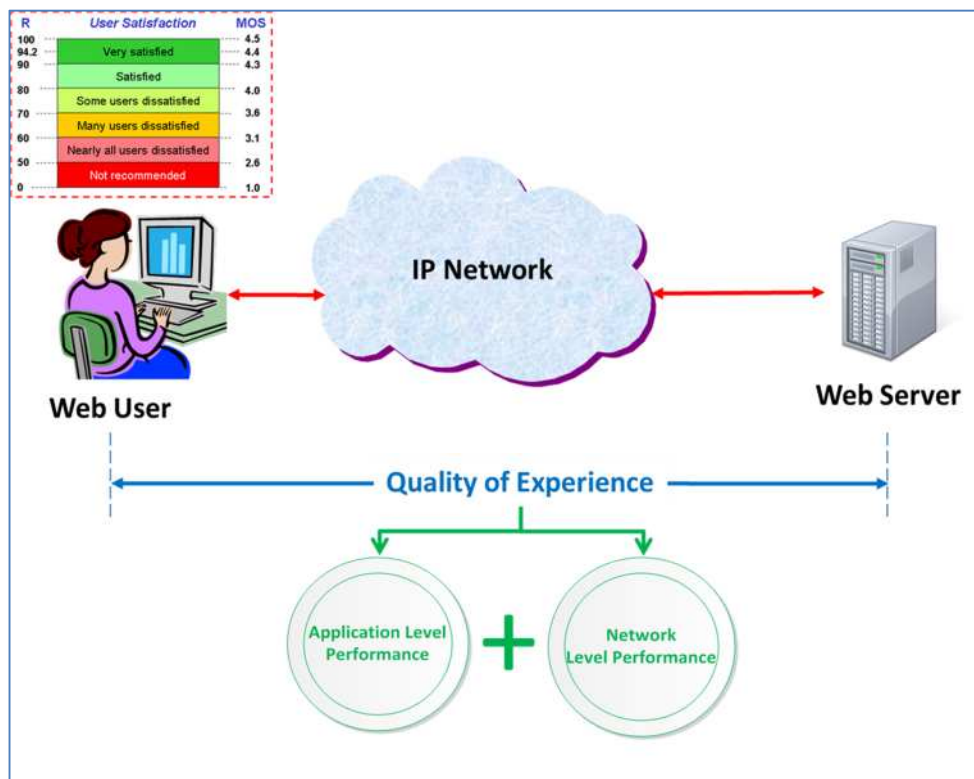


Figure 5-4: Model 1 Approach

It takes into account a variance within subject and group in which we mention that individual subjects may model intercepts and slopes that diverge considerably from population. Random effects can be extracted and saved in this model including their posterior variances. It uses unstructured covariance matrices of  $(D + RPS)$  or  $(1 + D + RPS)$  where all correlations are estimated.

$X$  is an  $n \times p$  dimensional fixed effects matrix where the structure of data and values of any covariate are used. Each fixed effect parameter of  $D$  and  $RPS$  has 5 levels ( $p_1 = 5$ ). In order to inform a linear regression, each factor used ( $p_1 - 1$ ) levels as the first level was a starting point. Therefore,  $p = 9$  represents 9 fixed levels of parameters (4 levels of  $D$ , 4 levels of  $RPS$  and the overall mean) in the fixed linear regression lines.  $n = 882$  observations. The dimension of  $X$  matrix is  $X_{882 \times 9}$ .

The  $n \times p$  dimensional fixed effect model matrix  $X$  is  $X_{n \times p}$  where the structure of data and values of any covariate are used as mentioned in Equation (2.8). Here,  $n$  is 882 (observations), and  $p$  is 9 (dimensions of fixed effect parameters of  $\beta$  are  $\{D, RPS\}$  respectively). The design matrix  $X_{(882 \times 9)}$  for model 1 with  $882 \times 9$  dimensions is of the form given in Figure 5-5. The structure of data and values of any covariate are used to create a matrix of  $X_{(882 \times 9)}$ .

The transpose of model matrix  $Z_{n \times q \times m}$ , dimensional random effect design model matrix of the fitted model 1 is mentioned in Equation(2.8). For this model,  $Z$  is a matrix of indicators of the level of grouping factor of ( $id : time$ ). Here,  $n$  is 882 (observations), and  $q$  is 9 (dimensions of  $(D + RPS)$ ), and  $m$  is total number of grouping factors, indexed by  $ij$ , thus,  $Z_{n \times q \times m} = Z_{882 \times 9 \times 738} = Z_{882 \times 6642}$  as described in Figure 5-6.

$$X_{(882 \times 9)} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 1 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Figure 5-5: Matrix  $X$  of Model 1

6642 x 882 sparse Matrix of class "dgCMatrix"

```
[1,] .. 1 .....
[2,] ..... 1 .....
[3,] ..... .. 1 .....
[4,] ..... .. .. 1
```

Figure 5-6: Matrix  $Z^T$  of Model 1

The  $9 \times 9$  matrix in Figure 5-7 is an estimate of the variance-covariance matrix of random effects associated with the grouping combined factor of  $(id : time)$ .

$$VC = \begin{bmatrix} 0.726327 & -0.5873241 & -0.4064881 & -0.473738 & -0.350695 & -0.6134636 & -0.7609828 & -0.500836 & -0.390269 \\ -0.587324 & 0.8989293 & 0.4810905 & 0.656783 & 0.385971 & 0.1881985 & 0.0702902 & 0.497971 & 0.389663 \\ -0.406488 & 0.4810905 & 0.6602133 & 0.440886 & 0.210028 & 0.0200262 & 0.1123939 & 0.334632 & -0.118380 \\ -0.473738 & 0.6567838 & 0.4408867 & 0.703738 & 0.351716 & 0.2739691 & -0.1380587 & 0.471953 & 0.288434 \\ -0.350695 & 0.3859710 & 0.2100280 & 0.351716 & 0.262960 & 0.2867361 & 0.1645012 & 0.396619 & 0.176047 \\ -0.613463 & 0.1881985 & 0.0200262 & 0.273969 & 0.286736 & 0.9930396 & 0.9340144 & 0.544981 & 0.588979 \\ -0.760982 & 0.0702902 & 0.1123939 & -0.138058 & 0.164501 & 0.9340144 & 1.8732946 & 0.251592 & 0.312269 \\ -0.500836 & 0.4979712 & 0.3346320 & 0.471953 & 0.396619 & 0.5449811 & 0.2515921 & 0.876243 & 0.325245 \\ -0.390269 & 0.3896632 & -0.1183800 & 0.288434 & 0.176047 & 0.5889791 & 0.3122694 & 0.325245 & 0.854764 \end{bmatrix}$$

Figure 5-7: Matrix of variance-covariance of random effects associated grouping factor of Model 1

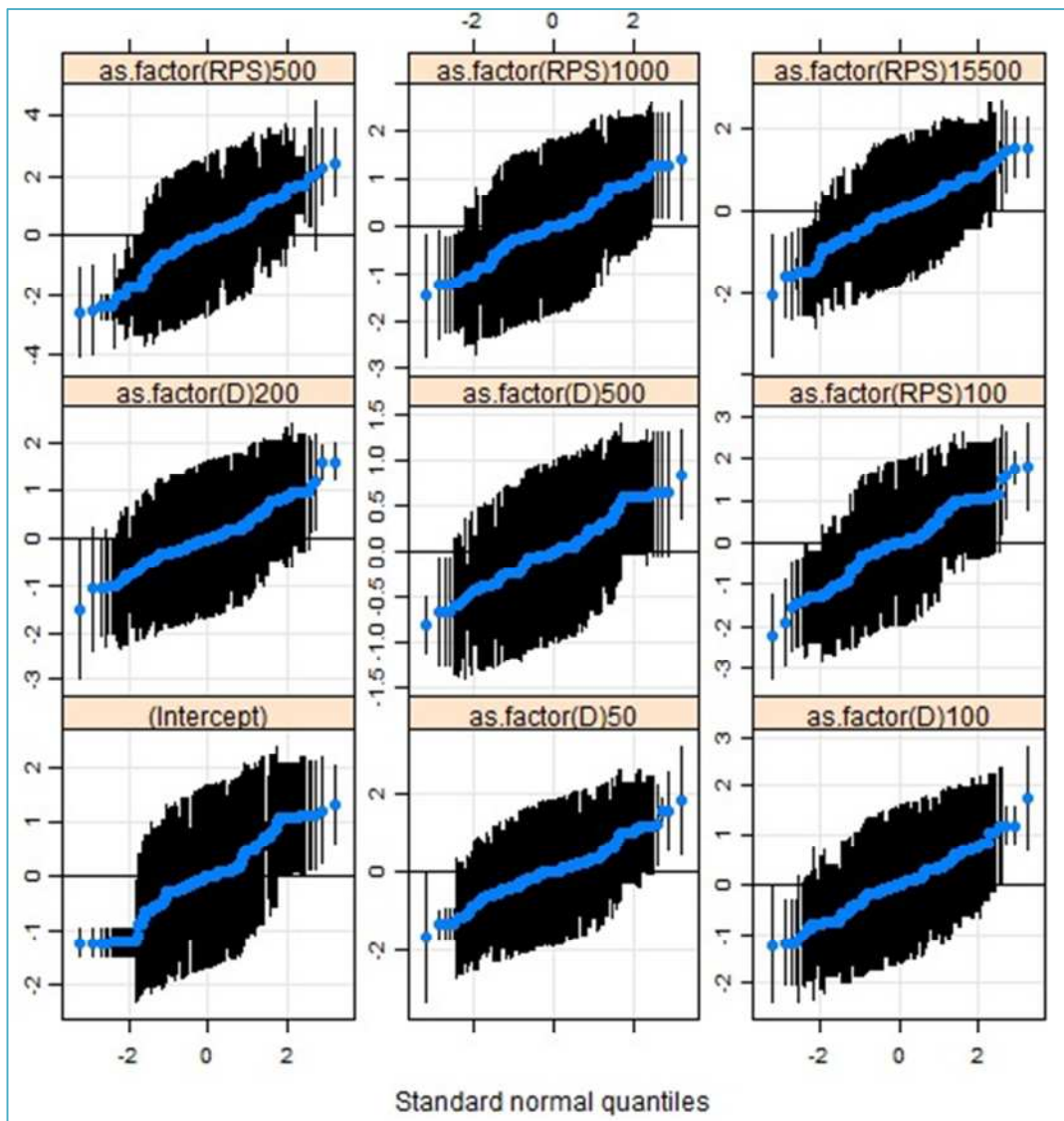


Figure 5-8: Caterpillar plot of normal probability of the conditional modes of the random effects for grouping (*id:time*) factor from Model 1

Figure 5-8 shows conditional modes of random effects (nested subject and time) from the fitted model, Model 1 described by Equation(2.42)

Figure 5-8 combined two results in one figure. The first result shows details on random effects estimates from the fitted model 1. This gives a list of  $b_i$  that corresponding to each of 6642 points of {id:time} on nice categories of intercept and each level of {D} and {RPS}. Thus, a list of  $b_i$  that corresponding to each 738 discrete points on each intercept, or each level of D, or each level of RPS is presented. The lowest value of  $b_i$  was -2.569, and the highest value of  $b_i$  was 2.439. Therefore y axis is value of random effects estimates. The highest values of each black line is upper values of 95% confidence levels, and the lowest values of each black line is lower values of 95% confidence levels, blue dot is an estimate of random each level of random effect of {id:time}.

Then the estimates of random effects of model 1 had been applied in a Q-Q plot, thus x-axis shows a standard normal population on the horizontal axis and y-axis show the random effects estimates quantiles. The Q-Q plot is used to compare a shape of distribution. The linearity of points suggests that data were normally distributed.

Model 1 described random effects of (D+RPS|(id:time)) which contains random intercepts and slopes. The random intercepts and slopes for each individual at each time point are established and they arranged in a normal probability plot in which the intervals created in this way overlap with the zero line, but some levels are clearly are either greater or less than 0 as shown in Figure 5-8. The estimate for the intercept, each level of D, and RPS is essentially 0, although the random effects {id:time} variance indicates that there is some variability in the intercepts between {id:time}. It is concluded that the random effect is necessary to be included in the model.

The 95% confidence interval for random effects is described by Equation(2.43), where SE is conditional standard deviations in each direction.

$$\text{Estimate} \pm 1.96 * \text{SE} \tag{2.43}$$

The precision of the conditional distribution of random effects in Figure 5-8 is indicated by a line that extends the 95% confidence interval.

### 5.3.5 Fixed effects for QoE of Web based services-Evaluation

To test for significance of the fixed effects, two methods have been used:

- (1) Log-likelihood or deviance comparison test using the maximum likelihood only
- (2) 95% Confidence intervals.

#### 5.3.5.1 Log-likelihood

Likelihood test is used to compare the fit of model 1 and a null model. It is based on the likelihood ratio that illustrates how many times the data are likely under the model 1 than with the null model. The logarithm function achieves its maximum value at the same points, thus, the log-likelihood is used in the maximum likelihood estimation. The log-likelihood is used to compute a p-value, and compares critical information to decide which model is superior and which model should be rejected in favour of an alternative model. The distribution of the likelihood test, based on Wilks' theorem [129], can be determined by computing the likelihood ratio ( $\Lambda$ ) for the data and comparing  $-2\log(\Lambda)$  to the chi-squared value corresponding to the approximation.

A significantly capable result is obtained when the chi-squared value is less than or equal to 0.05. The significant results are coded as '\*\*\*', '\*\*', '\*', '.' when the chi-square values are less than or equal to 0.001, 0.01, 0.05 and 0.1 respectively.

A null model (without fixed factors of  $D$  and  $RPS$ ) is tested and compared with a full model with fixed factors of  $D$  and  $RPS$  to see which one is a better fit to the data. The Model 1 and null model are described by Equations (2.42) and (2.44) respectively.

