DEVELOPING A HIGH DEFINITION VIDEO QUALITY OF EXPERIENCE MODEL BASED ON INFLUENTIAL PARAMETERS

A thesis presented in partial fulfilment of the requirements for the degree of

Doctor of Philosophy

in

Computer Science

at Massey University, Palmerston North,

New Zealand.

Syed Jawad Hussain

2015

Contents

List of Tables

List of Figures

Abstract

Multimedia applications are quickly becoming a basic necessity for both business and personal situations. They are being used on daily basis, for a big part of the population, it's becoming the main source of entertainment and communication. Internet based Television and IPTV are now being preferred over standard Television due to the added functionality that it provides. Users not only expect good quality they also require it to be available continuously without fail. The research reported in this thesis is focused on the "Development of a high definition video Quality of Experience (QoE) model based on influential parameters". We have proposed a model for predicting QoE for high definition videos. We sought and justified appropriate parameters for inclusion into our prediction model. Our analysis has helped us in identifying parameter quantization as the most influential parameter from content domain. Whereas the parameters such as packet loss, packet reorder and jitter were found to be equally influential on QoE from the network domain. We also were able to show that parameter buffer size was the most influential parameter from the customer premises processing domain. We also showed that by integrating these parameters from the three domains of content, network and customer premises processing has enabled us to predict QoE with greater accuracy. The model enables the service provider to predict QoE and act before user perception goes beyond a predetermined threshold level. By including all the most influential parameters from all three domains we have ensured that we were accounting for a complete end to end effect of a user's perception. We were also able to identify the optimal configuration for good quality. In addition, we determined configurations that reduced the output quality to unacceptable levels that should be avoided. We were also able to find out a mathematical relationships between these parameters.

We have discussed the current state-of-the-art knowledge for QoE in the three identified domains and then proceeded to identify the shortcomings of models and methods found in the current literature. We made use of Taguchi robust designs for reducing the run time of experiments. This also helped us in analyzing many different configurations but only conducting experiments on a subset of these configurations. For three experiments the Taguchi DoE technique helped us reduce the number of test combinations from 81 to only 9 combinations. For the fourth experiment we were able to reduce the

number of combinations from 256 to only 18 combinations. We were able to reduce known boredom and memory effects by careful design of our experiments. Moreover, we were able to analyze the results using an appropriate signal to noise ratio (SNR) analysis method. This method helped us identify the most influential parameters for QoE in the three identified domains and it also assisted in the identification of optimal configurations. The method utilize quality loss model for identifying optimum configurations. By this method we found the SNR for each configuration and those individual parameter configurations were selected which ensured the highest SNR values. This helped identify the optimum configurations from each domain. In addition, we used the House of Quality (HoQ) method to validate the results of our SNR analysis. We were able to show that for individual domain the HoQ method was able to verify the results of our SNR analysis. HoQ only failed when we tried to use it to verify the end to end video quality degradation. We identified that the reason for HoQ method's failure in generating acceptable results was the use of weights to identify the most influential parameters in situation were all parameters under investigation where already found to be influential. These heavy weight parameters canceled the effect of each other and the parameters became statistically insignificant.

We proposed a QoE prediction model by using the most influential parameters identified from each domain. Our results show that these influential parameters were the major reasons for degradation in quality. Hence, using such parameters can help in better prediction of video quality degradation.

We made use of ordinal logistic regression for developing the prediction model. We proposed a complete model which was able to predict quality upto an 88% accuracy. Later, we developed a reduced version of the model for low computing solutions which was able to predict quality with 84.5% accuracy.

Acknowledgements

I would like to express my deepest and sincere appreciation and gratitude to my supervisor Prof. Richard J. Harris for his constant and continuous support, encouragement, understanding, patience, advice, criticisms and opinions. His unconditional support was the main reason I was able to finish this study.

I would also like to thank Dr. Amal Punchihewa for his polite and comforting mentoring and support. He was able to discuss issues with such a calm manner that it atleast momentarily pulled me out of the research blues.

I also would like to thank Higher Education Commission Pakistan for funding my studies and support me during this PhD study. I am once again thankful to Prof. Richard J. Harris for his financial support for conferences and living expenses at the end of my studies in New Zealand.

I would also like to thank my friends at Massey University especially Dr. Stephen Lean for sharing ideas and long discussions.

Lastly, I would like to thank my family for their unconditional support and motivation. It would not be possible for me to finish this PhD study without my mother. Her continuous support, encouragement, prayers and financial help pushed me towards achieving this goal. My wife who was always there when I needed a shoulder for support. My three kids who made me forget my work issues the moment I entered my home. A Special thanks to my brother and sister who prayed for my success since forever. A very special thank you to my father Syed Sajjad Hussain (Late) who planted the idea for my higher studies. I dedicate my successes to him.

Chapter 1

Introduction

1.1 Brief Thesis Description

Due to rapid growth of Internet in the last decade, multimedia applications have became very popular. The inherent nature of the Internet Protocol (IP) has introduced issues in high quality service delivery [1]. Research has been conducted to improve this quality by either using the existing capabilities of IP or proposing new methods of overcoming these issues. Multimedia applications were developed with the intention of providing greater user experience by amalgamating video, voice, text and computer graphics. These capabilities also made such applications data intensive and as a result, these were always struggling to deliver a high quality service [2]. Services which were not ensuring high end user perception were losing business as clientele shifted from low quality services[3]. There was a need to empower the service provider to predict, the end user's perception of quality. This ability would help a service provider in continuously providing a high quality service and retaining its clientele[4]. Even in crisis situations, a service provider should be able to provide the best service possible.

Internet protocol television (IPTV) was selected as an example service for this research. Within IPTV, the video on demand (VoD) service was selected, as it is one of the frequently used services of IPTV. The idea was to look at the end to end video quality which encompasses video content production and processing (content domain), its delivery over the Internet (network domain) and processing of delivered content on customer equipment (customer premises processing - CPP domain) available within the customer's premises. To develop a model for predicting a user's perception we needed to identify the parameters which should be included in that model. We conducted a literature review to find out which parameters were reported to have an effect on video quality. These influential parameters were included in this study. In the next step, we identified the most influential parameter from each domain. This was done

by conducting subjective experiments and recording the user's feedback. The experiments were conducted using Taguchi robust designs. The Taguchi method ensured a small number of experiment runs compared to experiments covering every possible combination of possible parameters and proposed a method for identifying influential parameters. This method also enabled us to identify the optimal configuration for the selected parameters. An independent method called House of Quality (HoQ) which is part of Quality function deployment (QFD) was used to verify the results generated by the Taguchi method. Once the most influential parameters were identified, they were used to develop a prediction model. Ordinal logistic regression was used for developing this prediction model. Results of this model were verified on an independent data set acquired for the validation exercise. All the experiments were done on a test bed developed specifically for emulating the conditions of a VoD service.

1.2 Research Context and Background

By the mid-1990s the common users had access to higher processing speeds and larger memory capacities, enabling them to do more than ever before on their PCs. Content rich applications and games started to make their way into the market place. These applications effectively combined text, sounds, video, and graphic animation in exciting new ways. Applications able to make use of different media elements came to be known as multimedia. Multimedia applications were inherently demanding greater performance from the host computer and the network. With the advancements in Internet technologies, the World Wide Web also became a driving force for enhancement in multimedia applications. Many multimedia based services became popular and demanded resources for upgrading the access network and processing power of the host machine. Upgrading the host machine was easy but upgrading the access network required a lot of effort and resources. Improvements in the access network ensured that more demanding applications e.g. Voice over Internet Protocol (VoIP), IPTV, video conferencing etc could be made available to users. Bandwidth, or the amount of data that a network connection is able to carry, is important for multimedia transmission quality. Due to the availability of better and faster access networks, the applications evolved and so did the human perception towards these applications and their expectations from the network. A lot of resources and money have been invested in the upgrading of the access network and this has ensured that some more demanding applications can be provided to the end user (IPTV, Peer to Peer Streaming TV etc). Video broadcasting over the Internet, i.e., IPTV, is one of the most promising multimedia entertainment applications on the rise. The key to a successful IPTV system lies in the quality of its service. However, the recent success of Peer to Peer (P2P) IPTV services [5], such as PPStream, PPLive and CoolStreaming [6][7][8] reflects the fact that this domain is an

ever-evolving one. With the growth of the access network and the advent of multimedia applications mentioned above, service providers were eyeing up this lucrative area.

Multimedia have opened up new avenues for businesses and multimedia technology has become a powerful tool for companies as well as it introduced new applications for PC users. Apart from corporate controlled content there is a lot of multimedia content available which is user created and published on the Internet. The availability of multimedia applications was only possible due to the extensive use of advanced compression and transmission techniques. Different multimedia applications may have very different traffic characteristics and performance requirements. Users show different behaviours and needs which are application dependent. User requirements are also dependent on their access mode, their device, etc. Humans usually have different opinion about the same issue and the same is true in the case of multimedia services. Some users may not be satisfied on a performance level while others may think that it is acceptable. A service provider must meet these needs despite them being very heterogeneous. A service provider can't rely only on the traditional objective Quality of Service (QoS) parameters for the network, such as load and the packet loss rate on router interfaces to provide a user satisfying service. We need to find the relationship between QoS and user experience [9]. Typical QoS parameters only ensure that the network resources are not congested [10]. Apart from the fact they do not give any insight into end to end performance or user satisfaction. Researchers discovered that QoS had a direct effect on business and they related business metrics to QoS [11]. As multimedia evolved, it began to look at the user perspective which eventually evolved into the concept of Quality of Experience (QoE or QoMeX or QoX). Much research is being conducted in the area of QoE but still there is a need to demystify the many grey areas within the QoE domain. Major research within the QoE domain was done by measuring QoE for a service or identification of localized artefacts affecting quality.

Quality of Experience (QoE) is an extension of the QoS concept that encompasses content generation, managing networks, local loop access and content processing at the customer premises. Existing challenge is to find a quick and simple way to estimate QoE. It should still be reasonably accurate and able to account for the diversity of needs, habits and customs [12] of users. Poor quality perception would lead to loss of business. A service provider's ability to measure user's perceived QoE, can become the difference between market minnows and market leaders [13]. Quality is an essential requirement for successfully delivering a service and continuation of the service. There is a need for targeting for a specific quality level in the network planning exercise. In addition, service level agreements (SLA) must be complied with for continuation of a service. For such compliance ensuring quality and avoiding the pitfalls, which lead to quality degradation, becomes top priority. The knowledge regarding optimal quality

and the configurations which ensure graceful degradation in service must be researched. For estimating quality ahead of time, we need tools which may not be very accurate but can successfully predict within a certain level of confidence.