The null model is  $MOS \sim 1 + (1 | id : time)$  (2.44)

The comparison result for the two models is illustrated in Table 5-2 where the log-likelihood method has been used to compare the two models.

Table 5-2: Null model and Model 1

Model	Df	AIC	BIC	logLik	Chisq	Chi Df	Pr(>Chisq)
Null model	3	2085.6	2819.9	-1399.8			
Model 1	55	2199.5	2462.5	-1044.8	710.05	52	<2.2e-16***

Based on the values of the log-likelihood (logLik), the model with a higher value is better [130]. LogLink of model 1 is 355 higher than that of the null model. Based on the p-value  $< 2.2e-16$ ; the p-value is significant as it is  $\leq 0.001$  ('\*\*\*'). Thus it can be concluded that model 1 is significantly superior to the null model. Including both *D* and *RPS* as predictors and it is better than omitting them from the model.

### 5.3.5.2 95% confidence intervals

95 % confidence intervals have been estimated by Equation (2.43) in which the standard error is multiplied by  $\pm 1.96$  and added to the coefficient estimation. When the 95% confidence interval calculating procedures are repeated, 95% of the time it will contain the true parameter values.

Figure 5-9 shows the estimation of fixed effects factors of *D* and *RPS* of model 1, and their 95% confidence intervals. The blue line is an estimate of the fixed effect factors. The red and green lines are plotted for their 95% confidence intervals in which the red line represents the lower confidence value, and the green line represents the upper values of the 95 % confidence interval. As seen from the figure, the estimation of fixed effects factors of *D* and *RPS* is within the required 95% confidence intervals.

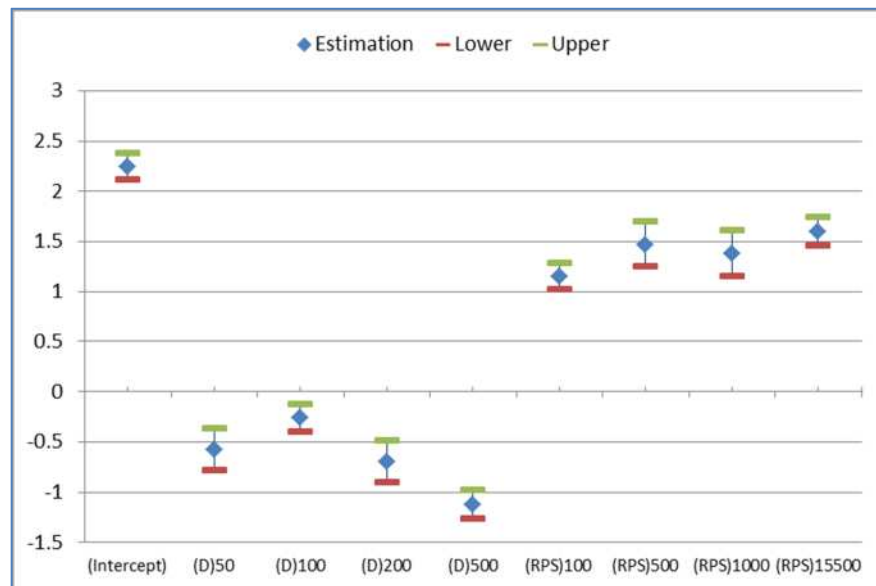


Figure 5-9: Estimation of fixed effects of *D*, and *RPS* and 95% confidence intervals of Model 1

Table 5-3: Ratio of Std. Error/Estimation for fixed factors of  $\{D, RPS\}$  of Model 1

Parameter	Estimation	Std. Error	Std. Error /Est.
(Intercept)	2.244481	0.067927	0.030264
(D)50	-0.57665	0.104453	-0.18114
(D)100	-0.26357	0.070936	-0.26913
(D)200	-0.69551	0.106353	-0.15291
(D)500	-1.12339	0.070844	-0.06306
(RPS)100	1.153231	0.067053	0.058143
(RPS)500	1.471292	0.113684	0.077268
(RPS)1000	1.381258	0.115966	0.083957
(RPS)15500	1.598749	0.071731	0.044867

**Std. Error /Estimation  $\leq 0.25$**

The authors in [131] give a caution that the 95% confidence intervals (95% CI) in Equation(2.43) are approximately valid only if the ratio of the Std.Error / Estimate is less than or equal to 0.25 ( $\leq 0.25$ ). To test the estimate of 95% CI, the ratio of Std.Error / Estimate for fixed effects factors of  $\{D, RPS\}$  it has been shown in Table 5-3 that where the ratio is small ( $\leq 0.25$ ), the estimation for 95% CI is valid.

### 5.3.6 Random effects method for QoE of Web based services-Evaluation

To specify correctly the covariance matrix for the random part of the model, it is important to understand random variation in the data, and to make sure that effects are reflected in the fixed part of the model. To test for random effects, likelihood ratio (deviance) tests and correlation parameters have been used.

For the random effects, it involves specifying ( $id : time$ ). This assumes that the random effects for each level of  $time$  with respect to a given user are independent and have the same variance. Suppose that we have 17 different users participating in the experiment. In the experiment, 50 sessions have been tested. The number of observations is 882. To specify particular assessment for QoE of Web based services, each user's assessment is divided amongst 50 different sessions/levels of time. It not only specifies each user assessment, but also the session time when the performance was changed, then the  $id$  factor is implicitly nested within  $time$ .

A vector of random effects is estimated for each combination of (*id : time*) . The length of vector is the level of factor *D* and *RPS* as a combination of (*D + RPS*) and we call the vector an estimated vector.

To evaluate whether adding this vector is necessary for fitting our data, we test and make a comparison between a model with the estimated vector and a model without the estimated vector.

Model with the estimated vector is described by model 1. Model without the estimated vector is built and described by Equation(2.45), called model(2.45), in which the estimated vector is not added.

$$MOS \sim D + RPS + (1 | id : time) \tag{2.45}$$

Table 5-4 compares model 1 and model (2.45) in terms of the estimated vector. The p-value (*p - value* = 0.01908 ≤ 0.05\*\* is significant) indicates that model added the estimated vector is superior compared to model without the estimated vector.

Table 5-4: Model (2.45) vs. Model 1

Model	Df	logLik	Chisq	Chi Df	Pr(>Chisq)
Model (2.45)	11	-1077.5			
Model 1	55	-1044.8	65.573	44	0.01908*

## 5.4 Model of QoE of Web based services-Evaluation

In this Section, the whole model is evaluated based on:

- 1) If the residuals are normally distributed;
- 2) Information Criteria;
- 3) What percentage of variance the model accounts for.

### 5.4.1 Residual distribution

Residuals are an important part of the model. They are elements of variation unexplained by the fitted model. It is expected that they are roughly normal and approximately independently distributed with a mean of 0 and some constant variance [132].

Figure 5-10 shows a distribution of residuals of Model 1 described by Equation(2.42). The dot plot is the collection of points, the residuals, along the left y-axis. The points on this plot form a nearly linear pattern, which indicates that the normal distribution is a good model for this data set.

As can be seen from Figure 5-10, the residuals of the model are close to a normal distribution, and approximately independently distributed with a mean of 0.

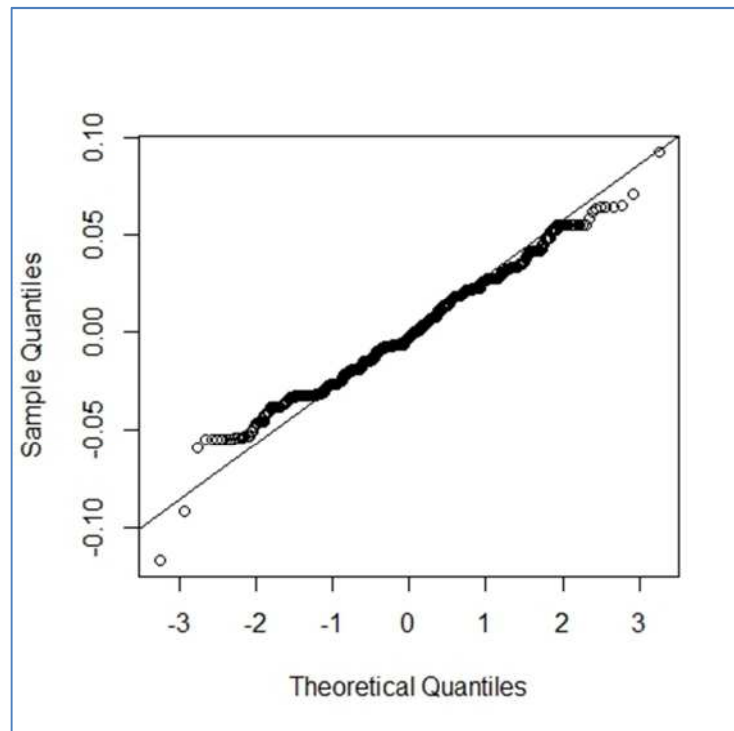


Figure 5-10: Residuals distribution of Model 1

#### 5.4.2 Information criteria

In this section, information criteria are used to make comparisons between different cases of the built models before choosing a model. The information criteria are based on the log of likelihood or deviance. They are not statistical tests, but rules for restricted maximum likelihood testing as mentioned in Equation(2.17) which relate to the dimension of the model, and effective sample size.

The parameter estimation criterion is the restricted maximum likelihood (REML). REML estimates  $\sigma^2$  in a fixed effects regression by  $\frac{SSR}{n-p}$  where SSR is the residual sum of squares.

The null model by Equation(2.44), the non-estimated vector-model by Equation(2.45), and full model 1 by Equation(2.42) are tested and compared.

Based on the information from Loglik illustrated in Table 5-2 and Table 5-4, the highest values of Loglik are chosen, Model 1 is superior with a p-value of  $0.01980 \leq 0.05$ .

Table 5-5 gives information about fixed estimation for model 1. The estimate of each  $D$  and  $RPS$  level is provided in the second column. To assess the significance of  $D$  and  $RPS$  as predictors, the t values are tested. The intercept is expected to have a high value since  $MOS$  scores are always a positive number, and it is expected that the intercept would be different from 0. Since the number of our observations is 882, the data is considered to be a large enough sample, thus it is expected that the t-distribution will approximate closely to the normal distribution.

Based on this point [119], those predictors are significant if their t-values are greater or equal than 2 ( $t_{value} \geq 2$ ) or less or equal than -2 ( $t_{value} \leq -2$ ), or the p-values calculated by Equation(2.47) are less or equal than 0.05 ( $p_{value} \leq 0.05$ ). The fourth column of Table 5-5 indicates the t-values of each level of  $D$  and  $RPS$ , and those values are higher than 2, or smaller than -2 or  $p$  values are less than 0.01. Thus, all fixed factors of  $D$  and  $RPS$  are deemed to be significant predictors.

The t-value has been calculated by dividing the Estimate with the Std. Error as described by Equation(2.46). The  $p$  value is calculated by the Equation(2.47):

$$t_{value} = \frac{\text{Estimate}}{\text{Std.Error}} \quad (2.46)$$

$$p_{value} = 2 * (1 - \text{pnorm}(t_{value})) \quad (2.47)$$

Table 5-5: Fixed effects of Model 1

Fixed effects	Estimate	Std. Error	t value	p value
<b>(Intercept)</b>	2.24448	0.06793	33.04	0.00000e+00
<b>(D)50</b>	-0.57665	0.01044	-5.52	3.38000e-08
<b>(D)100</b>	-0.26357	0.07094	-3.72	2.02649e-04
<b>(D)200</b>	-0.69551	0.10635	-6.54	1.00000e-10
<b>(D)500</b>	-1.12339	0.07084	-15.86	0.00000e+00
<b>(RPS)100</b>	1.15323	0.06705	17.20	0.00000e+00
<b>(RPS)500</b>	1.47129	0.11368	12.94	0.00000e+00
<b>(RPS)1000</b>	1.38126	0.11597	11.91	0.00000e+00
<b>(RPS)15500</b>	1.59875	0.07173	22.29	0.00000e+00

### 5.4.3 Correlation

Correlation is another method to evaluate the fitted model; it summarizes the strength of a relationship between two variables.

There are several correlation coefficients that can be calculated. The two most commonly used methods are the Pearson [133] and Spearman [134] correlation methods.

Pearson correlation tests require both variables to be measured on an interval or ratio scale. Spearman's correlation test requires at least a variable of ordinal data, and then the Spearman is required for calculation instead of Pearson. Each variable is ranked separately in order to put the value of the variable in order and numbering them. The lowest gets rank 1 and highest gets rank 5, for example, in the *MOS*.

The sample correlation of both Spearman and Pearson between observed values and fitted response in the property of least square regression are described by Equation(2.48). The Spearman correlation coefficient is defined as the Pearson correlation between ranked variables [135, 136].

The correlation coefficient is a number between +1 and -1 which shows the magnitude and direction of association between two variables. The magnitude is the strength of the correlation. The correlation is stronger when the value of  $R$  is closer to either +1 or -1. Direction shows how variables are related. The two variables have positive relationship if the correlation is positive; otherwise, the two variables have a negative relationship.

$R$  is defined as follows:

$$\begin{aligned}
 R &= \frac{\text{cov}(\text{fitted}, \text{observed})}{\sqrt{\text{var}(\text{fitted}) * \text{var}(\text{observed})}} \\
 &= \frac{\sum (\text{Observed}_i - \overline{\text{Observed}})(\text{Fitted}_i - \overline{\text{Fitted}}) / N}{\sqrt{\sum (\text{Observed}_i - \overline{\text{Observed}})^2 \sum (\text{Fitted}_i - \overline{\text{Fitted}})^2 / N * N}} \\
 &= \frac{\sum (\text{Observed}_i - \overline{\text{Observed}})(\text{Fitted}_i - \overline{\text{Fitted}})}{\sqrt{\sum (\text{Observed}_i - \overline{\text{Observed}})^2 \sum (\text{Fitted}_i - \overline{\text{Fitted}})^2}}
 \end{aligned}
 \tag{2.48}$$

Where :

$\text{cov}(\text{fitted}, \text{observed})$  = the covariance of fitted model and observed values of  $MOS$

$\text{var}(\text{fitted})$  = the variance of fitted modelling value

$\text{var}(\text{observed})$  = the variance of observed  $MOS$  value

$N$  = the number of cases

In this research, the  $MOS$  is on an ordinal scale, and then a Spearman test is applied rather than using a Pearson test.

Table 5-6: Correlation test

Correlation Method	$R$	$R^2$	Note
Pearson	0.9997	0.9998	Not used Not suitable for ordinal scale
<b>Spearman</b>	<b>0.9716</b>	<b>0.9441</b>	Used for $MOS$ in ordinal scale

As shown in Table 5-6, Spearman's  $R = 0.9716$  indicates a strong positive relationship between fitted mixed effect model, model 1, observed values, and the  $MOS$ . The coefficient of determination of  $R^2$  is 0.9441. Correlation between observed and fitted values is calculated to see what percentage of variance is accounted for by the model based on the Spearman test. Correlation between observed and fitted values of model 1 as calculated as 0.9441, meaning that it can explain 94.41% of the variance accounted for by the model.

Significance of model 1 is shown in Table 5-7 in which the Spearman value(s), p-value of the sample statistic, and correlation have been computed.

Table 5-7: Test of Significance of Model 1

Spearman's rank correlation rho		
Model 1	S	p-value
Value	3229472	<2.2e-16 “***”
Alternative hypothesis	true rho is not equal to 0	
Sample estimates:	rho: 0.9717592	

## 5.5 Discussion and Conclusions

As noted earlier, the data that is collected from users may be a potential source of error, since users are all different; this clearly occurs when the group of users is small-as in our case. Thus, we have prepared a model that takes random effects into account as a necessary feature for a potential source of a random effect.

As previously noted, the groups of users in the study are a local group of undergraduate and post-graduate students at Massey University, but the interest of this study for QoE assessment is not just about that group, it is about the effect presented by a general group of users, therefore there is a random effect representing an uncontrollable variation amongst groups that is required for the model.

The mixed effects model described here takes into account effects such as objective parameters like  $D$ ,  $RPS$  and the uniqueness of individuals themselves during the course of the experiments together with other potential covariates.

The linear mixed effects model shows a good correlation between fitted modelling values and the observed values. When the random variable is a combination of nested ( $id : time$ ) with the estimated shift vector of  $\{D, RPS\}$ , its correlation is 94.431%. It is seen that a linear mixed effects model has provided the most acceptable model for our experimental data.

## Chapter 6 An Enhanced Analytical Framework Involving Psychological Support

*Chapter 6 gives an analysis of the integrated model in which human perception related to State of Mind, and content effects has been integrated into the assessment. The model demonstrates an acceptable correlation between values of the fitted analytical model and observed user scores. The enhanced assessment helps to improve model performance and provides a deeper understanding of users' scores.*

### 6.1 Model 2 and model 3

To understand further changes in specific cases of users' perception and for un-matching cases of our intuition as described in Chapter 4, further metrics that are related to content and users' state of mind had been proposed into our customer QoE for Web browsing model.

The enhanced models:

Model2:

$$\mu_{ijkln} = \beta_0 + \beta_D D_i + \beta_{RPS} RPS_j + \beta_{Content} Content_k + b_{id_i} + b_{time_n} \quad (2.49)$$

Model 3:

$$\mu_{ijklmn} = \beta_0 + \beta_D D_i + \beta_{RPS} RPS_j + \beta_{Content} Content_k + b_{id_i} + b_{time_n} + b_{SOM_m} \quad (2.50)$$

The enhanced models, model 2 and model 3 are described by Equations (2.49) and (2.50), with more detail given in Equations (2.51) and (2.52). The models take into account the fixed effects of  $\{D, RPS, Content\}$  and random effects of  $\{id, time, SOM\}$ . The enhanced models allow an additive shift for each combination of subject and time ( $id : time$ ), and the variances are the variance of the effects of level  $i$  of  $\{D, RPS\}$  on a completed set of indicators for the level involving the random factor of  $SOM$ . Random effects for each combination of time and user identity ( $id : time$ ), and for  $SOM$  are allowed in the model.

$$MOS \sim D + RPS + Content + (D + RPS | id : time) \quad (2.51)$$

$$MOS \sim D + RPS + Content + (D + RPS | id : time) + (1 | SOM) \quad (2.52)$$

Model 2, described by Equation(2.49), is an enhancement of Model 1. It integrates the fixed effect of *Content* into the model. Model 3, described by Equation(2.50), is an enhancement of Model 2 in which a random effect of *SOM* is incorporated into the model.

## 6.2 Enhancements using Content

### 6.2.1 Model development

The purpose of this analysis is to provide an estimate of the impact made by the parameter of *Content* and to measure the reliability of its estimation. The analysis approach is described in Figure 6-1 where QoE for web browsing is assessed via metrics derived from Network, Application and Content.

The content has been tested as mentioned in Section 4.2.5. The result of this test is shown on Table 4-5: MOS vs. Content. Based on this result, we see that the content has a potential impact on users' scores especially when the users' scores are higher than a level of 4 or 5. Its effect is significant at the cases of (very) satisfied users, and is not indicated in the case of not-recommended and dissatisfied users.

In the scope of our research, *Content* is objectively reflected by different downloading times. However, it is a fact that users may have different perceptions for different kinds of content that they may (not) be interested in. The natural concept of content is not only fully reflected by differences in downloading time. However, its narrow concept, in terms of download time, is technically acceptable. Here, our ambition is to assess *Content* based on its narrow concept as described by different download times. Other concepts for content will need further study.

The model described by Equation(2.51), namely Model 2, has integrated a new fixed factor of *Content* into its specification.

In the approach described by Figure 6-1, we analyse the effect of both Networking and Application metrics that contribute to QoE measurement. The Networking metrics that are considered in this approach is delay, while the Application metric consists of

requests per second. The Application metric is controlled by the web-server. Based on real measurements of these two metrics and an integration of psychological effect of Content, we shall determine their impact upon the MOS. During the experiment, the influence upon web-related activity by users, such as a user’s patience and boredom, was also taken into account qualitatively, but not measured quantitatively.

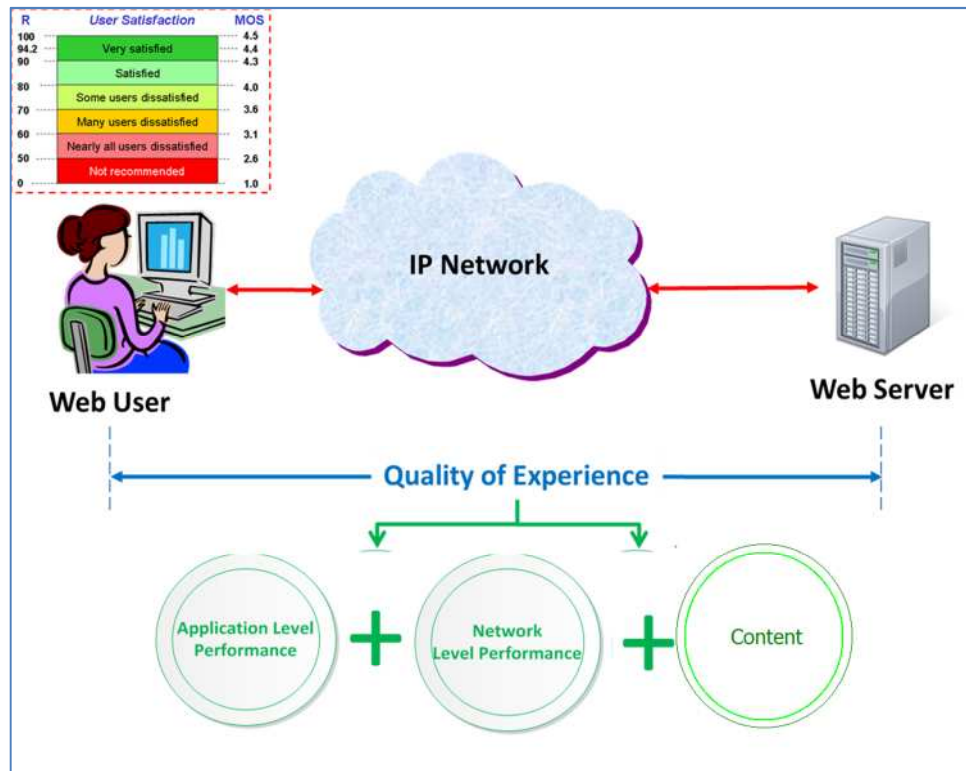


Figure 6-1: Enhancement approach using Content

### 6.2.2 Comparison of model 2 with model 1

Table 6-1 shows a comparison of two models (with/without *Content* ) structural assessment based on the likelihood test approach, Akaike information criterion (AIC) [125], and Bayesian information criterion (BIC) [137].

Table 6-1: Result of model structural assessment based on likelihood test approach

Model	Df	AIC	BIC	logLik	Chisq	Chi Df	Pr(>Chisq)
Model 1	55	2199.5	2462.5	-1044.8			
Model 2	58	2157.1	2434.4	-1025.5	48.438	3	1.718e-10***

<sup>1</sup>Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Note: <sup>1</sup>if p-value  $\leq 0.001, 0.01, 0.05, 0.1,$  and  $1,$  it is coded as '\*\*\*', '\*\*', '\*', '.', and '' respectively. The p-value is significant if p-value  $\leq 0.05.$

Based on the AIC, and BIC assessment of a model [138], a smaller value is preferred. Based on the log-likelihood, assessment of model, a larger value is better. Thus, model 2 with a fixed factor of *Content*, is superior to model 1 without *Content* as  $p \leq 1.718e - 10 \leq 0.001$  '\*\*\*' is significant.

### 6.2.3 Estimation of model 2 parameters

Figure 6-2 shows estimation of factors  $\{D, RPS, Content\},$  and their 95% confidence intervals for Model 2.

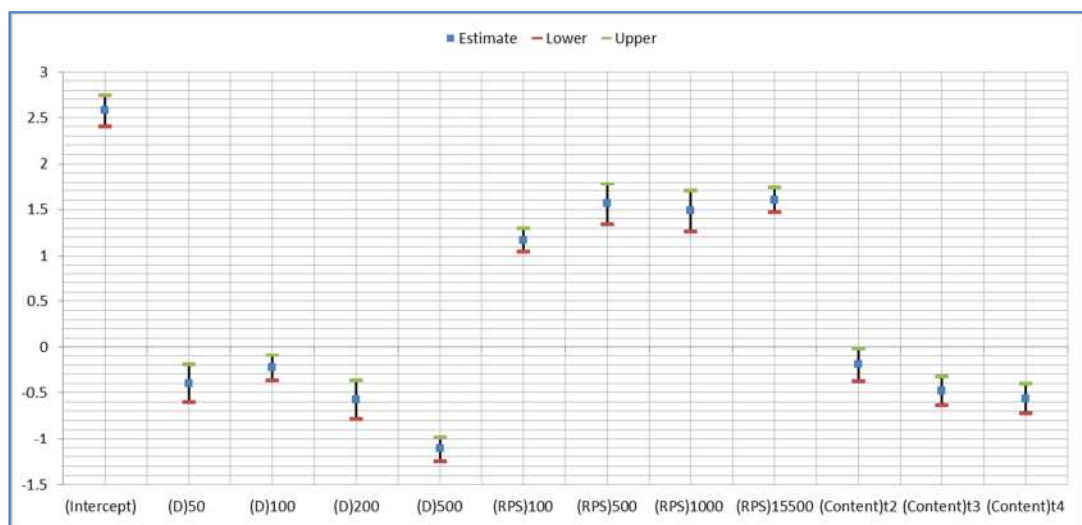


Figure 6-2: Estimation of fixed factor of *D, RPS* and *Content* in Model 2

In Figure 6-2, the blue dot is estimation of the factors, and red and green lines are upper and lower 95% confidence intervals respectively. As seen from the figure, the estimation of fixed effects is within the required 95% confidence intervals.

Table 6-2 gives information about estimation of *Content* in Model 2. The p-value is estimated by Equation(2.47). Those estimated values of *Content* at different levels are significant as their p-values are  $\leq 0.05.$  The t-values are  $\leq -2$  or  $\geq 2$  that satisfy the requirements for a normal distribution as mentioned in subsection 5.4.2.

Table 6-2: Mixed effects estimation of Model 2

Fixed effects	Estimate	Std. Error	t value	p value
(Intercept)	2.57506	0.08737	29.473	0.0000000000
(D)50	-0.39930	0.10402	-3.839	0.0001236023
(D)100	-0.22930	0.06922	-3.313	0.0009242945
(D)200	-0.57431	0.10669	-5.383	0.0000000733
(D)500	-1.11400	0.06840	-16.286	0.0000000000
(RPS)100	1.17220	0.06381	18.369	0.0000000000
(RPS)500	1.56169	0.11268	13.859	0.0000000000
(RPS)1000	1.48571	0.11293	13.157	0.0000000000
(RPS)15500	1.60351	0.06934	23.124	0.0000000000
(Content)type2	-0.19565	0.09296	-2.105	0.0353112328
(Content)type3	-0.47879	0.07977	-6.002	0.0000000019
(Content)type4	-0.56218	0.08046	-6.987	0.0000000000

**Note:** t-value $\geq$ 2 or t-value $\leq$ -2 ; p-value $\leq$ 0.05

The assessment shows an acceptable correlation between values of the fitted analytical model and observed user scores. The QoE estimates are aimed at providing service and planning, understanding user opinions, and they also target actual customer opinion prediction.