1.2.1 Evolution of QoE from QoS

Quality of service that's been delivered by IP based networks is usually described as "best effort" where there are no service guarantees. This was good enough for internet browsing but certainly not adequate for supporting multimedia applications, interactive or non-interactive ones. In order to make IP based networks able to provide acceptable performance for such applications, resource management mechanisms were proposed either by packet scheduling and priority mechanisms, or by load balancing and QoS routing. Resource allocation decisions are presently driven by QoS parameters and service level agreements [14]. Services and networks were always looked at from the perspective of QoS. It was not possible to capture a user's experience. Quality of Experience (QoE) is an extension of the QoS concept that reaches right back to the users and the content generator and takes into account the users' needs in designing, monitoring and managing networks. It has been described as "a consequence of a user's internal state (e.g., predispositions, expectations, needs, motivation, mood), the characteristics of the designed system (e.g., complexity, purpose, usability, functionality, relevance) and the context (or environment) within which the service is experienced (e.g., organizational/social setting, meaningfulness of the activity, voluntariness of use)" [15]. According to Brooks [16], QoE is a measure of user performance based on objective and subjective psychological measures of using a service or product. Though user experience embodies psychological measures, there is a need to express it in relation to the networks and equipment that influence user behaviour. QoE data need to possess user experience and technical measures so that we could express user experience, which is using a service with known levels of QoS. Hence we say that QoS is related to QoE because a QoE measure needs to be stated together with the technical conditions of a communication service if it is to be useful for stakeholders. If a service should be improved for customers or end-users, stakeholders need to know that the QoE level is not good enough and should be able to decide which one or more technical QoS parameters could be improved in order to achieve a higher QoE. Consequently, QoE should be expressed in QoS terms. According to Fiedler et al [17] the relationship between QoE and QoS is exponential, for this they used the exponential interdependency of quality of experience and quality of service hypothesis (IQX) they inserted the measured QoS values into a corresponding exponential formula, later their impact on QoE was assessed and analyzed where QoS parameters reflected the level of disturbance and QoE parameters represented level of satisfaction. According to Zapater et al [18] QoE

is a measure of end-to-end performance at the service layer and QoE takes into account how well a service meets the needs of customers. Whereas QoS is network centric. Khirman et al [19] identified that the fundamental assumption behind the traditional QoS approach is that the measured quality of service is closely related to the quality of experience (QoE) for the end-user. QoE as defined by ETSI TISPAN TR 102 479, is the user perceived experience of what is being presented by a communication service or application user interface [20]. This definition itself suggests some factors that influence the experience of a typical user. We note that some of this is highly subjective and takes into account many different factors beyond simple quality of service considerations, such as service pricing, the viewing environment, stress level and so on. According to ETSI TR 102 643 QoE is defined as "A measure of user performance based on both objective and subjective psychological measures of using an ICT service or product" whereas according to P.10/G.100 it's defined as the overall acceptability of an application or service, as perceived subjectively by the end user. QoSE (QoS experienced/perceived by customer/user) ITU-T E.800 A statement expressing the level of quality that customers/users believe they have experienced. QoE definitions are evolving as research continues for understanding it and its impact on how the next generation networks will be designed or planned. The future research directions identified by the white paper published during a seminar entitled "QoE: From User Perception to Instrumental Metrics" held at Schloss Dagstuhl May 1st to 4th, 2012 highlights the importance of multidisplinary research for Quality of Experience (QoE) [21] [22]. As QoE is user focused and encompass acceptability, delight and performance, it is seems that it will become the key role is service provisioning and management. Dagstuhl work also commented on migration of focus from QOS to QoE and the challenges of bringing together the user, technology, and business. A major challenge is that the qualitative user perception needs to be translated into quantitative input which should further be used for dimensioning, managing and controlling network and the deployed services. Dagstuhl paper also proposed the use of feedback relating to service acceptance, usage, cost, and quality for evaluating QoE. The research generated during and after the seminar is helping develop standards for QoE and for developing metrics and measurement techniques aimed at improving QoE prediction.

Quality of Experience (QoE) is an extension of the QoS concept that reaches right back to the users and the content generator and takes into account the users' needs in designing, monitoring and managing networks. Network planning needs to be more aware of QoE requirements in order to more fully take into account the needs of the end users. This can be achieved by carefully monitoring key network performance indicators and appropriately managing network elements. There is also a need for visualizing QoE as a requirement for such a planning exercise. For planning ahead of time we need tools

which may not be very accurate (initially) but can (ultimately) successfully predict resource requirements in the network. Network planning and design is an iterative process which encompasses topological design, network-synthesis, and network-realization, and is aimed at ensuring that a new network or service meets the needs of the subscriber and operator. The process can be tailored according to each new network and service. By network planning we want to achieve our goals through forecasting traffic demands and characteristics, dimensioning resources, traffic engineering the network flows and we want our network to degrade gracefully under failure conditions. Forecasting means estimating the expected traffic loads that must be supported by the network whereas dimensioning a new network or service involves determining the minimum capacity requirements that will meet a Service Level Agreement (SLA) for a client. Hence we need to plan for peak-hours traffic etc. Traffic Engineering involves such functions as changing traffic paths on the network to alleviate traffic congestion or accommodate more traffic demand. This technology is critical when the cost of network expansion is often prohibitively high and network load is not optimally balanced. Survivability criteria specify standards that require the network to maintain maximum network connectivity and quality of service under failure conditions. It has been one of the critical requirements in network planning and design. We want to achieve such objectives from the perspective of QoE.

1.3 Description of the Problem

Multimedia applications have attracted a lot of attention due to their apparent usability and utility and hence became a potential profit market for service providers. All service providers want to maximize their profits and to capture a major chunk of this business. Better applications and standards were in demand and since then a lot of work has been done in this area. IPTV has proven its worth and rapidly evolving traditional TV culture. This service should be reliable and always available. From a user's perspective they are implicitly expecting a better experience from the newly deployed service. The problem in providing this quality over the internet was the way in which internet communications were designed initially. They were not designed for such demanding real time applications; however, due to the significant investment in IP based networks, IP has become the core technology for future networks and promises the same level of reliability and consistency in quality have become an uphill task for service providers. It is important to mention the efforts made to enable IP networks to work with these upcoming demanding applications and these include, for example, class of service (CoS) support in the IntServ and DiffServ models by the IETF [23][24]. Expectations of better service imply that service providers need to improve their monitoring and measuring methods. It also means that, as these technologies evolve and more users are able to tap

into faster access networks, service providers will be in for another challenge i.e. that they will have to ensure that this service can be provided to an ever-expanding clientele at a minimum expected level that will increase over time. This gives rise to the issue that we wish to investigate and provide a solution to. The service provider is in need of a tool for predicting user's QoE and for configuring the parameters throughout the communication chain to provide the minimum expected quality. This capability should be available in real time which will enable the service provider to avoid situations where the clients decide to leave the service provider due to low quality. A service provider could measure it and automatically or manually improve on the quality or at least ensure a graceful service degradation if there is a problem.

1.4 Motivations for This Thesis

IPTV and all its variants are being used every day by millions of people. These multimedia services are being supported by tablets and smart phones. Service providers will only be able to stay in business if their service is continuously of good quality. For this they need to ensure that they know the pitfalls which degrade quality.

The point of concern is that the service providers are only looking at their own network i.e. QoS parameters and trying to predict the end to end service quality. The first issue that arises from such measurement or prediction is that a user is actually oblivious of a service provider's perspective. A user is only concerned with the output quality that he/she gets and decides to leave or retain a service on the basis of his or her perception i.e. QoE about the service being offered. QoS only looks into one third of the problem. The other two parts are related to content and CPP domains over which a service provider may have no control. A second issue arises from the scenario where a service provider is happy with the QoS quality whereas the user is unhappy with the provided service.

The earlier work in this area was focused on using content and network domain parameters for developing models. Moreover, from our literature review we were not able to identify scientific justifications for using a certain parameter for model development. There is a need to research this aspect of model development and to identify the reason behind each parameter selection. Then we also need to find out the inter-relationship of these parameters and their order of effect on the end to end quality.

The subjective evaluation techniques usually make use of a well-known video data set. Such content is neither rich in experience nor adequately rich in quantity [25]. These two aspects introduced boredom for the volunteers participating in the experiment. This scenario raises the question whether their feedback is valid or not? It has already been identified by researchers that boredom and memory effects influence subjective experiment results [26][27]. There is a need to fix the boredom issue as

well as to answer the question of validating the feedback. The issues of randomness in the subjective feedback provided by the users regarding the service quality need to be addressed as well. If the feedback is random it will invalidate any analysis or finding based on that data [28]. Statistical techniques need to be visited to ensure that the feedback is not random and, only then, the data should be used for analysis.

There has been a lot of work done in QoS and its been realized that a higher level of QoS is critical for delivering good service. However, it is not the only requirement for a successful and continuous service delivery. There is a need to look at the whole problem from users' perspectives and to find out the relationship between QoS parameters and user perception. A model could be a successful model if it can predict the user perception on the basis of QoS parameters. Effectively it's necessary to consider QoS from the content and CPP domains in addition to the network domain. We need to look into the content domain and the customer premises domain parameters for finding the answer to this question. We arrived at the following main questions that need to be researched.

1. How can we collect a true user perspective of video quality?
2. How can we justify inclusion of any specific parameters for QoE prediction?
3. How can we develop improved models for predicting QoE?
4. Identifying what are the influential parameters for model development?
5. What are the parameter configurations to avoid and parameter configurations for optimal service delivery.

A detailed literature review is included in Chapter 2. This discussion highlights the fact that video QoE is an active research domain. Multiple organizations are striving for developing standards for video quality assessment, service delivery and processing standards ETSI TS 102 034 V1.5.1 (2014-05). Moreover, a lot of work is being done on individual parameters to study their effect on video quality. Researchers found relationships between a single parameter and QoE. This exercise is good for highlighting the fact that these individual parameters affect overall quality. Unfortunately, for service deployment or continued service such research is not very useful. In the domain of video quality assessment we only found a handful of models, for predicting video QoE, which were developed on three or more parameters. There is a need to develop models which are able to predict video QoE and they include the most influential parameters from all three domains. In addition, generalized models are good but there is a need to develop application centric models. Different applications are affected by a varied set of parameters. This variation in parameters depends on the type of communication (interactive and non-interactive), type of content (high definition or standard definition)

and timing (real time or non-real time). For effective model development the literature identified the pitfalls as follows:

1. Varied classes of video content i.e. level of motion and complexity affect video quality, especially high motion. Models need to account for such variation within video.

2. Psychological parameters like boredom and memory also affect a user's perception of appropriate video quality.

3. Randomness of user feedback can invalidate the subjective experiment.

4. Including insignificant or unimportant variable in the modeling process only will increase the model complexity without gains in model accuracy.

To develop an effective model the above mentioned pitfalls should be avoided and research must be conducted to efficiently workaround these issues. The above discussion highlighted the reasons we set out to develop a model for video QoE prediction.

1.5 Thesis Aims and the Scope of Research

The above sections highlighted the problem domain and the need for developing this thesis. The principal objective of this work is to answer the question: Can an end-end QoE mathematical model be developed across the three domains comprising of the content domain, network domain and CPP domain, that connect customers to video / IPTV services? The following sections will also highlight the auxiliary research aims and objectives.

1.5.1 Research Aims

The research aim was to develop a predictive model for video quality encompassing the three domains. To collect data for model development we want to conduct subjective experiments. Our aim is to improve the data collection method and to validate the collected data before using it for model development. This will necessarily improve the model as the data will be a true representation of user's perception of video quality. The objectives of this research are as follows:

1. To develop a video database for improving subjective quality assessments against boredom effects.

2. To validate subjective assessment feedback against randomness.

3. To find justification for including parameters for model development

4. To identify optimal configurations for successful and continuous service delivery.
5. To identify configurations which should be avoided for successful and continuous service delivery.
6. To find out the most influential parameters from the three domains.
7. To identify parallel methods for identifying the most influential parameters (validation).
8. To develop an end to end model for predicting video QoE.
9. To validate the results generated by the prediction model.

In order to achieve the above mentioned aims we shall deploy the methodology discussed in detail in Chapter 3.

1.5.2 Scope

While conducting the literature review, we realized that there are many parameters which influence video quality. These parameters can be classified as either human psychological factors or technical parameters. It would have been a huge job to incorporate all these parameters in our research. Hence, there was a need to make certain assumptions and later, by using existing methods, reduce the dimensions of this problem. The following assumptions were made to reduce complexity to a manageable level.

- Bandwidth: we did not consider bandwidth as a candidate parameter because we assumed that as local access provisions have improved (and continue to improve) drastically we can assume it can provide us with the best possible service. In addition we also ignored the fact that most communication is taking place on Wi-Fi within the home network. We considered the communication to be on the wire where adequate resources are available for high definition quality.
- We made the best effort to use the same hardware for all the experiments in order to ensure against any unwanted effects of hardware variation.
- The volunteers who participated in the research were replaced for every experiment. Unfortunately, we had to request a few volunteers to participate in more than one experiment. It was necessary to rotate the volunteers so that they did not become experts of the experimentation process and start to provide us with biased feedback rather than a pure perception of video quality.
- Voice was provided for a better experience but was not accounted for in the feedback.

- Synchronization issues were also neglected for this study. Effects of synchronization on HD video and audio were not studied in this research.

After each experiment we had an informal discussion with the volunteers. This session was aimed at asking them informal questions regarding the experimental setup and the content that was included in the experiment. We were also interested in the apparent behaviour of users. We found out that almost all the volunteers were happy to participate in the experiment as the content were very recent and was quite interesting. But this activity was not including formal psychological investigation in order to evaluate the user's behaviour and validity of the feedback.

1.6 Summary of Contributions

The main contribution of this thesis are as follows

- Taxonomy of influential parameters for video QoE.
- Proposing a framework for parameter classification into domains
- Novel use of Taguchi robust design in end to end video quality assessment.
- Use of Fleisś Kappa method for checking video quality feedback randomness.
- Novel study of content domain parameters, analysis of their effect on QoE. Identification of content domain most influential parameter. Identification of optimal configuration for content domain parameters.
- Novel study of network domain parameters, analysis of their effect on QoE. Identification of network domain most influential parameters. Identification of optimal configuration for network domain parameters while playing varied content types.
- Novel study of customer premises processing (CPP) domain parameters, analysis of their effect on QoE. Identification of CPP domain most influential parameter. Identification of optimal configuration for CPP domain parameters while playing varied content types.
- Novel study of end to end QoE effect of most influential parameters from three domains of content, network and CPP. Identification of most influential parameters from end to end QoE perspective while playing varied content types.
- Novel use of House of Quality method for validating results of Taguchi signal to noise ratio analysis method.

- Analysis and modeling of QoE for end to end IPTV VoD service. This model included parameters from content domain, network domain and CPP domain. Purpose of such a model is to explain the relationship between parameters, to predict QoE MoS and helping in providing an optimal continuous service while playing varied content types.

1.7 Thesis Outline

This thesis describes an effort to understand the relationship between multiple parameters from content domain, network domain and CPP domain, all having a significant influence on the end to end QoE. Chapter 1 briefly introduces the thesis, a description of the problem, research background, motivation for the thesis, the methodology in brief, thesis aims and objectives and lastly provides a summary of contributions. It concludes with this Thesis outline.

Chapter 2 gives an overview of existing QoE research. Main focus of this review is on introducing the state of the art research in video QoE. A framework was proposed on the basis of three domains of content, network and CPP. Parameters were identified from earlier work which affect QoE in context of video communication. Gaps in current work were highlighted and research questions were developed on the basis of these identified gaps.

To introduce the general frame work adopted to answer the research questions detailed methodology is discussed in chapter 3. A discussion is included on method selection. These method are adopted within the general methodology for proposing solutions. Multiple parallel methods are considered and the most feasible methods are selected for this thesis. A discussion is included for experimental design and selection of an appropriate design for experiments conducted for this research.

Chapter 4 presents findings and analysis of the effects of multiple Quality of Experience (QoE) parameters in the content/compression domain. The objective of this component of the research is to investigate best candidate parameters for model development. The Taguchi robust design method is used to achieve this aim. This method also provided us with an analysis technique to identify the most influential parameters as well as the least influential ones for QoE. This chapter also highlighted the relationship between multiple parameters of the content domain, the most influential parameter from this domain and optimal configuration of parameters from this domain. The idea is to find out the optimal configuration of parameters from content domain which can ensure better quality video service delivery.

Adaptation of the above mentioned methodology enabled experimentation for network domain and CPP domain. Detailed discussion and results of experiments for network domain and CPP domain is included in chapters 5 and 6 respectively.

An end to end study of influential parameters is conducted on the basis of the results of previous experiments. Chapter 7 presents findings and analysis of the effects of multiple Quality of Experience (QoE) parameters across three domains i.e. content, network and customer premises processing. This chapter discusses the end to end video quality effect on human perception. The aim of this experiment is to identify the most influential parameter across the three domains. Their order of effect and the optimal configuration of these parameters which could ensure best effort service quality under the influence of noise factors. The parameters selected for this experiment are the most influential parameters found out in the previous experiments.

Chapter 8 discusses the model development and provides a discussion regarding the predictability of the model. An ordinal logistic regression model for predicting video MOS is developed. A minimum model is also developed for low complexity implementations. Though there is a compromise in accuracy with the minimum model. An independent data set is used for model validation. The results show that the model is able to predict the variations in the independent data set with less implementation complexity. The minimum model also performed well with the validation data set and results show that it can predict video QoE with reasonable accuracy and less implementation complexity.

Chapter 2

Literature Review on Quality of Experience

2.1 Chapter Overview

This chapter gives an overview of existing QoE research. The main focus of this review is on introducing the state of the art research in video QoE. A framework is proposed on the basis of three domains of content, network and CPP. Parameters are identified from earlier work which affect QoE in the context of video communication. Gaps in current work are highlighted and research questions are developed on the basis of these identified gaps.

2.2 Quality of Experience

Quality of Experience (QoE) is looking at service quality from user's perspective. Study of QoE can enable tool/model development which could predict human perception. Such a model could prove invaluable and can help service provide increase his revenues.

2.2.1 Definitions of QoE

Quality of experience (QoE), as defined by ETSI TISPAN TR 102 479, is the user perceived experience of what is being presented by a communication service or application user interface [20]. This definition itself suggests some factors that influence the experience of a typical user. QoE is a highly subjective measurement. It takes into account many different factors beyond simple quality of service considerations, such as service pricing, the viewing environment, stress level and so on. Werner [29] states that "IPTV is not a well-defined term and may be a source of ambiguity and sometimes confusion". Werner also explained the difference between "IPTV" and other IP based

TV services commonly known as "WebTV". Conventional television used to provide a guaranteed level of service, however; an increasing number of households that have broadband connections (which enable them to access video streaming and to download files), are now using the ubiquitous IP protocol. Unfortunately, these services have no service guarantees. The IPTV services that telecommunication companies aim to launch are based on the IP protocol. Thus there is an urgent need to assure the same quality as achieved in conventional television. Presentation of TV services via the Internet raises many new challenges to service providers since the Internet is based around a completely new paradigm for service delivery. This leads to a completely new set of factors that can influence a user's experience of the service. Our research attempts to identify the factors that influence QoE for a multimedia service such as IPTV.

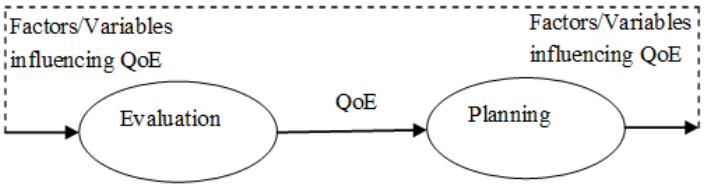

Figure 2.1: Utility cycle of QoE

2.2.2 QoE vs QoS

This section will discuss the relationship between QoE and QoS. How they are related to each other and what are the differences. The principal differences lies in the perspective of providers and users i.e. QoS is the perspective of a service provider whereas QoE is the user's perspective.

2.2.3 QoE for IPTV

IPTV is being deployed by many telecommunication companies all over the world. Wider acceptance by the public of IPTV as an alternative to current television services is dependent upon its promise of quality and reliability. Currently, to the best of our knowledge, the industry still lacks comprehensive quality planning and evaluation tools for IPTV. When the service is operational, QoE can be predicted for that service based on resources committed and factors that affect that service. Therefore, it is important to identify all possible factors that could influence QoE for users of an IPTV service.

An extensive literature review at the time of writing has revealed that no single researcher has surveyed all the end-to-end factors that influence QoE for an IPTV multimedia environment. Most research has been concentrated in only one of the following three domains; viz media content, network and user-end factors. In order

to construct a comprehensive framework to model QoE for IPTV it is necessary to consider all the factors affecting Quality of Experience in all areas i.e. acquisition, coding, delivery, customer premises processing including environment, navigation and user behaviour.

2.3 Artefacts and influencing factors

Literature review of the current state of the art research helped us identify the artefacts and factors/parameters which influence QoE. The following discussion provides a list which affect QoE and must be studied for developing predictive models.

2.3.1 Classification of influencing factors

We have reviewed the most referenced publications related to QoE. The research objectives of most of these publications were to identify the effects of certain parameter on human perception i.e. QoE. These publications made use of different application with varied levels of interaction. These applications were IPTV, VoIP, kiosk, web traffic and "YouTube"-like peer-to-peer services. As mentioned in the thesis scope, we were focused on QoE for IPTV but we looked at the literature discussing other applications as well. This helped us identify suitable parameters having an effect on QoE for IPTV. Over 21 publications were found that identified the factors listed below and a detailed commentary on these publications is included in Sections 2.4, 2.5 and 2.6.

1. Content
   (a) Video
      i. Dynamic range of the Y U V signals
      ii. Gamma correction factor
      iii. Bandwidth/slopes of the filters
      iv. Automatic Gain Control (AGC)
      v. Stability level for the camera clocking device
   (b) Audio
      i. Loudness
      ii. Number of channels
   (c) Coding
      i. Codec
      ii. GoP (Group of Picture)
      iii. Bit Rate

A. Variable Bit Rate
B. Constant Bit Rate

iv. Level of compression/Quality factor

v. Level of motion
A. High motion and panning
B. Medium motion
C. Head and Shoulder

vi. Artifacts introducing compression Techniques
A. Low pass filtering
B. Removal of HF components of temporal data
C. Removal of HF components in spatial domain
D. Block processing
E. Connection mismatch
F. Colour sub sampling
G. Colour Quantization

2. Delivery

(a) Jitter
(b) Packet loss ratio
(c) Bandwidth
(d) Packet loss distribution
(e) Queuing/De-queuing delays
(f) End to end Latency
(g) One way mean delay
(h) Channel Noise
(i) Protocol
(j) Path unavailability

3. Customer premises processing

(a) STB (Set Top Box)
i. Decoding performance
ii. Coding latency
iii. Channel Zapping latency
iv. Error correction

v. Synchronization (Lip Sync)

(b) Display

i. Resolution

ii. Refresh Rate

iii. Number of bits

iv. Display Technology

v. Form Factor

(c) Environment

i. Ambient Light

ii. Ambient Noise

iii. Vibration

iv. Wind

(d) Navigation and user behaviour

i. Structure of Navigation

ii. Multimodal inputs

Looking at these factors we realized that the end to end video communication chain can be divided into three groups or domains. Figure 2.2 shows these domains.

2.4 Content Domain Artefacts

Content domain consist of acquisition and coding. Acquisition artefacts can be avoided if skillful professional use high quality equipment. Coding artefacts are usually introduced as a compromise to the requirement of limited bandwidth and storage. The following discussion highlights the important factors.

2.4.1 Acquisition

There is an upper bound on the quality of the content as determined by the source of the original content. However, viewers may be accepting of poorer quality given the circumstances of the originating environment. One example may be for news, viewers would accept poor quality video if sophisticated equipment cannot be used to obtain the news footage. High commercial valued content such as sports coverage needs high performance optical systems and low noise image sensors for acquisition. Charge Coupled Devices (CCDs) frequently used in most digital cameras introduce a number of different noise like electronic fixed patterns and “dark” noise. Before acquisition, a digital camera performs dark reading and later subtracts it from the exposure signal.

Other reasons for noise introduction are the fill factor of pixels or the actual percentage of sensor elements. Smaller elements have a higher inherent noise ratio. Noise is also introduced during an analogue to digital conversion of image data. The noise level is high in situations with low light where the noise to signal ratio will be highest. Blooming or light spill-over is caused by photons spilling from one sensor element to another creating what can be a whole region of over-fill, resulting in highlight blow out and or weird colours in these areas. When a relatively small sensor array is used to create an image, pixilation becomes very apparent. Larger sensor arrays are more expensive but supply enough information to produce a more lifelike picture. "Christmas tree lights" is a variation of colour aliasing artefacts. On many sensors, filtration is applied with twice as many green pixels as red and blue and this is done to emulate human vision. This results, if blown up especially on diagonal lines, in an unreal mosaic of colours.