From the results, Content contributes an improvement in determining the users' perception. One example that can be applied in industry is a content control in regional service planning to improve the quality of service matching a user's demand.

### 6.3 Enhancements Using State of Mind

The purpose of this analysis is to provide an estimate of the parameter known as the state of mind and a measure of the reliability of its estimation. QoE for web browsing is assessed via metrics derived from the three areas of Network, Application and Content as well as taking in account the random effect of a users' state of mind.

As mentioned in 4.2.4, state of mind (*SOM*) has been analysed during the course of our experiments. Users' states of mind are a random distribution via time and under any circumstances. The model 3 described by Equation(2.52) has been integrated with a random effect for the state of mind.

Table 6-3: Results of model structural assessment based on likelihood test approach (cont.)

Model	Df	AIC	BIC	logLik	Chisq	Chi Df	Pr(>Chisq)
<b>Model 1</b>	55	2199.5	2462.5	-1044.8			
<b>Model 2</b>	58	2157.1	2434.4	-1025.5	48.438	3	1.718e-10***
<b>Model 3</b>	59	2130.4	2412.5	-1006.2	28.717	1	8.376e-08***

According to the likelihood as described in Table 6-3, Model 3 is one more parameter than model 2. In the favour of the “smaller is better” for both AIC and BIC, model 3 is superior to model 2. In favour of the “larger is better” of log-likelihood (logLik), model 3 is superior to model 2.  $p_{value} = 8.376e - 08 \leq 0.001$  '\*\*\*' shows a significant test ('\*\*\*') for model 3. Thus, model 3 is superior to model 2.

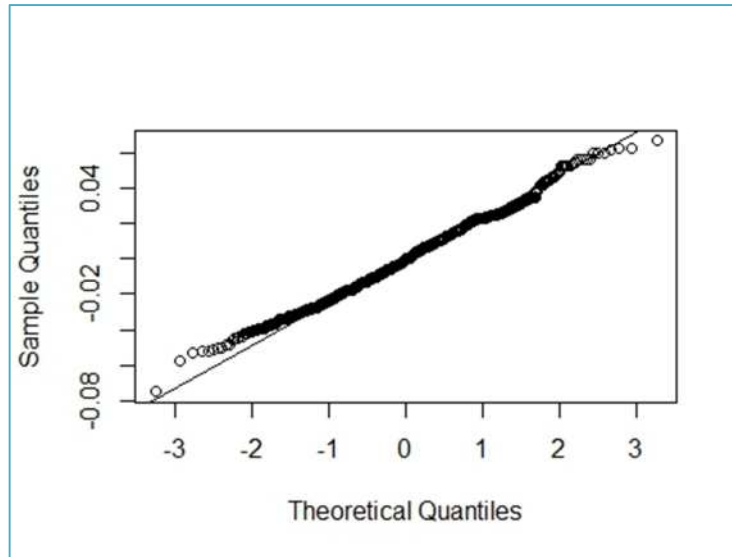


Figure 6-3: Residuals distribution of Model 3

Figure 6-3 shows a normal Q-Q plot and Q-Q line for model 3. It is roughly normal, and closer than that in model 1. It is approximately independently distributed with a mean of 0 which is much closer to 0 than that found for model 1.

The linear regression between observed and the fitted mean opinion score is shown in the figure with  $R^2$  at roughly 0.9441 (Spearman test) for model 3 giving the respective best fits for the data.  $R^2$  is calculated by Equation(2.48).

The conditional modes of the random effects for each of *SOM* are presented in Figure 6-4. It indicates normal probability of random variable of *SOM* and its 95% confidence intervals.

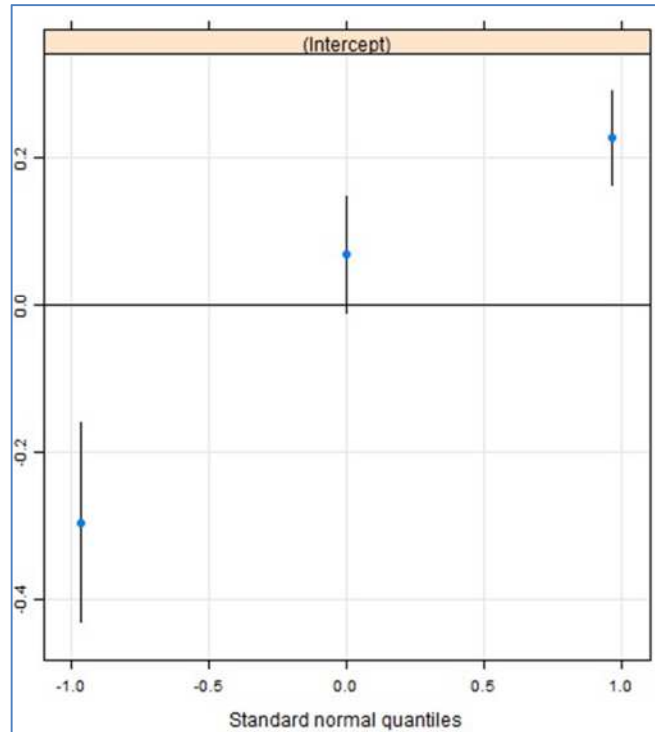


Figure 6-4: Conditional modes of the random effects for *SOM* factor

Estimated values for random effects of *SOM* in the Model 3 are shown Table 6-4. Those values are either greater or less than 0, and it concluded that the random effect of *SOM* should be introduced into model 3.

Table 6-4: Estimated values of random effect of *SOM*

<b>SOM</b>	<b>Intercept</b>
<b>Bored</b>	0.0682959
<b>Normal</b>	0.2261464
<b>Stress</b>	-0.2944423

The conditional modes of the random effects for each nested combination of (*id : time*) are presented in Figure 6-5. It shows the normal probability of random variables of subject and time (*id : time*) and its 95% confidence intervals.

Figure 6-5 combines two results into one figure. The first result shows details of random effects estimates from the fitted model 3. This gives a list of  $b_i$  that corresponds to each of the 6642 points of {id:time} on the intercept and each level of {D} and {RPS}. Thus, a list of  $b_i$  that corresponds to each of 738 discrete points on each Intercept, or each level of D, or each level of RPS. The lowest value of  $b_i$  was -2.429, and the highest value of  $b_i$  was 2.38. Therefore the y-axis is a value of the random effects estimates. The highest value of each black line is the upper value of the 95% confidence level, and the lowest value of each black line is the lower value of the 95% confidence level, while the blue dot represents an estimate of the average random value of the random effect of {id:time}.

Then, the estimates of the random effects are applied in a Q-Q plot, thus the x-axis shows a standard normal population on the horizontal axis and the y-axis shows the random effects estimate quantiles. The Q-Q plot is used to compare the shape of the distribution. The linearity of points suggests that the data were normally distributed.

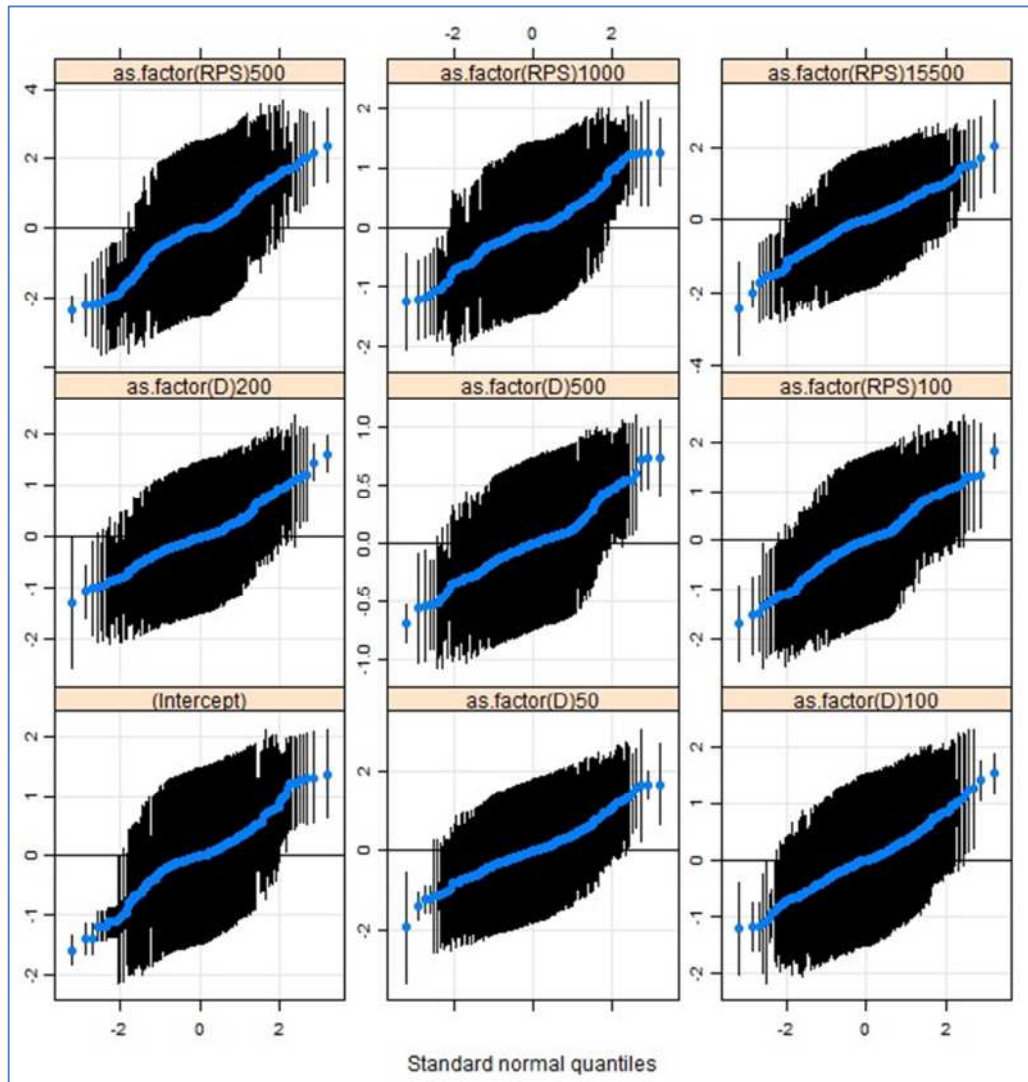


Figure 6-5: Caterpillar plot of normal probability of the conditional modes of the random effects for grouping (*id: time*) factor from Model 3

Figure 6-6 shows the estimation of fixed effects factors of  $\{D, RPS, Content\}$  in Model 3. The 95% confidence intervals are estimated by red and green lines in which the first represents the upper confidence interval values and the latter represents the lower bounds.

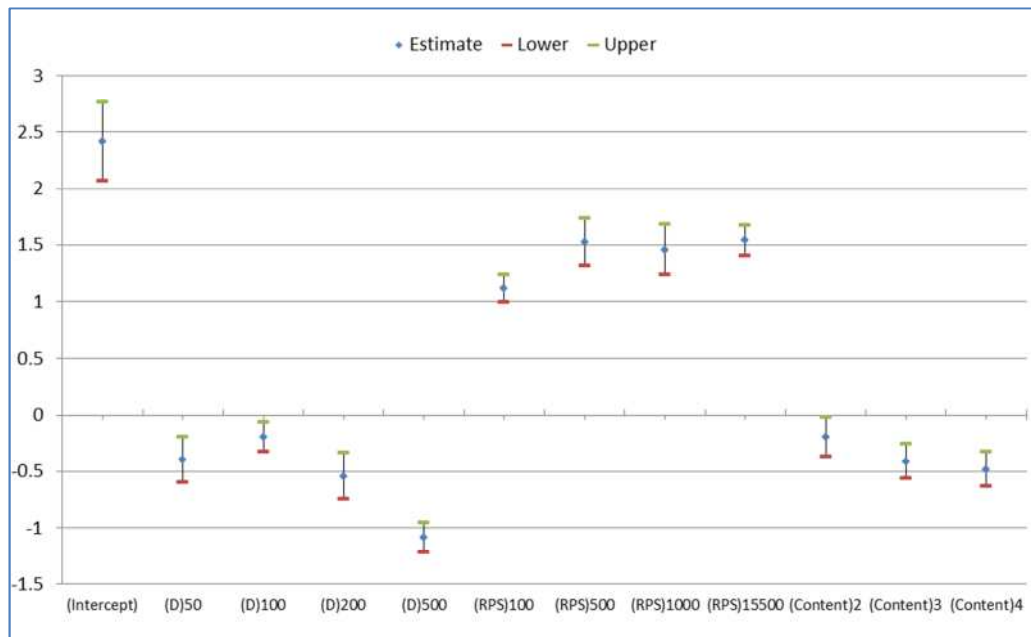


Figure 6-6: Estimation of fixed effects of *D*, *RPS* and *Content* and 95% confidence intervals of Model 3

## 6.4 Discussion and Conclusions

As noted earlier, the data that is collected from users may be a potential source of error because:

- 1) There is variability between users in how they respond to the experiments that must be accounted for in our analysis.
- 2) The responses from the same participants are likely to be correlated. For example, if the response from user of ( $id = 1.3$ ) is above the mean of response of the first test at time  $i$ , it is likely that (s)he will be above the mean of response of test 2 at time  $i + 1$ .
- 3) We observe 882 responses, but only from 21 users (with a total of 17 different users), if we make inferences about users in general, our effective sample size is 17 rather than 882.

Thus, we have seen the above issues and applied the mixed effect model which has taken the potential source of error into account, as a source of fixed and random effects. The fixed effects explain the response itself, while the random effects explain the variance of the response.