There are many different ways to expand an image's size, such as Linear, Bilinear and Bicubic methods. Bicubic interpolation is widely regarded as the best method. Camera manufacturers create their own interpolation systems specific for the task and hence create some undesired effects. Some loss of perceived sharpness occurs at the capture stage with almost all sensing devices, and digital techniques are used for compensation. 24-bit colour information is inadequate in some cases. Photographing a rose with vivid red colours would certainly require more bits for proper representation as the 24-bit palette has, in effect, just 255 levels of pure red to represent the flower's colours. That's why a capture system that uses a 30 bit or 36 bit representation is better as it offers more colour and tone choices. Hence we conclude our discussion in this section by noting that selection of hardware is very important for end to end quality and these factors must be considered for measurement of quality.

2.4.2 Coding

Coding in this context refers to both source coding and channel coding. To achieve bandwidth efficiency, higher levels of compression need to be achieved. This leads to compression artefacts such as blockiness, blur, ringing, mosquito noise, jitter etc. Error correction coding is performed at the cost of influencing the effective data rate. Higher levels of error detection and correction need more redundant bits. Advanced video coding techniques based on MPEG-4 part-10 have built-in error correction ability. Recent advances in scalable video coding can offer the capability to code only once to meet a wide range of viewer display requirements. There are number of research papers that have reported on the evaluation of compression artefacts such as blockiness, blur, ringing, mosquito noise [30] [31] and jitter [32] [33]. Coding in this context refers to both source coding and channel coding.

To assess image quality as perceived by a user in an automated manner there is a

need for online measurement systems. This avoids the usual costly (in terms of time and money) subjective techniques which are also a non-repeatable method for collecting subjective scores. Janowski et al. [34] used mapping models which were constructed using the Generalized Linear Model (GLZ). They are a generalization of the least squares regression in statistics for ordinal data. They were able to compute overall qualitative image distortions based on partial quantitative distortions from component algorithms operating on specified image features.

According to Wolf et al. [35] the current objective video quality measures achieved good prediction of subjective ratings. Cerqueira et al. [36] proposed a metric for measuring video artefacts and the results show that although correlation with subjective scores is quite high, the blockiness metric has a lower correlation. Maalouf et al. [37] propose a Reduced Reference (RR) perceptual Image Quality Measure (IQM) based on the grouplet transform and they claim that the proposed method performs well and has good consistency with subjective quality assessment. They performed rational sensitivity thresholding to obtain the sensitivity coefficients of both images based on human visual system properties. Narvekar et al. [38] presented work which is a no-reference objective sharpness metric based on a cumulative probability of blur detection. Their work also takes into account the Human Vision System (HVS) response to blur distortions and they claimed better performance for images that have background and foreground blur distortions which are different. Whereas Zhu et al. [39] worked on sharpness metric detecting both blur and noise based on image gradients. Their proposed metric behaves as an indicator of the signal to noise ratio but there is a need for prior estimation of noise variance. Mosquito noise is a compression noise which has temporal aspects for which Mantel et al. [30] presented a spatio-temporal and compression independent method to remove mosquito noise.

Work done by Ninassi et al. [40] developed a perceptual full reference video quality assessment metric which was focusing on the temporal evolutions of the spatial distortions. The technique that they used assimilated temporal variations of spatial distortions at the eye fixation level and whole video sequence into short and long term temporal pooling. Rahayu et al. [41] concluded that there is no objective model that comes out as a best performer from a statistical point of view for high quality data. They compared Peak Signal to Noise Ratio (PSNR), Multi Scale Structural Similarity (MS-SSIM) and Single Scale Structural Similarity (SS-SSIM). Hence SSIM was not able to perform better than PSNR as it does in the case of standard or low quality images. Staelens et al. [42] tested full length movies through subjective testing. Their aim was to study the scalability effects on user perception while using scalable video coding extensions of MPEG 4. The study reveals that users favor temporal scalability over quality scalability. Reiter et al. [43] conducted a study to evaluate the PSNR's

effectiveness in estimating relative subjective quality levels for different types of quality distortions. The results show that PSNR is not a reliable metric for doing such tasks. It can't be relied upon for assessing the combined effects of compression and transmission artefacts.

The most notorious artefacts which occur due to different techniques of compression must be analyzed from an end to end perspective. Blur and blockiness are introduced due to low pass filtering and block processing respectively. Ghosting appears due to multipart effects because of the mismatch between connections either electrical or optical. Colour bleeding occurs due to colour sub sampling and quantization of colour information whereas mosquito noise appears due to the removal of high frequency components in temporal video data i.e. between frames and ringing occurs due to removal of high frequency components in the spatial domain i.e. from a single frame.

2.5 Delivery Domain Artifacts

The term delivery refers to the IP based communication channel from the content server at the head end to the viewers' location where multimedia contents are viewed. Since the delivery of multimedia content such as IPTV is via an IP channel, it suffers from traditional shortcomings of the IP protocol unless special arrangements have been made. Multimedia services such as IPTV need service guaranteed delivery to maintain the required QoE.

It has been observed by Liu et al. [44] [45] that perceptual video distortion is caused by not only source coding artefacts and packet losses [46] [47] but also there is a joint impact of the two losses. Packet error has a spillover effect as the information content becomes corrupt whereas complete reference frames are required for decoding. Hence errors tend to propagate. At high Packet Error Rate (PER), its PSNR is almost constant. The following could be the reason behind it, at low error rates, few corrupt reference frame packets affect reconstruction of several other dependent frames; and at higher error rates, reference frame packets and the dependent frame packets are simultaneously corrupt. There is a difference in measurements between PSNR and Video Quality Metric (VQM) in different situations [48]. At higher PERs, VQM continues to rise. This is shows growing user dissatisfaction with increasing error rates. Human subjects were also able to distinguish between error spilling at low error rates and network errors at high errors where many frames are naturally corrupted. IP fragmentation breaks individual frames into packets. An I-Frame is usually divided into 16 packets with a fragmentation of 1024 bytes. With every lost packet, part of a frame is lost. By observing MPEG-4 headers it is obvious that an I-Frame affects the quality more than other frames. I-Frames are used for the reconstruction of other frames. Due to temporal redundancy in a sequence, successive I-Frames may be reconstructed or error

patching techniques are used to generate lost frames. If we lose the first reference frame there is play out degradation for the entire length of the group of picture (GOP) or a complete "white out" especially when the very start of a video frame is lost [49].

The effect of packet error is far less than the effect of a packet loss. User perceived degradation reduces as more intelligent error patching is applied to reconstruct frames using previous intact frames. For interactive messages, the data stream has very tight delay bounds; hence the time to deliver becomes an important criterion, whereas if queuing delays are uniformly present throughout the play out they have little effect on quality. PSNR remains constant due to pure delays alone, decreasing marginally at high delay, especially when combined with packet losses in the network, queue overflow at the collector node. Loss has a prominent effect as there is noticeable reduction of PSNR readings with the very introduction of loss. With the introduction of losses in combination with delay, VQM indicates stronger user dissatisfaction when there are few losses compared to no losses, while there is little change in opinion between a loss rate of 0.01 and 0.02.

Jitter, the variance of the per-packet delay, is caused by non-uniform sharing of the links and has direct implications on the receiver buffer capacity. High jitter can lead to both buffer overflow and underflow and these situations lead to distortions [50]. Due to jitter the video sequence sometimes sees many packets ahead, and sometimes very little competition [51]. PSNR reacts strongly to values of jitter exceeding 0.05 seconds, dropping rapidly but VQM suggests that human subjects can tolerate jitter levels of around 0.06 seconds.

2.6 Customer Premises Equipment Domain Artefacts

Customer premises processing in this context refers to factors that involve processing multimedia content from the service provider delivery point to the point of consumption. Generally, either copper or fiber connects to a set-top box or media box which processes the multimedia content. Decoded video is interfaced to a video display and the audio content is presented to a multi-channel audio system. Future systems may include other features such as vibrating chairs as part of home theater systems.

Since there is a wide range of different displays, audio and viewing environments, each viewer may perceive different quality from the same multimedia content. Rapid advances in display technology are continually leading to a higher viewing experience. CRT, projection, large tiled displays, plasma, LCD and LED are common types of display that are based upon a wide range of operating principles. They offer different contrast; dynamic range, resolutions, and frame refresh rates etc. These different display technologies introduce somewhat different artefacts due to compression distortions [52]. According to [53] for high resolution materials the perceived quality is higher for

CRT displays when compared to LCD. Pinson et al. [54] says that a CRT high resolution monitor can probably be used to emulate the subjective experience of viewers utilizing an LCD low resolution monitor. Pechard et al. [55] explored the effects of distortion in reference to display size. He used a subjective approach to measure QoE for H.264 distortion over different display sizes. They claimed that the ideal distance selected by users was 8H instead of 3H which is recommended for HD (where H is the Height of the video). They used 21 subjects and there is no mention about the homogeneity of the sample which could make the sample skewed towards one type of viewing experience. They also stated that image size and distortions influence user perception but distortion is the predominant factor once HDTV is compared with low quality Standard Definition Television (SDTV) whereupon in case of high quality SDTV image effect become more important. For low distortions large image size gets a positive perception whereas for high distortion levels it's other way around. Which sums up to the fact that for HDTV to become a standard more service quality is to be offered as people prefer SDTV as it reduces the visual impact of distortions.

Pechard et al. [56] compared Absolute Category Rating (ACR) and Subjective Assessment Method for Video Quality (SAMVIQ) subjective methodologies concluding that SAMVIQ is more accurate for higher resolutions. There has been isolated work on highlighting artefacts introduced due to the advanced technology used in these types of displays e.g. the synchronization mismatch in large tiled displays. There have been some expensive hardware solutions available in the market but they are not economically viable for research institutes or medium to small organizations. The work done by [57] is an effort towards helping in the design of low cost large ultrahigh resolution tiled displays where this artefact of mismatch in synchronization can be mitigated. The authors used subjective testing to come up with a threshold value in msec which can be translated into Double Stimulus Impairment Scale (DSIS) scores. He was also able to show the relative difference in synchronization mismatch in displays with bezels than ones without bezels. His other findings were that people prefer tiled with smaller or no bezels and that the discomfort threshold is larger with bezels compared to without bezels. The discomfort threshold depends on the level of motion and it decreases as the number of tile increases. The work could be more effective if objective measures were used to predict the DSIS in which case the model could be applied in real time.

The ambient environment affects the perceived quality of experience as for test purposes the ITU has come up with some subjective testing recommendations i.e. ITU-R BT.500-11 and BT.710-4 which specify the ambient environment e.g. illumination conditions, viewing distance and display parameters for specific type of Multimedia (SD and HD Television). According to [58] the Sensory Effect Description Language (SEDL) is being standardized by ISO/IEC MPEG. This descriptive language is used

for triggering sensory effects while experiencing multimedia content. In this paper they only concentrated on determining the colour of light effects form the content. There is a need to study user perception where sensory effects are triggered. Though it seems to increase the quality of experience further but the effects of quality losses like jitter, packet losses and/or SEDL losses are yet to be studied and may affect the user experience drastically.

Selection of buffer size within the set-top-box also effect video quality play at the user end [33]. There is a need to set the buffer size with the optimal setting which could work with different types of video content [50] [59].

In [60] discussed the study conducted to compare unimodal and multimodal remote controls. They claim that the results indicate multimodality improves user experience even though there is either little or no increase in system usability. The paper focuses only on hedonic attributes, stimulation and identity. They also identified that multimodality interfaces are providing universal access as different user group barely showed differences and that multimodality benefited older users [61]. We sum up our discussion about customer premises equipment by saying that selection of customer premises equipment impacts upon the perceived quality and the upcoming models for measuring end to end QoE need to consider these parameters such as display type and their underlying technology. Users will be making the decision, based on cost or performance, for buying or selecting such equipment but service providers need to come up with benchmark equipment that can promise users good quality. Thus the benchmarks could vary between service providers depending upon their implementations.

2.7 Standards

In the past few years, a great deal of effort has been put into standardization of QoE measurement methodologies. Due to these efforts, standards for voice and video and multimedia have started to appear. ITU-T recommendations P.910 and G.1081 were examined for this study. P.910 discusses the procedures from test setup to subjective analysis. It also recommends the type of devices and a minimum set of capabilities that must be present so that the quality degradation which is observed by the subject may only be due to coding and delivery rather than due to poor capturing and display equipment. It also recommends how the content may be organized to control the flow of tests as otherwise they might introduce fatigue and boredom in the subjects which could lead to skewed measurements.

These points raised in the recommendation highlight the psychological problems which are associated with subjective testing methodologies. G.1081 specifies the performance monitoring points for an IPTV service. These points are in different domains and cover content provider, service provider, network provider and the end user. The

framework we have proposed is in accordance with this recommendation. Based on this framework we also highlight the parameters which should be considered for measuring degradation in quality.

Apart from the above standards, ITU-T recommendation J.144 and ITU-R BT.1683 discuss the testing procedure and specifications for objective perceptual video quality measurement techniques for digital cable television in the presence of a full reference where J.249 works with a reduced reference. J.144 and BT.1683 are recommendations for models which relate to compression errors. ITU-T recommendation J.247 and J.246 works with objective perceptual multimedia video quality measurement with a difference of full reference and reduced reference. There are a few projects which are underway in Video Quality Expert Group (VQEG) especially involving the Reduced Reference and No Reference TV (RRNR-TV) project which takes into account compression errors and transmission errors.

2.8 Issues with Subjective Data Acquisition

Subjective assessment introduces many issues due to the inherent property of being subjective[62]. These issues need to be resolved in order to effectively use such methods for meaningful evaluations [63]. The issues we face with subjective assessment of the video quality concern whether we are getting a random response or the true user perception. Random response is not good for our study and we need to know the level of randomness in the feedback. The following two points are important for our subjective data collection and analysis:

- How to capture true perception
- Randomness of feedback

For capturing the true perception we need to investigate the reasons people get bored [64]. The reason was identified as loss of interest as stated by Weinstein et al. For this reason there is a need to make an effort to make the experiments interesting. The other half of the problem is to check the randomness of the feedback. Random response means that volunteers are not assessing the actual content and providing false feedback.

2.9 Finding the most influential parameters

We also want to identify the most influential factors from the data we collect. This is one of the main objectives of this study as well.

- Influential parameters

- Optimal parameter combinations

To develop an effective model we need to identify the most influential parameter from each domain. Firstly, we need to identify the parameters which have a significant influence on user's QoE. This will establish the scientific basis for inclusion of parameters in the model. Secondly, an explosion in the state space for the modelling if there are many parameters per domain. Thus if there were 3 from each domain we would produce a state space of 3 x 3 x 3 = 27 states and thus it becomes nearly impossible to calibrate such a model. We would prefer to chose a single parameter per domain (the most influential one) to keep the state space manageable. Then there is a need to analyze the subjective feedback to identify the change in perception due to the change in parameter configuration. Multiple methods can be used for this purpose like regression, Morris method [65], Taguchi robust design and so many other statistical techniques. A literature review on this topic identify the methods available for this purpose and discuss the advantages and disadvantages of these methods has been published by [66]. Moreover, we were also interested in finding out the optimal solution from the set of available configurations.

In the Taguchi method, optimal solution implies that we are interested in determining the best or most suitable combinations of control factor levels under the influence of noise factors. The best or most suitable combinations are those which maximize the SNR under the influence of noise factors. The SNR are log functions of desired output characteristics. The experimental design ensures that all control factors are balanced and equally represented in the experiment and the number of experiments are kept to a minimum [67]. All this ensures that minimum resources are utilized for the experimentation process. This will enable us to configure the service for the best possible quality under the current conditions. These questions need to be answered by this study.

2.10 Research Questions

From the above discussion we were able to identify gaps in the existing knowledge which lead to the following research questions that will be addressed in this thesis:

1. Acquiring participants' true perception about video quality
   (a) Agreement from the group of participants about video quality
2. Finding parameters which affect video quality
   (a) Logical division of the end to end communication into suitable domains.
   (b) Finding out the influential parameters from each of the identified domains.

(c) Develop model for QoE across the three principal domains that we have identified on the basis of the most influential parameters from each domain

(d) Finding parameter combinations (configurations) which can promise consistent performance

3. Develop models that can predict video quality

2.11 Summary and Conclusion

It can be observed that most of the research work in the past has focused on a limited number of factors that have been identified and presented in this introduction. Traditionally QoE has been measured subjectively which has got its benefits; however, this process can be slow, expensive and laborious as well as being difficult to achieve for mainstream services such as IPTV. Hence it's required that objectively observable measures be used for predicting user experience. If these directly measurable parameters can be located and linked in an appropriate manner to user QoE then this provides a platform for the development of models that can be used to infer performance as well as possibly providing methodologies and tools for planning networks that deliver IPTV that meets customer needs. Although all the influencing factors do not have the same impact on QoE, attempts will be made to incorporate as many key influencing factors as possible into our model. Multimedia service providers such as IPTV can then adapt and improve this model for their specific applications.

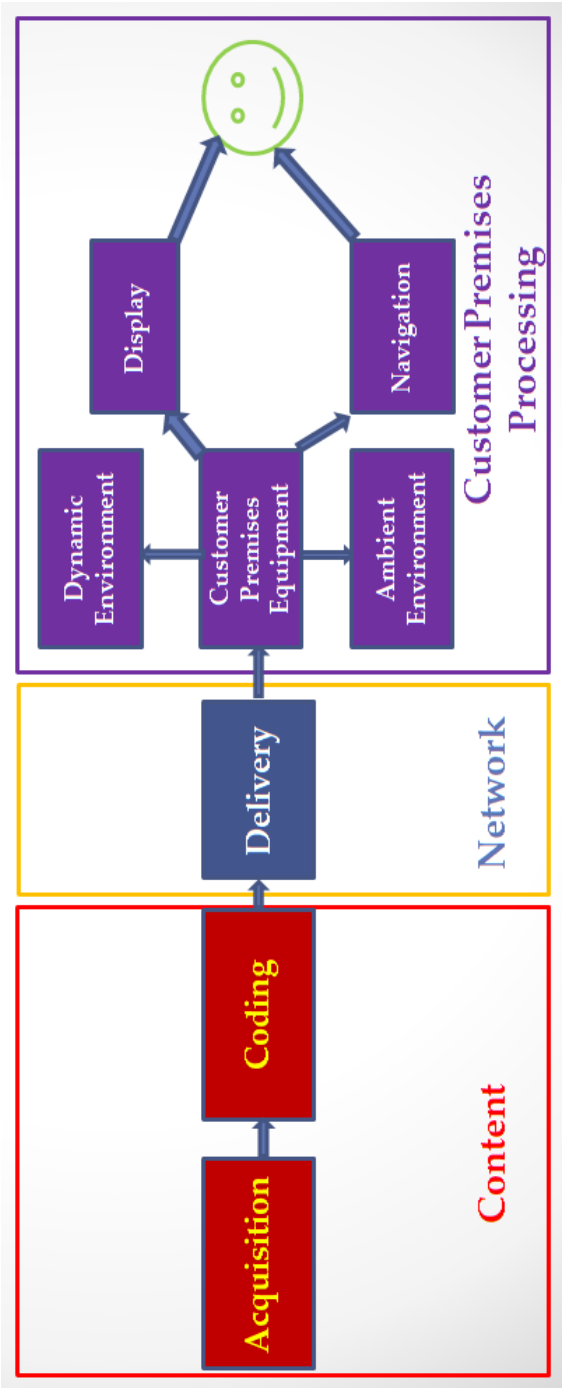

Figure 2.2: A framework highlighting factors within domains which affect IPTV service delivery

Chapter 3

Framework and Experimental Design

3.1 Chapter Overview

This chapter introduces the methodology based on standard scientific principles and adopted to answer the research questions underpinning this thesis that were identified in Section 2.10 above. Moreover, this chapter includes discussion on method selection. In some cases, more than one possible methodology is available and, after evaluation of the advantages and disadvantages of such methods the most feasible methods were selected for this research.

3.2 Framework

1. For capturing the true perception we employed multiple methods and techniques.
   (a) Taguchi method was used for design of experiment (DoE)
   (b) We used ITU P910 standard as a general guideline for conducting subjective testing
   (c) Fliess´ Kappa and correlation among raters was used for finding the true perception
2. For finding parameters which affect video quality we used two methods in tandem
   (a) Taguchi method was used for signal to noise ratio analysis
   (b) HoQ (House of Quality) method of QFD (Quality Function Deployment) was used for finding influential parameters
3. For developing model for QoE we used ordinal logistic regression

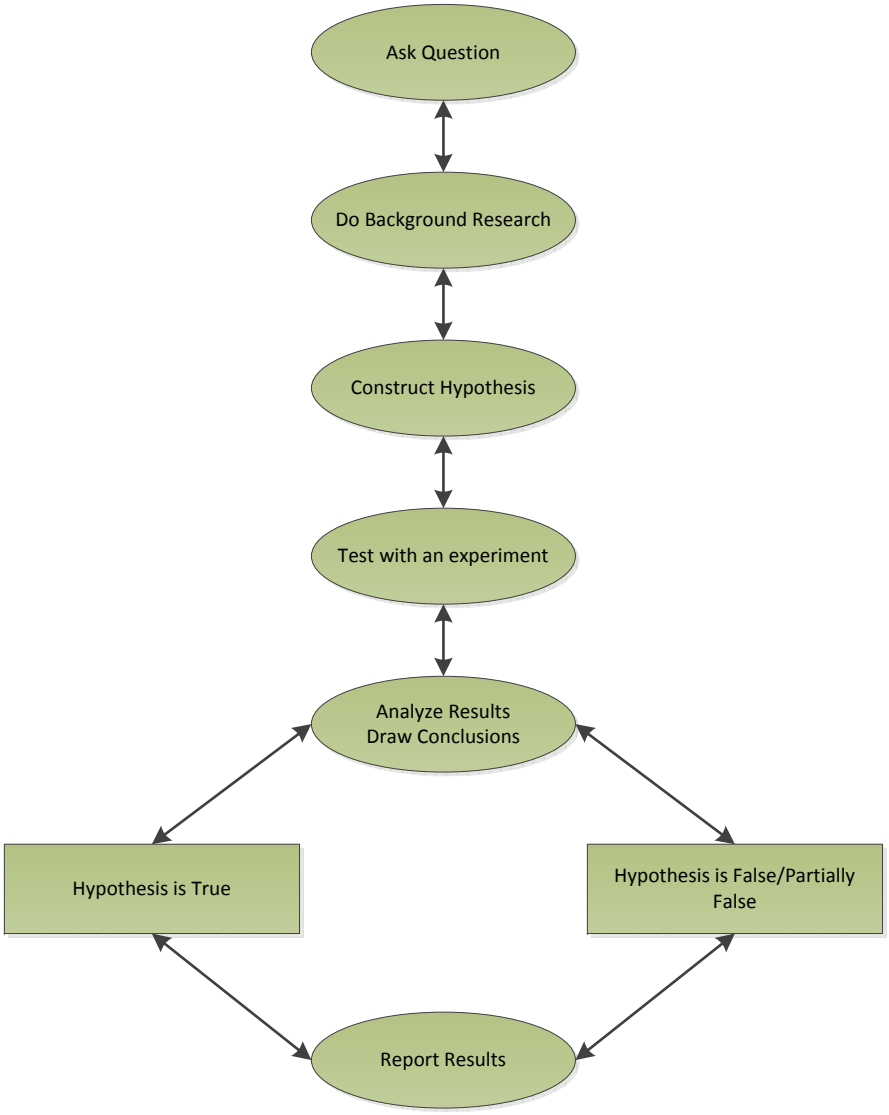

Figure 3.1: Experiment flow diagram

The framework we developed, for handling the above research questions, follows a basic scientific methodology, viz: Asking questions, conducting background research, constructing a hypothesis, testing using experiments, analyzing results and drawing conclusions. Wilson et al. [68][69][70]. Figure 3.1 shows the flow diagram for a general approach used for the scientific methodology. Figure 3.2 shows the framework that was followed for this research. The first task was to come up with an appropriate set of research questions. In the literature review we found a large set of parameters had been studied by researchers and we reviewed the various methods that they used to conduct their research. From our survey, we were able to identify gaps in current knowledge and from this we developed questions that could be used to close some of these gaps. We wanted to find reasons for video degradation. In other words, we wanted to know the effect of identified parameters on video quality. In addition, there was a need to predict video quality as this capability can enhance a service provider's revenue. This can be achieved by investigating the impact of changing these parameters on the perceived human perception of video quality by the users.