In this chapter, we presented an integration of the human factor of state of mind as well as the uniqueness of individuals themselves during the course of experiments into

model 2. It has become clear that *SOM* is not easy to capture in a realistic situation although it is needed to support complex analysis from both a psychological and machine learning perspective. *SOM* changes over time for each individual, this leads to the observation that shows changing *MOS* scores even though the controllable effects of *D* and *RPS* remain fixed. *SOM* is integrated as a random factor because user's state of mind change randomly during the experiments between subjects and within each subject. In the scope of our research, *SOM* is subjectively measured by users' feedback.

In our terminology, the *Content* is considered as objective factor based on download time. However, it is a fact that, content is a user specific factor in terms of particular users. Although, the definition of *Content* in our experiment has been rather narrowed in comparison to the actual concept of content; however, its effect is still indicated in Section 4.2.5, the content metric for QoE of Web based services. Therefore, that factor is necessary for QoE of Web based services. Both *SOM* and *Content* have been integrated in Model 3.

Based on results described in Sections 6.1 and 0, the integration of *SOM* and *Content* remains the correlation between fitted values and observed values, giving a value of  $R = 0.9716$  and  $R^2 = 0.941$ . However, more importantly, the integrated factors improve the model performance based on information criteria where maximum of log likelihood, dimension of model, and effective sample size have been used to assess the result.

## Chapter 7 Summary, Conclusions and Future Work

*This chapter summarizes the main conclusions from our research work and proposes future work to extend the analysis of QoE of Web based services. Future work includes an objective assessment of state of mind and an extension of the content concepts involving QoE assessment of Web based services.*

### 7.1 Summary

The thesis started with the concept of QoE and background needed for the study. Research problems have been described as: (1) what are the objective parameters to be measured? and (2) how to measure them for the case of web browsing. To answer those questions, a combination in our knowledge from establishment of QoE in web based services has been identified and the current state-of-the-art in QoE modelling for web based services have proceeded in which the combination are filled by: (1) Using Taguchi Robust design method to mitigate time and cost required in subjective evaluations (2) Defining objectively measurable technical inputs for QoE of web based services which are based on both networking and application performance (3) Understanding the psychological effects such as boredom, state of mind and other effects related to content.

The literature has been reviewed and discussed in Chapter 2. In this chapter, there were many different types of assessments reviewed for QoE in general and QoE of Web based services in particular and those are (1) subjective quality evaluation, (2) objective quality evaluation and (3) objective-subjective correlations for measuring QoE. The current status and trends of QoE of Web based services has been discussed. The chapter ends with summary and discussion about modelling QoE for web based services.

Experimental design, test bed, web resources, and data collections for our experiments were proposed in Chapter 3. The experiments were designed following the Taguchi method approach. Web resources have been designed by us with four different types of

content on four different sites. Data collection has been performed in a test bed located at Massey University. A session known as the controlling session was introduced to reduce boredom and memory effects which occurred during the course of our experiments.

Chapter 4 represents the evaluation of each metric involved in a framework; then, the framework for QoE of Web based services was proposed. The scientific approach requirements for such a model are (1) the QoE model should correlate well with MOS scores. (2) Observations should be taken in an objective manner to infer QoE and understand both subjectively and objectively the psychological factors. The analytical modelling proposed framework is an integration of comprehensive metrics collected from subjective, objective parameters and psychological effects.

Chapter 5 investigates the applicability of a mixed effects model for predicting QoE in World Wide Web based multi-media services. The mixed effects model brings all effects together to yield the *MOS* in which objective parameters of  $\{D, RPS, Content\}$ , other subjective parameters and the uniqueness of individuals themselves during the course of experiments. The mixed effects model decomposes the outcomes of an observation into fixed effects (population mean), and random effects (group specific and correlation structure of variations amongst users). There are three models that have been analysed in Chapter 5 and 6 which are (1) Model 1 with fixed effects of  $\{D, RPS\}$  and random effects of  $\{id, time\}$  (2) Model 2 with fixed effects of  $\{D, RPS, Content\}$  and random effects of  $\{id, time\}$  (3) Model 3 with fixed effects  $\{D, RPS, Content\}$  and random effects of  $\{id, time, SOM\}$ .

Chapter 6 also gives enhancements to Models 2 and 3 in which the effects of  $\{Content, SOM\}$  are integrated into the assessment of QoE for Web based services. The objective metric for QoE evaluation has been based on networking perspectives from WWW and metrics based on the application layers and human perception. The subjective metric is related to a state of mind (SOM) factor and the variance of each individual amongst the group. Models are developed to account these factors and other potential covariates to customer QoE during the course of our experiments.

## 7.2 Conclusions

The main objective of the thesis was both to propose and develop of a new QoE model for web based services. The final QoE model of Web based services uses a comprehensive set of metrics in order to construct a full and detailed understanding of users' experiences in a web based service environment. The analysis and validation of QoE models of Web based traffic based on both objective parameters from networking, application, and content perspectives, and subjective user state of mind have shown the efficacy of the proposed models. A comprehensive set of metrics has been evaluated on the overall customer QoE.

The thesis also proposed a framework for comprehensive assessment of QoE for Web based services which is based on a comprehensive set of metrics such as requests per second derived from application characteristics, delay derived from networking, state of mind and individual difference form human derived metric, and content-time from *CT* derived metric.

The literature review revealed the shortcomings of earlier and recently proposed QoE models. We have highlighted the limitations of traditional methods for the assessment of QoE as subjective MOS. This method is clearly cost and time consuming for obtaining user scores. We have reviewed existing ITU-T recommendations in which web browsing was involved, such as G.1030. It is clearly demonstrated that this recommendation should be updated as there is a limitation in this work that relates to current network performance and in understanding of user behaviour. We also reviewed the current status and trends for QoE in Web based services. We demonstrated that these models have certain drawbacks under current conditions for packet traffic in general and web traffic in particular, and they provide only a rough approximation of a user's QoE. We have identified shortcomings in our knowledge for establishment of QoE in Web based services and we have proceeded to identify and discuss the state-of-the-art in QoE modelling for Web based services.

The experiment has been designed by using Taguchi quality method approach which ensures that a quantitative number of observations are required. We have utilized the orthogonal arrays of  $L_{18}$ ,  $L_{25}$ , and  $L_{36}$  using the Taguchi approach to construct our experiments in order to characterize the application as well as estimate network

performance metrics in our QoE assessment model for web-based systems. We proposed further experiments with our controlling session to reduce the boredom effect that may impact on their subjective assessment. As a result, a session known as a controlling session was introduced which enables users to be interested in surfing provided web pages, to have enough time to discover the web pages and to achieve the tests with being less stressed and bored. The controlling session lasts for a maximum 120 minutes, and each session represented a maximum of three minutes. From the evaluation mentioned in Chapter 5, it is concluded that there is a general agreement between the users' perception and increased values for both objective metrics of  $\{D, RPS\}$  and we called it "the rule of nature".

However, to understand further changes in specific cases of users' perception and of un-matching cases of our "rule of nature", we proposed two further metrics that are related to content and state of mind into our customer QoE for Web browsing assessment.

The objective metric of content has been investigated showing its effects on customer QoE. This metric has been evaluated objectively by using the download time of each type of content. Specifically, in our scope of the research, we found that in some cases of (very) satisfied users, the users pay more attention and enjoy content more rather than the actual performance of networking and application. By another way, content is more important than connectivity in this case. Otherwise, content is less important than connectivity. The subjective metric of state of mind has been captured subjectively by users' feedbacks showing a random change via customer QoE of Web browsing. We found that the misinformation effect happened when users are in "stressed" mood.

The integration of  $\{SOM, Content\}$  explained our existing abnormalities with users' scores of  $\{4,5\}$  when the evaluation was based only on both networking and application layers as  $\{D, RPS\}$ . Evidently, at those levels of users' scores, we found that *Content* has affected a user's perception. In this case, users pay more attention to content rather than connectivity.

We have also investigated the applicability of a mixed effects model in predicting QoE in World Wide Web based multi-media services. An analysis is presented on objective factors such as *D*, *RPS* and *Content* and human factors like *SOM* and the

uniqueness of individuals themselves that may impact on outcomes of observations as fixed effects and random effects. The fixed effects model takes into account population mean as well as the response itself. Random effects take into group specifics and account for the correlation structure of variations amongst users. The objective metric for QoE evaluation has been based on networking perspectives from WWW and metrics based on the application layer and human perception. The subjective metric is related to state of mind factor and variance of each individual amongst the group. A model is developed to account for these factors and other potential covariates to QoE assessment during the course of our experiments. As noted earlier, there is a random effect representing an uncontrollable variation amongst groups that is required in the model. Thus, we have prepared models that take random effects into account as a necessary feature for a potential source of random effects. The models are developed to account for these factors and other potential covariates in QoE assessment during the course of the experiment.

Three models demonstrate an acceptable correlation between values of the fitted analytical model and observed user scores. The integration of psychological factors as *Content* and *SOM* (the third model) has improved the model performance in terms of information criteria where the dimension of the model, effective sample size, and log likelihood have been used to assess the results.

In conclusion, the three models enables users to feedback their opinion of their Web interactions by revealing the way in which they are interacting with the system rather than by explicitly providing their opinion via a totally subjective approach such as through a questionnaire. Three models also define and present the building of a measurement models that represent a good correlation between objectively observed parameters and subjective parameters. It is demonstrated that by integration of further human perception factors involving content and state of mind into the assessment of QoE of web based services, the overall model performance has been improved. The models demonstrate an acceptable correlation between values of the fitted analytical model and observed user scores. The in depth analysis and modelling QoE of web based services certainly helped to get a better understanding on how web surfing is experienced by users.

### 7.3 Future Work

The work presented in this thesis develops comprehensive metrics for QoE of Web browsing including subjective assessment of the user's state of mind. It is difficult to evaluate user's state of mind as there is no fool-proof way of assessing this parameter, even though it has an impact on a user's score. Thus, an objective assessment of state of mind should be considered for further work.

The effect of *Content* proposed in this thesis shows a clear impact on customer QoE. However, in this work, the *Content* is limited to the concept where different kinds of *Content* have been represented by different download times. Further context and the concepts of content have not been tested in this work. Thus, an extension of content concept will be considered for the future work.

With regard to the context of content, future work should investigate how interesting the content is for the users. For example, using web mining [139] to filter such information as general usage statistics, statistics common entry point for users, usage cluster or frequent items for defining mined knowledge source interesting as content of a web page. Once interesting levels of content are defined, those levels can be applied in our model to the overall QoE of web based services.

The outcome of our proposed model can be applied by service providers undertaking network planning. For example, by running our model, the results will explain the overall QoE in terms of networking, application, content and SOM. After running the model, service providers can provide their services taking into account specific requirements of regional services. For example, some area that is specific for quality in terms of content rather than just in terms of networking quality.

This work presented a model for QoE of Web browsing in general, it could be carried out for a particular application of web browsing such as video, YouTube, or peer-to-peer traffic in any future work.

## List of Abbreviations

AIC	Akaike Information Criterion
ASQ	American Society for Quality
BIC	Bayesian Information Criterion
BLUP	Best Linear Unbiased Predictor
CI	Confidence Intervals
Df	Degree of Freedom
EPSN	Edge Peak Signal to Noise ratio
ETSI	European Telecommunications Standards Institute
Euro-NF	European Network of Future
FTP	File Transfer Protocol
GLS	Generalized Least Squares
HTML	Hypertext Makeup Language
HTTP	Hypertext transfer Protocol
IC	Information Criteria
ICT	Information and Communication Technology
IEEE	Institute of Electrical and Electronics Engineers
IETF	Internet Engineering Task Force
IOS	Internetwork operating system
IP	Internet Protocol
ISP	Internet Service Provider
ITU	International Telecommunications Union
ML	Maximum Likelihood
MLE	Maximum Likelihood Estimation
MOS	Mean Opinion Score
NetEm	Network Emulator
PESQ	Perceptual Evaluation of Speech Quality
PSQM	Perceptual Speech Quality Measure
QoE	Quality of Experience
QoS	Quality of Service

REML	Restricted/Residual Maximum Likelihood
RFC	Request for Comments
SMTP	Simple Mail Transfer Protocol
SOM	State of Mind
SSR	Residual Sum of Square
TCP	Transmission Control Protocol
URI	Uniform Resource Identifier
URL	Uniform Resource Locator
VQEG	Video Expert Group
WWW	World Wide Web