From the research questions it is obvious that we are interested in looking at video quality from an end to end perspective. As discussed in the Chapter 2 we found literature which either discussed the effect of a single parameter on quality or looked at a combination of parameters from a specific domain. There was a need to investigate larger set of parameters and to look at the combined effect of these parameters within and across domains. Keeping in view this observation we realized that we need to divide the end to end chain of communication into logical domains. In Hussain et al. [71] we divided the whole communication chain into three logical domains. These are: the content domain, the network domain and the customer premises processing domain as shown in Figure 2.2. We need to consider for each domain the parameters affecting video quality.

We selected the parameters which were identified in the literature as having some effect on video quality as discussed in Chapter 2. In most of the work that we came across in the existing literature usually it analysed the effect of a single parameter. This technique, although effective in identifying the individual isolated effect of each parameter, does not reflect the true impact of multiple configurations with multiple parameter combinations. We planned our research to tackle this gap and study the effect of a combination of parameters from each domain identified in Figure 2.2. Later, we also studied the combined effect of parameters across multiple domains to obtain a true end-to-end view of QoE for video. We were able to identify around 22 parameters/artefacts that potentially have an impact on video quality. These were across all three domains. To develop a model based on 22 parameters is a huge task. Such a model would be too complex and might not be implementable. It would be almost

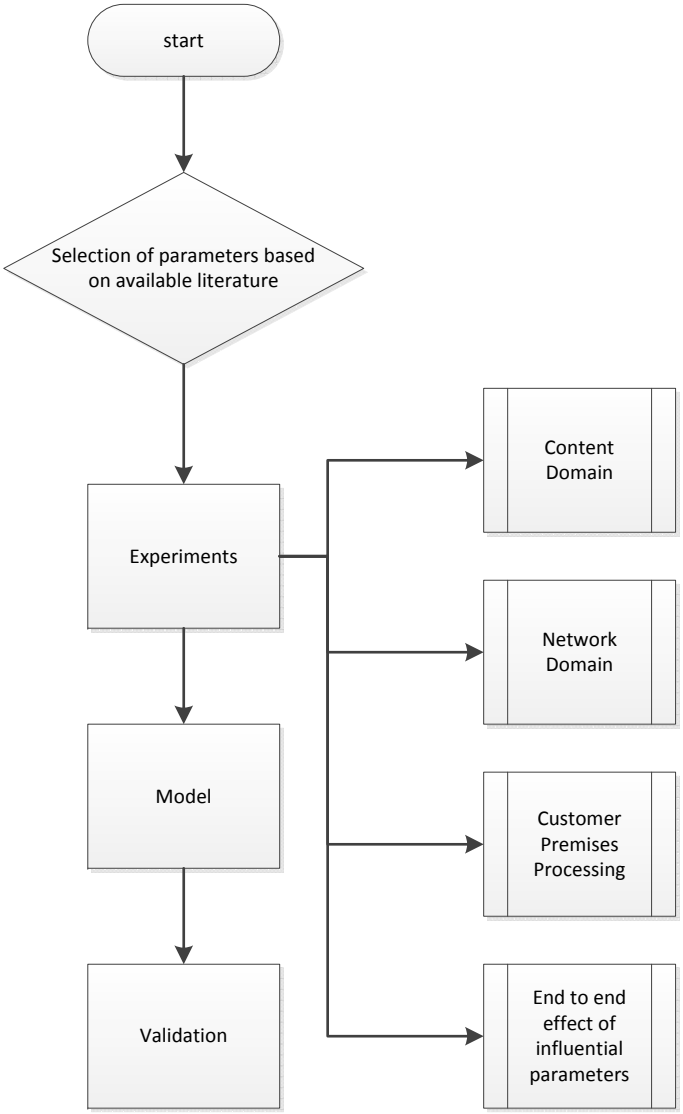

Figure 3.2: Methodology

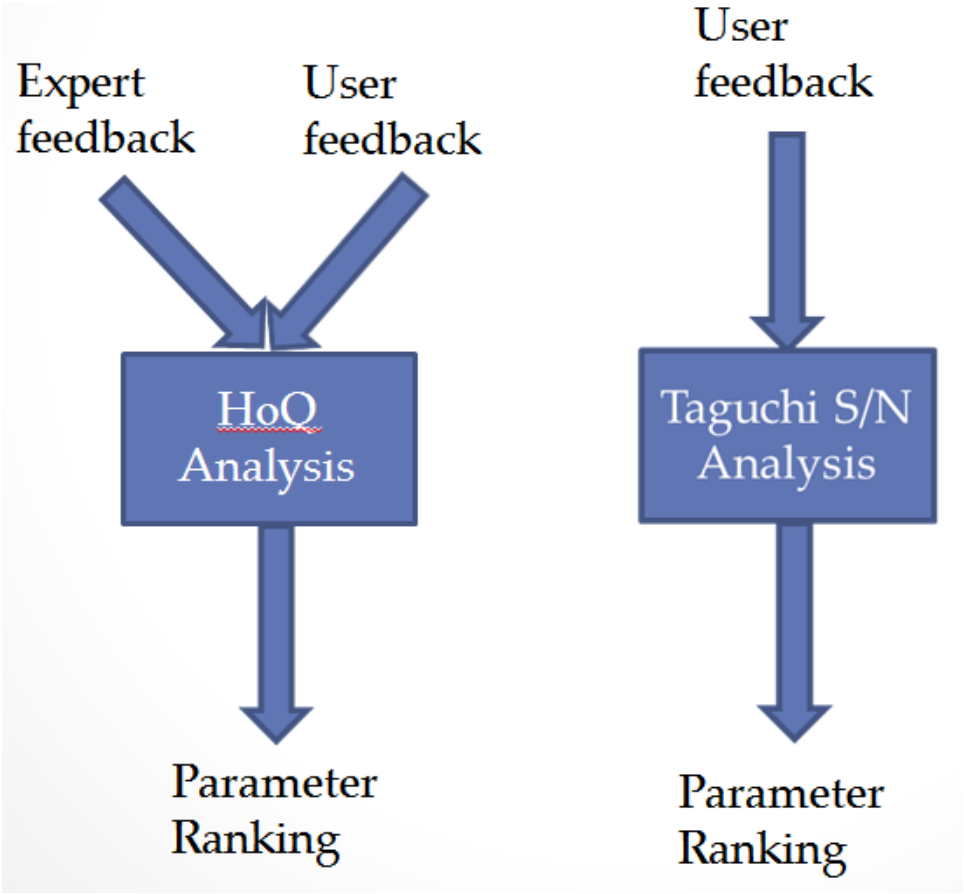

Figure 3.3: Comparison between Taguchi method and HoQ method

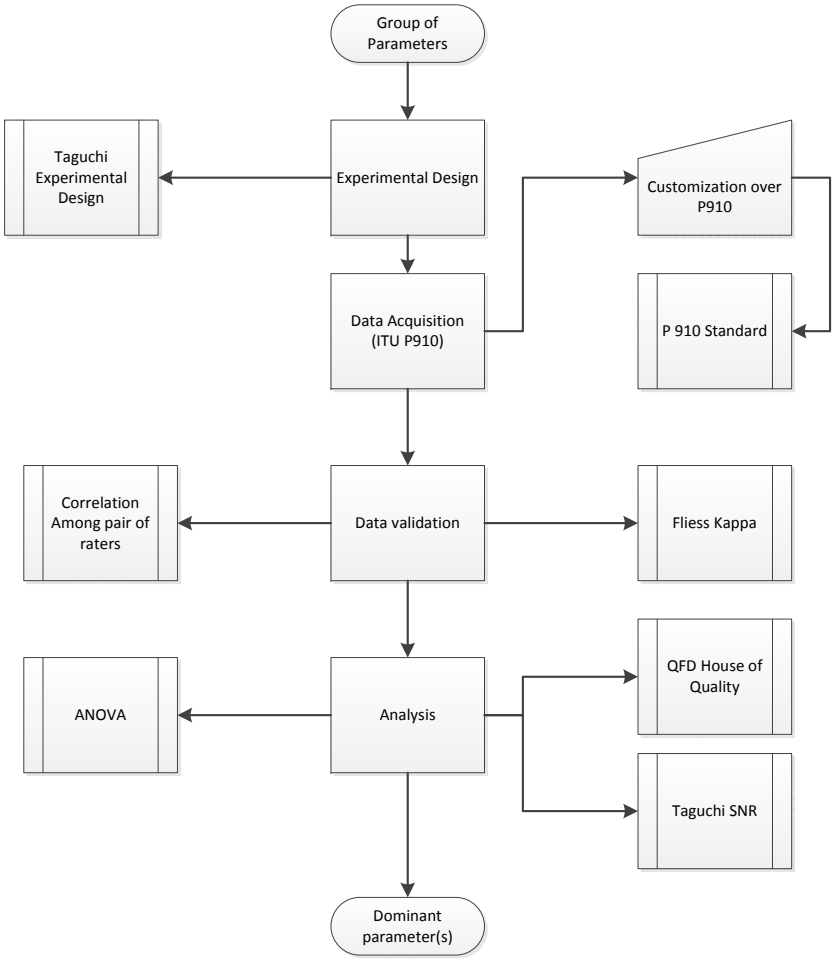

Figure 3.4: Experiment flow diagram

impossible to calibrate and perform a sensitivity analysis or a dimensioning exercise using such a complex model. Accordingly, we wish to reduce the number of parameters to a manageable number for our model. In an effort to reduce model complexity we need to limit it to just a few parameter values. We wish to investigate the effect of combinations of parameters from the various domains and to determine which parameters we can either be removed from the model altogether or possibly use a single default value.

3.3 Experiment

The second part of our framework involves conducting experiments to validate our hypothesis/question. We want to develop our experiments in such a way that they can provide us with insight into the relationship between human perception and video quality. There are two means of measuring human perception regarding video quality. Firstly we could directly ask volunteers questions which would provide us with some measure of human perception. However, this method is subjective and may introduce an expensive experimentation process in terms of time, effort, complexity and monetary value since large sample sizes of volunteers would be required in order to obtain useful results. The second method involves predicting human perception from objective parameters. This is a better method for real time prediction or evaluation of human perception if appropriate linkages can be determined. If objectively measurable parameters are available to service providers it would enable to monitor and manage the QoE of their customers. However, there still needs to be a benchmark or ground truth value for evaluating such an objective method and that needs to come from an appropriate calibration between the objective and subjective methods.

3.3.1 Limitations and Requirements

There arise three important requirements for our investigation. There is a need to introduce a method which could reduce the requirements of time, effort and funds involving any subjective method. Secondly, a method is required that could generate quality benchmark data. Thirdly, we need a method which could inform us about which parameters should be used for predicting human perception. The number of parameters should be as small as possible but they should be able to predict human perception as accurately as possible, in order to identify optimal parameter configurations that ensure quality.

Mitigating Psychological Effects

During the literature review we came across many problems which reflect undesired effects on observer's perception and his/her ability to reliably provide accurate feedback. The source of most of these undesired effects lies in human psychology. ITU documents (e.g. ITU-R BT.500-11) discuss in detail the methods used for subjective quality assessment. The Double Stimulus Continuous Quality Scale (DSCQS) method is widely accepted as an accurate test method with little sensitivity to contextual effects. The issue with double stimulus methods like DSCQS is that they provide only a single quality score for a specific video sequence which is 10 seconds long. To effectively relate that single score to an objective quality method generating a score in real time

is a problem in itself. Though double stimulus methods do have their fair share of disadvantages as far as implementation is concerned, we still decided to use a customized version of this method for our current work. Our focus here is to identify parameters having first order and second order effects on quality. Single stimulus methods give rise to memory effects according to [72] the observers have non-symmetrical memory in that they are quick to criticize degradation in video quality but slow to reward improvements. Using a single stimulus increases the variation in results due to contextual effects, boredom, and reaction times to quality changes. According to [73] memory and boredom do affect an observer's capability to accurately perceive quality. Work done by [74], Aldridge et al. [75] highlights the forgiveness effect, recency effects and the boredom effect. Whereas Wakeman et al. [76] discussed effects concerning task difficulty on subject image quality. The forgiveness effect is where observers tend to forgive quality degradation once it's followed by a significant length of better quality video play. The recency effect occurs when observers are asked to assign a quality score at the end of a video and they would be influenced by the recent quality rather than the average quality or the maximum quality during playback. The boredom effect occurs when observers become disconnected from the content shown during playback. Apart from the above mentioned effects, personal preferences affect an observer's ability to assess quality. To effectively measure the observer's perception of quality it's necessary to mitigate, or reduce boredom, recency and forgiveness effects to a minimal level so that these undesired psychological effects are controlled by carefully designing any experiments.

Thus, we need to consider all the above mentioned requirements before we design our experiments. We decided to employ appropriate experimental design methods.

3.3.2 Experimental Design

The primary goal in a scientific research project is to show the statistical significance of an effect that a particular factor exerts on the dependent variable. In industry, the primary focus of experimental design is to identify unbiased main effect estimates with the minimum possible number of observations.

The purpose for employing experimental design is to achieve the following primary benefits:

- Reduce time to design/develop new products & processes
- Improve performance of existing processes
- Improve reliability and performance of products
- Achieve product & process robustness

- Perform evaluation of materials, design alternatives, setting component & system tolerances, etc.

The benefit of employing experimental design would suffice the needs described in Section 3.3. A methodology for designing experiments was proposed by Fisher in [77]. His earlier work was concentrated on agricultural applications. Later, his work was adopted generally in physical and social sciences. A special focus at that time was industrial applications. According to fisher's philosophy, an experimental design method should provide the following functionality:

Randomization: Randomization ensures that the experiment is a rigorous, "true" experiment.

Replication: Measurements are subject to variation and uncertainty. Measurements are repeated to help identify variation. This ensures true effects of treatments. This increases the reliability and validity of the experiment.

Blocking: Arrangement of experimental units in to similar groups. This reduces known but irrelevant sources of variation. Increase precision of estimation of the source of variation.

Orthogonality: Orthogonality concerns the forms of comparison (contrasts) that can be legitimately and efficiently carried out. Contrasts can be represented by vectors and sets of orthogonal contrasts are uncorrelated and independently distributed if the data are normal. Because of this independence, each orthogonal treatment provides different information to the others.

Factorial experiments: Use of factorial experiments instead of the one-factor-at-a-time method. These are efficient at evaluating the effects and possible interactions of several factors (independent variables). All enable the reduction of the required time for conducting an experiment.

There are two general issues which are addressed by experimental design

- How to design an optimal experiment.
- How to analyze the results of an experiment.

A number of methods are used to identify the optimum settings for the different factors that affect the production process.

Fractional Factorial Designs: In certain experiments we consider the factors having only two levels. Using full factorial design the number of experiment runs

will increase exponentially. Fractional factorial designs sort out this issue. Interaction effects are ignored in fractional factorial designs. This is done with the assumption that the dependent variable exhibits linear behaviour. Also higher order interactions are ignored.

Maximally un-confounded and minimum aberration Designs: 2^{k-p} fractional factorial designs are often used in industrial experimentation because of the economy of data collection that they provide. The 2^{k-p} maximally unconfounded and minimum aberration designs techniques will successively select which higher-order interactions to use. Things to consider in designing any 2^{k-p} fractional factorial experiment include the number of factors to be investigated, the number of experimental runs, and whether there will be blocks of experimental runs.

Box-Behnken and mixed level factorial Designs: Standard designs with 2 and 3 level factors can be generated. Generally its a combination of the procedures described in the context of 2^{k-p} and 3^{k-p} designs. These designs are efficient but they are not necessarily orthogonal with respect to all main effects. If ANOVA is used for analysis there is no need for orthogonality of the design.

Central composite and non-factorial Designs: The 2^{k-p} and 3^{k-p} designs all require that the levels of the factors are set at certain levels. In many instances, such designs are not feasible as some factor combinations are technically not feasible. For reasons related to efficiency it is often desirable to explore the experimental region of interest at particular points that cannot be represented by a factorial design.

Latin square Designs: Latin square designs are used when the factors of interest have more than two levels and it is known that there are no interactions between factors.

Taguchi Method: Taguchi design methods are set apart from traditional quality control procedures and industrial experimentation in various respects [78]. Of particular importance are:

- The concept of quality loss functions
- The use of Signal-to-Noise Ratios (SNR)
- The use of orthogonal arrays

We shall discuss in detail these important attributes of Taguchi design in Section 3.3.2.

Discussion of the Taguchi Method

We were interested in the capabilities of Taguchi design and we wanted to investigate the pros and cons of this method. We found out that, generally, Taguchi design is being widely used in industry, although within the statistician community his method is often quite debatable. The discussion centres on the claims made by Taguchi and statisticians generally not agreeing to some of these claims.

Orthogonal Arrays: The purpose of design is to conduct experiments in a way that we could investigate the effect of variables separately from each other. When the effect of variables is completely separate from one and other they are orthogonal to each other. Taguchi claimed that his method utilizes orthogonal arrays and ensures that the variable effects can be analyzed separately. The general perception in the industry is that this capability is unique to Taguchi design, on the contrary, all the traditional design methods are orthogonal. In the literature such designs are known as "saturated designs". Taguchi used the term L experimental design arrays. These are actually standard 2 and 3 level factorial designs with varying degrees of saturation. Plackett-Burman, fractional factorial and Latin square are examples of similar designs. Statistically there is not much of a difference between Taguchi design and standard designs and hence can be used interchangeably.

Noise factors: Taguchi introduced terminology of control and noise factors. Control factors are those factors which could be controlled for the experiment whereas noise factors are uncontrollable factors. The Taguchi method considers a known variable/factor as noise if, for a certain design, you don't want to control a certain variable for running the design. In the statistics literature, noise refers to unknown effects. As the variable is unknown it can't be controlled . Randomization of the experimentation process ensures that effects of these unknown variables will not mix with the effect of one of the control factors. There is a need to understand the difference between standard definition of noise in statistics and Taguchi's explanation of noise. They are two different issues. Though in industry it's important to consider few variables as Taguchi noise variables and then run the design to find the optimum variable configurations.

Design Structure: Taguchi introduced terms of inner array and outer array. These two designs are, in actuality, traditional factorial designs. Statisticians argue that generally this technique will introduce inefficiency. Such design will require more runs than necessary. The gains of such a design are to analyze the effects of control factors (process factors) under the limitation of noise factors. The same purpose could be achieved by considering all the variables as equal and analyzing the effects. Once we don't want to control one variable, so that it is

contributing to the overall noise of the process, we can use analysis of design to identify optimum control variable configurations achievable under noise variable limitations and identify the variation in process output.

Robust Design: A design is said to be robust if it can handle "damage" and produce useful results. The damage could be of two types

- One or more experiments in the design fail to produce measurable results
- There were shifts, changes or omissions with respect to the level settings of the variables in the design.

Robustness in Taguchi design has got nothing to do with the above mentioned two types of robustness. Taguchi design will provide results which will identify optimum process settings. These process settings will ensure product output whose properties will remain at or near optimum even under the influence of all the noise and process variables. In traditional design the same can be analyzed by using Box-Meyers analysis.

The above discussion signify the fact that Taguchi designs were not a novel innovation by Dr. Taguchi. He introduced these already tested and established design techniques with new nomenclature. His work was admirable as he was one of the pioneers who introduced DoE in the industrial world and was able to showcase the strength of DoE in improving quality. His ideas were not very well received in the statistical world as these ideas were already known and some research was already done in identifying the problems of DoE and certain enhancements were already been proposed.

Pros and Cons of Taguchi Method

The following are the advantages of using Taguchi designs

1. Using loss function for process optimization
2. Introduced improvement in finding the optimum processes settings under the limitation of known but uncontrolled variation
3. Being able to evaluate the impact of noise variables
4. Introduces the need to develop an effective response variable
5. Taguchi showed that generally the output variable Y is based less on the mean and more on some measure of its variation
6. Introduced concept of separation between sources of variation that we can control from sources of variation against which the process or product must be robust.

The following are the disadvantages of using Taguchi designs

1. Inefficient experiment design as it introduces more experiment runs than required
2. Requires the investigator to possess in-depth knowledge of the subject matter, which denies an opportunity to an unknown significant variable to affect the analysis. (This can be fixed by ensuring randomization of the experimentation process)

Advantage of using a statistical process control is that it will help eliminate defects after manufacture. In the case of Taguchi design maximizing quality and minimizing loss ensures that defects will be eliminated during the manufacturing process. Optimization can be done by using techniques like factorial design, central composite, Box-Behnken, etc. Taguchi design is simple but provides comparable results with most other sophisticated design methods. Taguchi designs are easy to implement if the experimentation team comprises of domain experts whereas other methods require learning and a larger sample to learn to provide worthwhile output. Taguchi design like most other statistical methods is based on a sound scientific mechanism which helps to evaluate and implement improvements in products and services. Taguchi design is a robust design. We want to employ a method that is:

- Robust in nature
- Provides analysis techniques which could serve the purpose of ranking parameters
- Identify the most influential and least influential parameters
- Provide insight into optimum configuration for end to end video communication
- This exercise will help us in developing a model for end to end video communication

Taguchi Design of Experiment

In the present work, the Taguchi design of experiment approach was implemented to study the effects of multiple compression parameters which influences the users' QoE. In order to find the relationship between parameters within the compression domain and overall quality, there was a need to identify the most influential parameters. At the same time it's important to identify the least effective parameters. A video quality model can then be developed based on the most influential parameters and either using default values for the least effective parameters or neglecting them altogether. Another reason to employ the Taguchi design was to reduce the time required to complete the subjective experiment. If an experiment was conducted for 4 parameters each having 3

levels using standard factorial design it would have taken around 81 sets of conditions i.e. 3^4. We considered the orthogonal array proposed by Taguchi. These are equivalent to fractional factorial designs and Taguchi proposed several of these orthogonal arrays depending upon the number of factors and their levels. For the above mentioned example experiment i.e. for 4 parameters and each parameter having 3 levels Taguchi proposed an L9 array. Details regarding the standard orthogonal arrays are discussed in many books covering robust designs or DoE in general [79]. This L9 array requires only 9 conditions to be tested. This reduces the number of conditions to a significantly smaller number and makes the experiment more manageable and cost effective. By employing the Taguchi design we studied the combinations of compression parameters (control factors) and user feedback in terms of Mean Opinion Scores (MOS) under the influence of several noise factors. Control factors are factors that can be configured in a controlled environment such as a test-bed, whereas noise factors are factors which are known but we don't want to control them for the purpose of experiment. Taguchi implies that a process is consistent in performance if it exhibits insensitivity to the influence of the noise factors. Orthogonal arrays are used to come up with different combinations of control factors. Orthogonal arrays represent the smallest possible matrix of combinations of the control parameters. This technique is good for minimizing repetitions, cost and time. In contrast to factorial design, Taguchi experiment design studies the effects of multiple parameter variations simultaneously. An orthogonal array is selected on the basis of the number of parameters and levels per parameter. After setting up the data according to Taguchi recommended arrays, test runs were performed randomly. The signal to noise ratio was used to analyze results to determine the influence of individual factors and the order of each effect.