## Bibliography

- [1] T. R. Chandrupatla, *Quality and Reliability in Engineering*. Rowan University, New Jersey, 2009.
- [2] ASQ. *American Society for Quality (ASQ): Glossary-Entry: Quality*, retrieved 2013.
- [3] A. A. Kuehn and R. L. Day, "Strategy of product quality", *Harvard business review*, 1962.
- [4] J. M. Juran, *Juran on leadership for quality*: Free Press, 2003.
- [5] J. A. Dotchin and J. S. Oakland, "Theories and concepts in total quality management", *Total Quality Management*, vol. 3, pp. 133-146, 1992.
- [6] K. Brunnström, *et al.*, "Qualinet White paper on Definitions of Quality of Experience and Related Concepts", ed Dagstuhl, 2012.
- [7] S. J. Watson, "An analysis of the concept of experience", *Journal of advanced nursing*, vol. 16, pp. 1117-1121, 2006.
- [8] C. Meyer and A. Schwager, "Understanding customer experience", *Harvard business review*, vol. 85, p. 116, 2007.
- [9] K. Kilkki, "Quality of Experience in Communications Ecosystem", *Journal of universal computer science*, vol. 14, pp. 615-624, 2008.
- [10] ITU-T, "Definition of Quality of Experience (QoE) Ref.: TD 109rev2 (PLEN/12)", ed, 2007.
- [11] ETSI, "Human Factors (HF); Quality of Experience (QoE) requirements for real-time communication services", European Telecommunications Standards Institute (ETSI), France, 2009-2012.
- [12] ITU-T, "Rec. P.800 : Methods for subjective determination of transmission quality", ed, 1996.
- [13] ITU-R, "Rec. BS.1116-1: Methods for the subjective assessment of small impairments in audio systems including multichannel sound systems", ed, 1997.
- [14] ITU-R, " Rec. BT. 710-4: Subjective assessment methods for image quality in high-definition television", ed, 2000.
- [15] ITU-R, "Rec. BS. 1534-1: Method for the subjective assessment of intermediate quality levels of coding systems", ed, 2003.
- [16] ITU-R, "Rec. BT. 500-11: Quality of Experience Requirements for IPTV Services", ed, 1974-2002.
- [17] ITU-T, "Rec. P. 830: Subjective performance assessment of Telephone-band and wide band digital codecs", ed, 1996.
- [18] ITU-T, "Rec. P.910: Subjective video quality assessment methods for multimedia applications", ed, 1999.
- [19] ITU-T, "Rec. P.911: Subjective audiovisual quality assessment methods for multimedia applications", ed, 1998.
- [20] ITU-T, " Rec. P.805: Subjective evaluation of conversational quality", ed, 2007.
- [21] ITU-T, "G107: The E-Model, A Computational Model For Use in Transmission Planning", ed, 2003.
- [22] ITU-T, "Rec. P861: Objective quality measurement of telephone-band (300-3400 Hz) speech codecs", ed, 1998.
- [23] ITU-T, "Rec. P862: Perceptual evaluation of speech quality (PESQ): An objective method for end-to-end speech quality assessment of narrow-band telephone networks and speech codecs", ed, 2001.
- [24] M. Fiedler, T. Hossfeld and T.-G. Phuoc, "A generic quantitative relationship between Quality of Experience and Quality of Service", *Network, IEEE*, vol. 24, pp. 36-41, 2010.
- [25] Euro-NF. Euro Network of Excellence: The European Community for research, technological development and demonstration activities. Available: [http://euronf.enst.fr/en\\_accueil.html](http://euronf.enst.fr/en_accueil.html) (last accessed 12/2012) [Online].
- [26] T. Hobfeld, R. Schatz and S. Egger, "SOS: The MOS is not enough!", in *Quality of Multimedia Experience (QoMEX), 2011 Third International Workshop on*, 2011, IEEE, pp. 131-136.
- [27] M. Fiedler, T. Hossfeld and P. Tran-Gia, "A generic quantitative relationship between Quality of Experience and Quality of Service", *Network, IEEE*, vol. 24, pp. 36-41, 2010.
- [28] E. Ibarrola, F. Liberal, I. Taboada and R. Ortega, "Web QoE Evaluation in Multi-agent Networks: Validation of ITU-T G.1030", in *The Fifth International Conference on Autonomic and Autonomous Systems, 2009. ICAS '09.* , 2009, pp. 289-294.
- [29] ITU-T, "Rec. G1030: Estimating end to end performance in IP networks for data application", ed, 2005.
- [30] H. Schulzrinne, "25 years of quality of service research-where next", in *Keynote at the 18th IEEE International Workshop on Quality of Service (IEEE IWQoS 2010)*, ed. Beijing, China, 2010.
- [31] ITU-T. (2013-2016). *SG 12- Performance, QoS and QoE*. Available: <http://www.itu.int/en/ITU-T/studygroups/2013-2016/12/Pages/default.aspx> (last accessed 01/2013).
- [32] T. Hoßfeld, *et al.*, "The memory effect and its implications on web QoE modeling", in *Proceedings of the 23rd International Teletraffic Congress*, 2011, ITCP, pp. 103-110.
- [33] Microsoft. Download Center: "Windows Internet Explorer 8 for Windows XP". Available: <http://www.microsoft.com/en-nz/download/details.aspx?id=43> (last accessed 12/2012) [Online].
- [34] StatCounter. (2012). *Top 5 Browsers on December 2012*, StatCounter Global Stats. Available: <http://gs.statcounter.com/> (last accessed 12/2012).
- [35] F. Pukelsheim, *Optimal design of experiments*, vol. 50: Society for Industrial Mathematics, 2006.
- [36] T. Berners-Lee, "WWW: Past, present, and future", *Computer*, vol. 29, pp. 69-77, 1996.
- [37] E. R. Babbie and L. Benaquisto, *Fundamentals of social research*: Cengage Learning, 2009.

- [38] G. Albaum, "The Likert scale revisited", *Journal-Market research society* vol. 39, pp. 331-348, 1997.
- [39] Oracle, "Best Practices for Improving Survey Participation ", March 2012.
- [40] ITU-T, "Rec. P. 835: Subjective test methodology for evaluating speech communication systems that include noise suppression algorithm", ed, 2003.
- [41] ITU-T, "Rec. P.880: Continuous evaluation of time-varying speech quality", ed, 2005.
- [42] ITU-R, "Rec. BS.562-3: Subjective assessment of sound quality", ed, 1990.
- [43] ITU-T, " Rec. P.563: Single Ended Method for Objective Speech Quality Assessment in Narrow-Band Telephony Applications", ed, 2004.
- [44] ITU-R, "Rec. BS.1387-1: Method for Objective Measurements of Perceived Audio Quality (PEAQ)", ed, 2001.
- [45] ITU-T, "Rec. J. 246: Perceptual visual quality measurement techniques for multimedia services over digital cable television networks in the presence of a reduced bandwidth reference", ed. Geneve, Switzerland, 2008.
- [46] ITU-T, "Rec. J. 247: Objective perceptual multimedia video quality measurement in the presence of a full reference", ed. Geneva, Switzerland, 2008.
- [47] K. Brunnstrom, D. Hands, F. Speranza and A. Webster, "VQEG validation and ITU standardization of objective perceptual video quality metrics", *Signal Processing Magazine, IEEE*, vol. 26, pp. 96-101, 2009.
- [48] A. M. Rohaly, *et al.*, "Video quality experts group: Current results and future directions", in *Proceedings-Spie the International Society for Optical Engineering*, 2000, International Society for Optical Engineering; 1999, pp. 742-753.
- [49] K. Hyun Jong, *et al.*, "The QoE Evaluation Method through the QoS-QoE Correlation Model", in *The Fourth International Conference on Networked Computing and Advanced Information Management*, 2008, NCM '08, pp. 719-725.
- [50] S. A. Moebs, "A learner, is a learner, is a user, is a customer: QoS-based experience-aware adaptation", in *The 16th ACM international conference on Multimedia*, Vancouver, Canada, 2008.
- [51] S. Khirman and P. Henriksen, "Relationship between Quality-of-Service and Quality-of-Experience for Public Internet Service", in *The 3rd workshop on Passive and Active Measurement*, Fort Collins, Colorado, USA, 2002.
- [52] P. Brooks and B. Hestnes, "User measures of quality of experience: why being objective and quantitative is important", *Network, IEEE*, vol. 24, pp. 8-13, 2010.
- [53] ITU-T, "Rec. Y.1541 Network performance objectives for IP-based services", ed, 2002.
- [54] ITU-T, "Rec. G1010: End-user multimedia QoS categories", ed, 2001.
- [55] L.-y. Liu, W.-a. Zhou and J.-d. Song, "The Research of Quality of Experience Evaluation Method in Pervasive Computing Environment", in *The 1st International Symposium on Pervasive Computing and Applications*, 2006, pp. 178-182.
- [56] L. Janowski and Z. Papir, "Modeling subjective tests of quality of experience with a generalized linear model", in *The International Workshop on Quality of Multimedia Experience, QoMEX 2009* San Diego, California, U.S.A, 2009, IEEE, pp. 35-40.
- [57] S. Khirman and P. Henriksen, "Relationship between quality-of-service and quality-of-experience for public internet service", 2002.
- [58] I. Pais and M. Almeida, "End user behavior and performance feedback for service analysis", in *The 13th International Conference on Intelligence in Next Generation Networks*, Bordeaux, France, 2009, IEEE, pp. 1-6.
- [59] I. Din, N. A. Saqib and A. Baig, "Passive Analysis of Web Traffic Characteristics for Estimating Quality of Experience", in *Globecom Workshops, 2008 IEEE*, 2008, pp. 1-5.
- [60] ITU-T, "Rec. G1040: Network contribution to transaction time", ed, 2006.
- [61] ITU-T, "Rec. Y.1540 Internet protocol data communication service –IP packet transfer and availability performance parameters", ed, 2011.
- [62] Cisco, "Cisco Visual Networking Index: Forecast and Methodology 2010-2016", 30 May 2012.
- [63] D. Moltchanov, " Problems arising in evaluating perceived quality of media applications in packet networks", in *The 18th ITC Specialist Seminar on Quality of Experience* Berlin, Germany, 2008.
- [64] D. C. Rubin, S. Hinton and A. Wenzel, "The precise time course of retention", *Journal of Experimental Psychology: Learning, Memory, and Cognition; Journal of Experimental Psychology: Learning, Memory, and Cognition*, vol. 25, p. 1161, 1999.
- [65] N. T. Siebel and S. Maybank, "The advisor visual surveillance system", in *ECCV 2004 workshop Applications of Computer Vision (ACV)*, 2004, Citeseer.
- [66] X. Tao and G. Shaogang, "Video Behavior Profiling for Anomaly Detection", *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 30, pp. 893-908, 2008.
- [67] G. Taguchi and Y. Wu, *Introduction to off-line quality control*: Central Japan Quality Control Assoc., 1979.
- [68] W. Welch, T. Yu, S. M. Kang and J. Sacks, "Computer experiments for quality control by parameter design", *Journal of Quality Technology*, vol. 22, pp. 15-22, 1990.
- [69] T. KWOK-LEUNG, "An overview of Taguchi method and newly developed statistical methods for robust design", *Iie Transactions*, vol. 24, pp. 44-57, 1992.
- [70] Y. Wu and A. Wu, *Taguchi methods for robust design*: ASME press, 2000.

- [71] G.Taguchi, S.Chowdhury and Y.Wu, *Taguchi's quality engineering handbook*: John Wiley & Sons, 2005.
- [72] R. K. Roy, *A primer on the Taguchi method*: Sme, 2010.
- [73] R. Mandl, "Orthogonal Latin squares: an application of experiment design to compiler testing", *Communications of the ACM*, vol. 28, pp. 1054-1058, 1985.
- [74] L.T.Nguyen, R. Harris and J.Jusak, "A Pilot Study to Access Quality of Experience Based on Varying Network Parameters and User Behaviour", in *2011 International Conference on Telecommunication Technology and Applications*, Sydney, Australia, 2011, IACSIT Press, Singapore, pp. 30-34.
- [75] L. T. Nguyen, R. Harris, J. Jusak and A. Punchihewa, "Modelling of Quality of Experience for Web Traffic", in *2010 Second International Conference on Network Applications Protocols and Services (NETAPPS)*, 2010, IEEE, pp. 84-89.
- [76] Ubuntu. (Retrieved 2013). *Ubuntu Server 12.04.1 LTS*. Available: <http://www.ubuntu.com/download/server>
- [77] Cisco. (Retrieved 2013). *Catalyst 1900 Series Installation and Configuration Guide*. Available: <http://www.cisco.com/en/US/products/>
- [78] Cisco. (Retrieved 2013). *Cisco 1710 Security Access Router*. Available: <http://www.cisco.com/en/US/products>
- [79] Cisco, *Cisco IOS Configuration Fundamentals Command*. San Jose, USA: Cisco Systems, Inc., 2007.
- [80] C. Semeria, "Understanding IP addressing: Everything you ever wanted to know", *NSD Marketing, 3Com Corporation*, 1996.
- [81] S. Hemminger, "Network emulation with NetEm", 2005, Citeseer, pp. 18-23.
- [82] S. Hemminger, "Netem-emulating real networks in the lab", in *Proceedings of the 2005 Linux Conference Australia, Canberra, Australia*, 2005.
- [83] R. Love, *Linux kernel development*: Addison-Wesley Professional, 2010.
- [84] Y. Benita, "Kernel korner: analysis of the HTB queueing discipline", *Linux Journal*, vol. 2005, p. 13, 2005.
- [85] M. Devera and D. Cohen, "HTB Linux queueing discipline manual", ed: May, 2002.
- [86] E. Gelenbe, G. Pujolle and J. Nelson, *Introduction to queueing networks*: Wiley Chichester, 1998.
- [87] OpenSource, "Mod\_qos", *Sourceforge.net*, vol. version 10, February 2012.
- [88] Apache, "Apache HTTP server. Available: <http://httpd.apache.org/docs/2.2/> ", vol. 2.2, Retrieved 2012.
- [89] R. Barnett, "Apache Web Server 2.2. 0".
- [90] GNU. (Retrieved 2007). *GNU General Public License*. Available: [www.gnu.org/licenses/gpl.html](http://www.gnu.org/licenses/gpl.html)
- [91] R. Stallman, "The GNU operating system and the free software movement", 1999.
- [92] Ubuntu. (Retrieved 2012). *Package: libapache2-mod-qos(9.57-1)[universe]*. Available: <https://launchpad.net/ubuntu/+source/libapache2-mod-qos/9.57-1>
- [93] D. Callan, "Content is king", *AKA Marketing.com*. Available: [www.akamarketing.com/content-is-king.html](http://www.akamarketing.com/content-is-king.html), 2006.
- [94] P. Morville and L. Rosenfeld, *Information Architecture for the world wide web: designing large-scale web sites*: O'Reilly Media, Incorporated, 2006.
- [95] I. S. Graham, *The HTML sourcebook*: John Wiley & Sons, Inc., 1995.
- [96] D. Raggett, A. Le Hors and I. Jacobs, "HTML 4.01 Specification", *W3C recommendation*, vol. 24, 1999.
- [97] M. Kofler and B. Öggl, *PHP 5.3& MySQL 5.1: Grundlagen, Programmier Techniken, Beispiele*: Pearson Education, 2008.
- [98] R. Lerdorf, K. Tatroe and P. MacIntyre, *Programming Php*: O'Reilly Media, Incorporated, 2006.
- [99] F. Ricca and P. Tonella, "Analysis and testing of web applications", in *Software Engineering, 2001. ICSE 2001. Proceedings of the 23rd International Conference on*, 2001, IEEE, pp. 25-34.
- [100] A. MySQL, "MySQL database server", *Internet WWW page*. Available: <http://www.mysql.com> (last accessed 9/2011), 2004.
- [101] M. Viswanathan, "Measuring speech quality for text-to-speech systems: development and assessment of a modified mean opinion score (MOS) scale", *Computer Speech & Language*, vol. 19, pp. 55-83, 2005.
- [102] I. Pais and M. Almeida, "End user behavior and performance feedback for service analysis", in *The 13th International Conference on Intelligence in Next Generation Networks (ICIN 2009)* 2009, IEEE, pp. 1-6.
- [103] T. Ciszowski, *et al.*, "Towards Quality of Experience-based reputation models for future web service provisioning", *Telecommunication Systems*, pp. 1-13, 2009.
- [104] M. T. Rose, *The open book: a practical perspective on OSI*: Prentice-Hall, Inc., 1990.
- [105] R. Fielding, *et al.*, "RFC 2616", *Hypertext Transfer Protocol-HTTP/1.1*, vol. 2, p. 2.2, 1999.
- [106] P. Barford and M. Crovella, "Generating representative web workloads for network and server performance evaluation", *ACM SIGMETRICS Performance Evaluation Review*, vol. 26, pp. 151-160, 1998.
- [107] V. Cardellini, E. Casalicchio, M. Colajanni and M. Mambelli, "Enhancing a web-server cluster with quality of service mechanisms", in *21st IEEE International Performance, Computing, and Communications Conference, 2002.* , 2002, IEEE, pp. 205-212.
- [108] A. Orebaugh, G. Ramirez and J. Beale, *Wireshark & Ethereal network protocol analyzer toolkit*: Syngress, 2006.

- [109] L. T. Nguyen, R. Harris and J. Jusak, "Analysis of networking and Application Layer derived metrics for web Quality of Experience", in *Consumer Communications and Networking Conference (CCNC), 2012 IEEE*, 2012, pp. 321-325.
- [110] C. D. Fisherl, "Boredom at work: A neglected concept", *Human Relations*, vol. 46, pp. 395-417, 1993.
- [111] G. Ryle, *The concept of mind*: University of Chicago Press, 1949.
- [112] P. M. Spacks, *Boredom: The literary history of a state of mind*: University of Chicago Press, 1995.
- [113] L. B. Wexner, "The degree to which colors (hues) are associated with mood-tones", *Journal of applied psychology*, vol. 38, p. 432, 1954.
- [114] J. M. Cruz-hernandez, "Method and Appratus for Generating Mood-based Haptic feedback", ed: EP Patent 2,350,776, 2011.
- [115] I. A. Essa and A. P. Pentland, "Coding, analysis, interpretation, and recognition of facial expressions", *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 19, pp. 757-763, 1997.
- [116] E. F. Loftus, "Planting misinformation in the human mind: A 30-year investigation of the malleability of memory", *Learning & Memory*, vol. 12, pp. 361-366, 2005.
- [117] A. Odlyzko, "Content is Not King", *First Monday*, vol. 6, No. 2, 5 February 2001.
- [118] N. M. Laird and J. H. Ware, "Random-effects models for longitudinal data", *Biometrics*, pp. 963-974, 1982.
- [119] R. H. Baayen, D. J. Davidson and D. M. Bates, "Mixed-effects modeling with crossed random effects for subjects and items", *Journal of memory and language*, vol. 59, pp. 390-412, 2008.
- [120] C. Van Bael, R. Baayen and H. Strik, "Segment deletion in spontaneous speech: A corpus study using mixed effects models with crossed random effects", in *Proceedings of interspeech*, 2007, pp. 2741-2744.
- [121] J. Fox, "Linear mixed models", *Appendix to An R and S-PLUS Companion to Applied Regression*, URL <http://cran.r-project.org/doc/contrib/Fox-Companion/appendix-mixed-models.pdf>, 2002.
- [122] M. J. Lindstrom and D. Bates, "Nonlinear Mixed Effects Models", *Biometrics*, vol. 46, pp. 673-687, 1990.
- [123] R. D. Bock and M. Aitkin, "Marginal maximum likelihood estimation of item parameters: Application of an EM algorithm", *Psychometrika*, vol. 46, pp. 443-459, 1981.
- [124] A. P. Verbyla, "Modelling Variance Heterogeneity: Residual Maximum Likelihood and Diagnostics", *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 55, pp. 493-508, 1993.
- [125] H. Akaike, "A new look at the statistical model identification", *Automatic Control, IEEE Transactions on*, vol. 19, pp. 716-723, 1974.
- [126] M. W. Browne, "Generalized least squares estimators in the analysis of covariance structures", 1973.
- [127] D. M. Bates, "lme4: Mixed-effects modeling with R", Available: <http://lme4.r-forge.r-project.org/book> (last accessed 12/2012), 2010.
- [128] L. T. Nguyen, R. Harris, A. Punchihewa and J. Jusak, "Application of a Mixed Effects Model in Predicting Quality of Experience in World Wide Web Services", in *2012 Fourth International Conference on Computational Intelligence, Modelling and Simulation (CIMSIM)*, 2012, IEEE, pp. 316-321.
- [129] F. Mosteller, "Samuel S. Wilks: Statesman of Statistics", *American Statistician*, vol. 18, No. 2, pp. 11-17, April, 1964.
- [130] J. W. Harris and H. Stocker, *Handbook of Mathematics and Computational Science*. New York: Springer-Verlag, 1998.
- [131] T. Snijders and R. Bosker, "Multilevel modeling: An introduction to basic and advanced multilevel modeling", 1999.
- [132] NIST/SEMATECH. (2012). *e-Handbook of Statistical Methods*. Available: <http://www.itl.nist.gov/div898/handbook/> (last accessed 12/2012)
- [133] J. L. Rodgers and W. A. Nicewander, "Thirteen ways to look at the correlation coefficient", *The American Statistician*, vol. 42, pp. 59-66, 1988.
- [134] C. Spearman, "The proof and measurement of association between two things", *The American journal of psychology*, vol. 15, pp. 72-101, 1904.
- [135] J. L. Myers and A. D. Well, *Research design and statistical analysis*: Lawrence Erlbaum, 2002.
- [136] M. Mittlböck and T. Waldhör, "Adjustments for R<sup>2</sup>-measures for Poisson regression models", *Computational statistics & data analysis*, vol. 34, pp. 461-472, 2000.
- [137] J. M. Bernardo and A. F. M. Smith, "Bayesian theory", *Measurement Science and Technology*, vol. 12, p. 221, 2001.
- [138] A. Chakrabarti and J. K. Ghosh, "AIC, BIC and Recent Advances in Model Selection", in *Philosophy of Statistics*, M. G. Dov, et al., Eds., ed Amsterdam: North-Holland, 2011, pp. 583-605.
- [139] R. Cooley, "The use of web structure and content to identify subjectively interesting web usage patterns", *ACM Transactions on Internet Technology (TOIT)*, vol. 3, pp. 93-116, 2003.
- [140] P. Teetor, *R cookbook*: O'Reilly Media, Incorporated, 2011.
- [141] R. Stallman, "The GNU project", ed, 1998.
- [142] R. Stallman, "Gnu general public license", *Free Software Foundation, Inc., Tech. Rep*, 1991.
- [143] D. Bates and M. Maechler, "lme4: Linear mixed-effects models using S4 classes. R package version 0.999375-35.2010", Available: <http://CRAN.R-project.org/package=lme4> (last accessed 12/2012), 2010.
- [144] D. Bates and M. Maechler, "matrix: A Matrix Package for R", *R package version 0.99875-2*. Available: <http://CRAN.R-project.org> (last accessed 9/2012), vol. 15, 2007.
- [145] D. Sarkar, *Lattice: multivariate data visualization with R*: Springer, 2008.

- [146] P. Dalgaard, *Introductory statistics with R*: Springer, 2008.
- [147] A. Gelman and J. Hill, "Data Analysis Using Regression and Multilevel/Hierarchical Models", *Cambridge University Press*, 2006.
- [148] A. Gelman and Y.-S. Su, "arm: Extract Standard Errors of Model Coefficients", Available: <http://CRAN.R-project.org/package=arm> (last accessed 3/2013), 2013.

## Appendices

### A. Publications

- (1) An Enhanced Analytical Framework Integrating aspects of Network and Application performance, Psychology, and Content in QoE Assessment for Web browsing, Le Thu Nguyen,, Richard Harris, Amal Punchihewa, in International Journal of Computer and Information Technology (ISSN: 2279-0764), Volume 02, Issue 05, September 2013.
- (2) L. T. Nguyen, R.J. Harris, and A. Punchihewa, "Assessment of Quality of Experience for Web Browsing As Function of Quality of Service and Content", in The fifth International Conference on Ubiquitous and Future Networks 2013-Invite (ICUFN 2013-Invite), 2013, IEEE.
- (3) L. T. Nguyen, R. Harris, A. Punchihewa and J. Jusak, "Application of a Mixed Effects Model in Predicting Quality of Experience in World Wide Web Services", in 2012 Fourth International Conference on Computational Intelligence, Modelling and Simulation (CIMSIM), 2012, IEEE, pp. 316-321.
- (4) L.T. Nguyen, R. Harris and J. Jusak, "Analysis of networking and Application Layer derived metrics for web Quality of Experience", in IEEE Consumer Communications and Networking Conference (CCNC), 2012 IEEE, pp. 321-325.
- (5) L.T. Nguyen, R. Harris and J. Jusak, "A Pilot Study to Access Quality of Experience Based on Varying Network Parameters and User Behaviour", in 2011 International Conference on Telecommunication Technology and Applications, Sydney, Australia, 2011, IACSIT Press, Singapore, pp. 30-34.
- (6) L. T. Nguyen, R. Harris, J. Jusak and A. Punchihewa, "Modelling of Quality of Experience for Web Traffic", in 2010 Second International Conference on Network Applications Protocols and Services (NETAPPS), 2010, IEEE, pp. 84-89.
- (7) L. T. Nguyen, R. Harris, and J. Jusak, A poster paper of "Modelling of Quality of Experience for Web Traffic", in 2010 New Zealand Computer Science Research Student Conference (NZCSRSC), 2010, Wellington.

- (8) L. T. Nguyen, R. J. Harris and J. Jusak, A poster paper "Inferring User Behaviour by Observing Packet Traffic", in New Zealand Post Graduate Conference (NZPGC 2009), Wellington.

## B. Model Specification

Model		Fixed				Random	
		Objective factor					Subjective factor
		Delay	Requests per second	Content	Time	User identity	State of Mind
Basic model	Model 1	✓	✓		✓	✓	
Enhanced model	Model 2	✓	✓	✓	✓	✓	
	Model 3	✓	✓	✓	✓	✓	✓

C. Orthogonal Array

Table A. 1:  $L_{18}(2^1 \times 3^7)$

Number of Run	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	
1 two-level factor	1	1	1	1	1	1	1	1	1	2	2	2	2	2	2	2	2	2	
	2	1	1	1	2	2	2	3	3	3	1	1	1	2	2	2	3	3	3
7 three-level factors	3	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3
	4	1	2	3	1	2	3	2	3	1	3	1	2	2	3	1	3	1	2
	5	1	2	3	2	3	1	1	2	3	3	1	2	3	1	2	2	3	1
	6	1	2	3	2	3	1	3	1	2	2	3	1	1	2	3	3	1	2
	7	1	2	3	3	1	2	2	3	1	2	3	1	3	1	2	1	2	3
	8	1	2	3	3	1	2	3	1	2	1	2	3	2	3	1	2	3	1
	8	1	2	3	3	1	2	3	1	2	1	2	3	2	3	1	2	3	1

Table A. 2:  $L_{25}$  (2 five-level factors)

No	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
A	2	2	4	4	5	5	5	1	5	3	3	5	1	1	1	2	2	1	4	4	3	2	4	3	3
B	1	4	2	5	5	4	3	4	1	5	4	2	1	2	5	3	5	3	4	3	3	2	1	1	2

Table A. 3:  $L_{25}(5^2 \times 2^1)$  (2 five-level factors and 1 two-level factor)

No	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
A	2	2	2	2	4	4	4	4	5	5	5	5	5	5	1	1	5	5	3	3	3	3	5	5	1
B	1	1	4	4	2	2	5	5	5	5	4	4	3	3	4	4	1	1	5	5	4	4	2	2	1
C	3	4	3	4	3	4	3	4	3	4	3	4	3	4	3	4	3	4	3	4	3	4	3	4	3
No	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50
A	1	1	1	1	1	2	2	2	2	1	1	4	4	4	4	3	3	2	2	4	4	3	3	3	3
B	1	2	2	5	5	3	3	5	5	3	3	4	4	3	3	3	3	2	2	1	1	1	1	2	2
C	4	3	4	3	4	3	4	3	4	3	4	3	4	3	4	3	4	3	4	3	4	3	4	3	4

Table A. 4:  $L_{36} (3^2 \times 4^1)$  (2 three-level factors and 1 four-level factor)

No	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36
A	1	1	1	1	1	1	1	1	5	5	5	5	5	5	5	5	3	3	3	3	1	1	1	1	5	5	5	5	3	3	3	3	3	3	3	3
B	1	1	1	1	5	5	5	5	1	1	1	1	5	5	5	5	4	4	4	4	4	4	4	4	4	4	4	4	1	1	1	1	5	5	5	5
C	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4

#### D. Restricted/ Residual Maximum likelihood (REML)

The problem of Maximum Likelihood (ML) estimator is to estimate intercept and slope  $\beta$  in the equation as below:

$$Y \sim N(X\beta, \sigma^2) \quad (3.1)$$

If there is no  $\beta$  in Equation(3.1), the problem is solved. Thus, REML is applied to avoid having any  $\beta$  in Equation(3.1) by finding special matrix  $K$  with dimension of  $n \times (n-2)$ . Special mean of orthogonal of  $K'X = 0$

Thus,  $K'Y = K'X\beta + K'\varepsilon = 0 + K'\varepsilon$

Then, the distribution for  $K'Y = N(0, \sigma^2 K'K)$  (3.2)

Equation(3.2) is no longer depended on  $\beta$ , then applying ML on  $K'Y$  give unbiased estimator for  $\sigma^2$ .

Verbyla in [124] gave an interpretation for REML method which partition likelihood into two independent parts:

$Y_1 = K_1'Y$  is related to fixed effects, and  $Y_2 = K_2'Y$  is related to residual (Zero expectation).  $K_1$  is  $n \times p$  matrix of full column rank.  $K_2$  is  $n \times [n-p]$  matrix of full column rank.  $K_1'X = I$  and  $K_2'X = 0$

In conclusion, REML is maximum likelihood estimation (ML) but replacing:

- $Y$  with  $KY$
- $X$  with  $0$
- $Z$  with  $K'Z$
- $V$  with  $K'VK$  where  $K'X = 0$

## E. R language and codes

In this research, R[140] is used. R is GNU project [141] and freely available under GNU license[90, 91, 141, 142]. Its source code is primarily written in C, Fortran and R. There are some required packages served for model development and testing such as lme4[127, 143], Matrix[144], lattice [145], stats[146], and arm[147, 148].

Packages required:

```
Loading required package: lme4
Loading required package: Matrix
Loading required package: lattice
Loading required package: stats
Loading required package: arm
```

Model development:

```
> model=function(type)
+   { require(lme4)
+     if(type==1){
+       model1<-
lmer(MOS~as.factor(D)+as.factor(RPS)+(as.factor(D)+as.factor(RPS)|as.factor(id):as.factor(time)), REML=TRUE, data=mixdatatext)
+       return(model1)
+     }else{
+       if(type==2) {
+         model2<-
lmer(MOS~as.factor(D)+as.factor(RPS)+as.factor(Content)+(as.factor(D)+as.factor(RPS)|as.factor(id):as.factor(time)), REML=TRUE, data=mixdatatext)
+         return(model2)} else
+         { model3<-
lmer(MOS~as.factor(D)+as.factor(RPS)+as.factor(Content)+(as.factor(D)+as.factor(RPS)|as.factor(id):as.factor(time))+(1|as.factor(SOM)), REML=TRUE, data=mixdatatext)
+         return(model3)}
+       }
+     }
> model1<-model(1)
> model2<-model(2)
> model3<-model(3)
```

Estimation model parameters: random effects vector ( $b$ ), covariance matrix of random ( $D$ ), matrix of random effects ( $Z$ ), and matrix of fixed effects ( $X$ ).

Random effects vector estimation ( $b$ ):

```
bEstimate=function(model)
+   {
+     b1<-as.matrix(ranef(model)$"as.factor(id):as.factor(time)")
+     b2<-as.matrix(ranef(model)$"as.factor(SOM)")
+     print(b1)
+     print(b2)
+     return()
+   }
bEstimate(model3)
```

Covariance matrix of random ( $D$ ):

```
> Dmatrix=function(model)
+   {
+     D<-diag(VarCorr(model)$"as.factor(id):as.factor(time)")
+     return(D)
+   }
> Dmatrix(model1)
> Dmatrix(model3)
```

Model matrix of random effect ( $Z$ ):

```
> Zmatrix=function(model)
+   {
+     Z.spare<-t(model@Zt)
```

```

+       Z<-as.matrix(Z.spare)
+       return(Z)
+     }
> Zmatrix(model1)
> Zmatrix(model3)

```

Model matrix of fixed effect( $X$ ):

```

> Xmatrix=function(model)
+ {
+   X<-model.matrix(model)
+   return(X)
+ }
> Xmatrix(model1)
> Xmatrix(model3)

```

Conditional modes of random effects plot:

```

> conditionplot=function(model){
+   model.ranef<-ranef(model, postVar=TRUE)
+   return(qqmath(model.ranef))
+ }
> conditionplot(model1)
> conditionplot(model2)
> conditionplot(model3)

```

Residual plots:

```

> residualplot=function(model){
+   qqnorm(residuals(model), main="Residual Plot ")
+   qqline(residuals(model))
+   return()
+ }
> residualplot(model1)
> residualplot(model2)
> residualplot(model3)

```

Calculation  $t$ ,  $p$  value as function:

```

> t_pvalue=function(mode)
+ {
+   tvalue<-fixef(model)/sqrt(diag(vcov(model)))
+   pvalue<-2*(1-pnorm(abs(tvalue)))
+   return(cbind(round(tvalue,10), round(pvalue,10)))
+ }
> t_pvalue(model1)
> t_pvalue(model3)

```

Testing Restricted/Residual log likelihood of model 1, model 2 and model 3 following REML (logLik):

```

> loglik.model=function(model){
+   loglik.model<-logLik(model, REML=TRUE)
+   return(loglik.model)
+ }
> loglik.model(model1)
> loglik.model(model2)
> loglik.model(model3)

```

Confidence interval calculation of fixed effect as function of CIFixed:

```

> CIFixed=function(model)
+ {
+   vc<-vcov(model,useScale=FALSE)
+   Estimate<-fixef(model)
+   StandardError<-se.fixef(model)
+   tvalue<-Estimate/sqrt(diag(vc))
+   upper<- Estimate+1.96*StandardError
+   lower<-Estimate-1.96*StandardError
+   return(cbind(lower, upper))
+ }
> CIFixed(model1)
> CIFixed(model3)

```

Testing  $R$  and  $R^2$  with method of Spearman:

```

> R_Rsquare=function(model){
+   R<-cor(fitted(model), MOS, method="spearman")
+   Rsquare<-cor(fitted(model), MOS, method="spearman")^2

```

```
+     return(cbind(R, Rsquare))
+   }
> R_Rsquare(mode11)
> R_Rsquare(mode12)
> R_Rsquare(mode13)
```

## F. Web design and codes

Codes for handle form to record the users' score and other information as "handleform2.php" in the experiments:

```
<?php //predefine variables//
if (!isset($_SERVER))
{
    $_GET    = &$HTTP_GET_VARS;
    $_POST   = &$HTTP_POST_VARS;
    $_ENV    = &$HTTP_ENV_VARS;
    $_SERVER = &$HTTP_SERVER_VARS;
    $_COOKIE = &$HTTP_COOKIE_VARS;
    $_REQUEST = array_merge($_GET, $_POST, $_COOKIE);
}

$Comments_1=$_POST['Comments_1'];
$NAME=$_POST['NAME'];
$EMAIL=$_POST['EMAIL'];
$COMMENT=$_POST["COMMENT"];
$Opinion_1=$_POST["Opinion_1"];
$Opinion_12=$_POST["Opinion_12"];

$userip=(($_SERVER['X_FORWARDED_FOR'])?$_SERVER['X_FORWARDED_FOR']:$_SERVER['REMOTE_ADDR']);
$time = date("F j, Y, g:i:s a");
$ip=GetHostByName($_SERVER['REMOTE_ADDR']);

//connect to database
$dbhost='localhost';
$dbuser='root';
$dbpass='culkin79';
$dbname="result";
$connx = mysql_connect("$dbhost","$dbuser","$dbpass");
if(!$connx)
{
    die("DB connection failed: " . mysql_error());
}

mysql_select_db($dbname);

//insert result to database

$insert_query1 = sprintf("INSERT INTO result_table(ipaddress, time, userfeedback, stateofmind) ." VALUES ('%s','%s','%s','%s')",$ip,$time, $Opinion_1, $Opinion_12);

$insert_query2 = sprintf("INSERT INTO comment(username, email, comment)" ." VALUES ('%s','%s','%s')",$NAME,$EMAIL, $COMMENT);

mysql_query($insert_query1) or die ('error updating database 1');
mysql_query($insert_query2) or die ('error updating database 2');
echo "Your feedbacks/comments have been saved!!!";

mysql_close($connx);

?>
```

Basic default format of web design for a web page “index.html” in the experiments:

```
<html lang="en">
<head>
<meta http-equiv="Content-Type" content="text/html; charset=UTF-8">
<meta name="Generator" content="Serif WebPlus 10.1.3.039">
<title>Massey</title>
<style type="text/css">
<!--
body {margin: 0px; padding: 0px;}
.Website-Body-Text-P
{
margin:0.0px 0.0px 0.0px 0.0px; text-align:center; font-weight:400;
}
.Normal-P
{
margin:0.0px 0.0px 0.0px 0.0px; text-align:left; font-weight:400;
}
.Website-Body-Text-C
{
font-family:"Verdana", sans-serif; font-size:11.0px;
line-height:1.18em; color:#ffffff;
}
.Normal-C
{
font-family:"Times New Roman", serif; font-weight:700;
font-size:24.0px; line-height:1.13em; color:#000080;
}
.Normal-C0
{
font-family:"Times New Roman", serif; font-size:16.0px;
line-height:1.25em; color:#000080;
}
-->
</style>
<script src="_wp_scripts/jsFlashVer.js" type="text/javascript"></script>
<script type="text/javascript"><!--
var blankSrc = "_wp_scripts/blank.gif";
--></script>
<script src="_wp_scripts/jspngfix.js" type="text/javascript"></script>
<script src="_wp_scripts/jsRollover.js" type="text/javascript">
</script>
<script type="text/javascript">
<!--
PPImgInit('tg_71', '_wp_generated/wpa26489f2.png', '_wp_generated/wpdebd4186.png', '', ''
,0,0);
PPImgInit('tg_72', '_wp_generated/wp2016bc80.png', '_wp_generated/wp5479d30.png', '', ''
,0,0);
PPImgInit('tg_73', '_wp_generated/wp40bf68a9.png', '_wp_generated/wp9c54a2c7.png', '', ''
,0,0);
PPImgInit('tg_74', '_wp_generated/wp6fda0b16.png', '_wp_generated/wp6b0011b7.png', '', ''
,0,0);
PPImgInit('tg_75', '_wp_generated/wp5f7cae97.png', '_wp_generated/wpd22c1426.png', '', ''
,0,0);
//-->
</script>
</head>

<body link="#000080" vlink="#ff8000" alink="#ff8000" text="#000000"
style="background: #ffffff; height:800px;">
<center><div style="position:relative;width:800px;">
<div style="position:absolute; left:47px; top:80px; width:132px; height:34px;">
<a href="index.html" onMouseOut="PPImgAction('out','tg_71')
onMouseOver="PPImgAction('over','tg_71')"></a></div>
<div style="position:absolute; left:179px; top:80px; width:132px; height:34px;">
<a href="research.html" onMouseOut="PPImgAction('out','tg_72')
onMouseOver="PPImgAction('over','tg_72')"></a></div>
<div style="position:absolute; left:311px; top:80px; width:132px; height:34px;">
<a href="studentlife.html" onMouseOut="PPImgAction('out','tg_73')
onMouseOver="PPImgAction('over','tg_73')"></a></div>
<div style="position:absolute; left:443px; top:80px; width:132px; height:34px;">
<a href="admission.html" onMouseOut="PPImgAction('out','tg_74')
onMouseOver="PPImgAction('over','tg_74')"></a></div>
<div style="position:absolute; left:575px; top:80px; width:132px; height:34px;">
```

```

<a href="feedback.html" onMouseOut="PPImgAction('out','tg_75')"
onMouseOver="PPImgAction('over','tg_75')"></a></div>
<div style="position:absolute; left:50px; top:756px; width:700px; height:34px;">
</div>
<div id="txt_42" style="position:absolute; left:50px; top:764px; width:700px;
height:17px;-moz-box-sizing:border-box;box-sizing:border-box; overflow:hidden;">
<P class="website-Body-Text-P"><span class="website-Body-Text-C">© 2010 Nguyen's
project </span></P>
</div>
<div id="vid_22" style="position:absolute; left:549px; top:22px; width:203px;
height:49px;">
<script language="JavaScript" type="text/javascript">
var hasRequestedVersion = DetectFlashVer(8, 0, 0);
if (hasRequestedVersion)
{
AC_FL_RunContent(
"src", "free-flash-clock-183",
"play", "true",
"loop", "false",
"width", "203",
"height", "49",
"salign", "R",
"scale", "ShowAll",
"id", "vid_22",
"quality", "High",
"wmode", "transparent",
"bgcolor", "#FFFFFF",
"name", "vid_22",
"type", "application/x-shockwave-flash",
'codebase',
'http://fpdownload.macromedia.com/get/flashplayer/current/swflash.cab',
'pluginspage', "http://www.macromedia.com/go/getflashplayer"
);
}
else
{
var alternateContent = 'An old version of the Flash plugin was detected.
<strong><a href="http://www.macromedia.com/go/getflash/">Please upgrade your Flash
plugin.</a></strong>';
document.write(alternateContent);
}
</script>
</div>
<div id="txt_52" style="position:absolute; left:61px; top:156px; width:680px;
height:602px;-moz-box-sizing:border-box;box-sizing:border-box; overflow:hidden;">
<P class="Normal-P"><span class="Normal-C0">CONTENT</span></P>
</div>
</div></center>
</body>
</html>

```