Recommendations for Successfully Employing Taguchi Method

In order to use Taguchi method the following steps would typically be followed:

- Flow chart the process
- Identify potential sources of variation
- Categorize sources of variation into "controllable" and "not considered for control"
- Determine what the "Y" metric should be
- Validate the measurement system
- Identify factors for the DOE
- Determine how to minimize variation on factors not in the DOE

Noise array or outer array should never include a variable that could be controlled in normal use.

Current use of Taguchi DoE

Taguchi technique has been widely used for optimization and identification of effect of parameters. In the field of polymer plastic technology an attempt was made by [80] to investigate the effect of molding variables on sink marks of plastic injection molded parts by deploying the Taguchi technique. The aim of this research was to reduce the impact of sink marks in injection molded parts and highlight the significance of each parameter on a sink mark. In the field of material processing technology, there is a need to maximize the wear resistance of steel [81]. For this purpose the Taguchi method was utilized for optimization of cryogenic treatment to maximize the wear resistance. Authors were able to identify optimum levels of parameters on the basis of maximum S/N ratio. In the field of assembly planning robust back-propagation neural network engines were used with Taguchi method for finding optimal solutions [82]. In industrial engineering, deployment of Taguchi method for optimization and identification of influential parameter is a normal practice. These domains are wide and varied in application but generally all these disciplines require the generating of output quality. Taguchi method has helped in ensuring output quality. For our study, these capabilities of the Taguchi method can be utilized for identifying influential parameters for an IPTV service quality.

Customization

There was specific customization for each experiment. These customizations are explained in detail within the relevant chapters of this thesis.

1. Content domain experiment was done on the test bed while keeping ideal configurations for network domain parameters and ideal configurations for CPP domain parameters. We only changed the content domain parameter configurations as per the L9 array.

2. Network domain experiment was done on the test bed while keeping ideal configurations for content domain parameters and ideal configurations for CPP domain parameters. We only changed the network domain parameter configurations as per the L9 array.

3. CPP domain experiment was done on the test bed while keeping ideal configurations for content domain parameters and ideal configurations for network domain parameters. The test bed included various displays and processing power. These configurations were done as per the L9 array.

4. End to end experiment was done and all the parameters included were changed as per L18 array. We developed additional video for this experiment as per the L18 configuration for the content domain parameters. Moreover, the network configuration as well as the CPP parameter configurations were also handled as per the L18 array.

3.3.3 Data Acquisition

Subjective and objective data was acquired for this experiment. The following sections discuss in detail the process of data acquisition.

Subjective Data Acquisition

As we identified in Section 2.8 of the literature review that there were certain issues that need to be handled before we can use subjective feedback [83]. We employed the following techniques for resolving these anomalies. In order to resolve the lack of interest of volunteers, we utilized the video content which was extracted from the latest movies and reduced the size of all clips to 10 sec. This ensured volunteers' interest in the experimentation. An effort was made to introduce multiple genres of videos to keep most of the volunteers interested in our experiment. This technique seemed to work as the informal discussion at the end of the experiment identified that most volunteers were keen on continuing with the experiment till the end. In the same section we discussed the need to assess the level of randomness of the feedback. We investigated the work done in the field of subjective assessment in the social science domain [84] [85]. Fliess´ Kappa has been regularly used for identifying inter-rater correlation or concordance among the group of raters. This capability enables the researcher to identify if the group was assessing the same content [86]. We made use of Fliess´ Kappa and were able to identify the level of concordance of our volunteer group for the content they were assessing. If the concordance was high we were confident that the feedback was not random. These two techniques helped us in capturing the true perception of the volunteers. The basic aim of conducting the subjective assessment was to set the base line for objective assessments.

For conducting the subjective assessments we generally followed the ITU standards document for subjective experiments [87]. Volunteer feedback was recorded using software that we developed. Software was developed in C# for handling video play as well as feedback capture. For analysis purposes, the feedback data was written to an Excel sheet from all the volunteers and for all the videos. It was developed using the VLC library for playing videos. It stored feedback from users on a .csv file on the server. Software was also used for randomization of the experimentation process. Feedback was in the form of Mean Opinion Scores (MOS). A Likert scale of 1-5, where 1 was

Table 3.1: HoQ Questionnaire

House of Quality Questionnaire		
Index	Requirements	MOS
1	Smooth Video (No Jitter)	
2	Video with No Blockiness (Not pixelated)	
3	Video with No blurring	
4	Synchronization of voice and video	
5	Meets acceptable quality (Overall)	
6	Time to load before playback	
7	Acceptable Video Quality for high motion (Camera panning or fast moving content)	
8	Max Video Quality for low motion (Head and shoulder)(Fixed camera)	
9	Acceptable Video Quality for Complex scenes (Many objects/lots of colours/Details)	
10	Acceptable Video Quality for simple scenes (Few objects/few colours/Not much detail)	
11	Continuity of service	

worst and 5 was excellent quality, was used for the MOS.

For the House of Quality (HoQ) method, the volunteer feedback was captured using a questionnaire as shown in Table 3.1. Each volunteer was asked to give their feedback using the Likert scale of 1-5 where 1 was deemed to be Unimportant and 5 was Very Important. They assessed each need or demand mentioned in questionnaire as per their perception.

Objective Data Acquisition

For capturing objective data we used different methods. Each of these methods is explained in detail with each experiment.

Participants/Subjects

We conducted two types of subjective assessments.

- Video quality assessment done using the Taguchi method
- House of Quality assessment

For both these experiments we requested volunteers from the university student population. They were from different nationalities, and culture. Each group comprised of 16 volunteers according to [88]. The age group was between 19 and 44 years. Gender distribution was ensured to be around 60% male and 40% females.

3.3.4 Data Validation

Before we use the data collected for analysis we need to ensure its validity. The following sections discuss the methods used for data validation.

Correlation Between Pairs of Volunteers

The first analysis that we wanted to perform was to validate the feedback we have recorded from the users/volunteers. The feedback/response from volunteers can only be considered valid if it is not random. Our aim is to capture perceived quality via feedback. All the observers were assessing the same media so that the scores should not be random. For this purpose, we would calculate the correlation between observers. Hence each observer's score will be checked against every other observer for each video. The results will provide us with a pairwise correlation. Higher correlations would suggest that the pair of volunteers assessed the same video and their perception about the video quality is comparable within that pair. Although higher correlations among pairs of observers would not be sufficient to say that there was "considerable agreement among observers generally as a group".

Fliess´ Kappa

We are also interested in finding out the level of agreement within the group. For a video based experiment we wanted to be sure that the responses are perception based feedback. Pairwise correlation was not enough to provide the complete picture. Taken from the psychological literature, a statistical method called Fliess´ Kappa has been used for finding group agreement in our exercise. This group-based agreement is called "concordance" [89]. It enables us to calculate concordance among any number of observers for a fixed number of items. Fliess´ Kappa i.e. k can be calculated by using Equation 3.1.

$$k = \frac{\bar{P} - \bar{P}_e}{1 - \bar{P}_e} \tag{3.1}$$

The factor $1 - \bar{P}_e$ gives the degree of agreement that is attainable above chance, and the factor $\bar{P} - \bar{P}_e$ gives the degree of agreement actually achieved above chance. Where k=1 is absolute agreement and k=0 means no agreement at all. Value of k help select a respective level of agreement and for interpretation of results as proposed by [90]. Kappa can be affected by prevalence but still provides more information than the raw proportion of agreement.

3.3.5 Ranking of Parameters

We employed two methods for ranking of parameters. SNR is the standard method prescribed by Taguchi for analyzing data acquired by using Taguchi Design. The second method used for ranking of parameters was the HoQ matrix from the QFD method. These two methods are governed by two completely different methodologies for ranking of parameters. Figure 3.3 shows the different inputs required by each method for generating the required ranks. For the Taguchi method we only requested volunteers for their feedback in response to the videos shown to them. For the HoQ matrix we required feedback from volunteers regarding the importance or priority of requirements/user demands. At the same time domain experts were required to provide their subjective feedback with reference to technical importance or relevance for each requirement/user demand. In Section 3.3.5 and 3.3.5 we shall explain both methods in more detail.

SNR Analysis

The signal to noise ratio analysis proposed by Taguchi is considered to be one of the innovations introduced by him. Generally, customer satisfaction is high for a product or a service performing on target whereas a product or service that fails within a tolerance range causes a quality loss. This quality loss is generally represented by a quadratic loss function. The product or service quality can be enhanced by minimizing the loss function. Taguchi used a logarithmic function which was the ratio of mean performance to variation in mean performance due to known but uncontrolled variables. This method enables the researcher to decide the best values or levels for the control factors. Hence, optimum performance can be delivered under the influence of noise factors. SNR ratio is the main measure used for optimization of processes and products. SNR represents sensitivity to variability and is a required measure for optimizing the robustness of a process or product [91].

Typically a loss function is used for parameter estimation and in optimal control the loss is considered as the penalty for failing to achieve a target value. Taguchi proposed SNR which provides a loss function and ensures that this design is robust. SNR was used for optimization purposes as well. For the signal to noise ratio, Taguchi considered average quality characteristics, standard deviation and a target quality value, where the standard deviation is caused by noise variables. There are three variations available for SNR calculations that are nominally the best if you want to move the average to a target. If the aim of the experiment is to reduce the average then Smaller is Better can be used. For maximizing the average, then Larger the Better can be used. For our experiments we employed the Larger the Better formula for calculating SNR using Equation 3.2. There is an effort to maximize the SNR and minimize the loss. Hence SNR is a predictor of quality loss that highlights the effect of noise factors with reference

to the product sensitivity. Use of a logarithm improves the additivity of the function. For robustness, the aim is to minimize sensitivity to noise by identifying combinations of control factors that maximize the SNR.

$$SNR = -10\log(\frac{1}{n}\sum_{i}\frac{1}{y_i^2}) \tag{3.2}$$

House of Quality

Quality Function Deployment (QFD) is widely used in manufacturing and business domains. QFD empowers the manufacturer or business by providing them with a matrix to relate customer needs to design, development, engineering requirements and service deployment. QFD involves understanding customer requirements, thinking in terms of quality systems, human psychology and current domain knowledge. It allows a manufacturer to maximize the positive quality that adds value to the product and provides a comprehensive quality system for providing customer satisfaction.

QFD is a platform which provides a systematic approach of employing comprehensive development processes for:

- Understanding "true" customer needs from the customer's perspective
- Prioritizing needs as per user requirements
- Help to decide which features to implement first to increase customer satisfaction
- Determining what level of performance to deliver
- Providing a matrix to relate customer needs with design, development and engineering requirements and service deployment

QFD is a comprehensive quality system that systematically focusses on production or service delivery with reference to customer satisfaction. It ensures that all the different groups working within the same organization work in collaboration to achieve the common goal of producing quality products. This is done by considering explicit and implicit requirements, identifying quality improvement areas and business opportunities. QFD employs a HoQ matrix to align customer input with expert feedback to identify organizational priorities for delivering quality products or services. This matrix which is at the core of QFD is the basic design tool. Where QFD is used for focusing skills from within an organization, House of Quality matrix is used for relating customer needs with design and delivery of products and services. For our research we are interested in finding relationships between parameters affecting quality and customer satisfaction in relation to video quality. We want to use House of Quality to

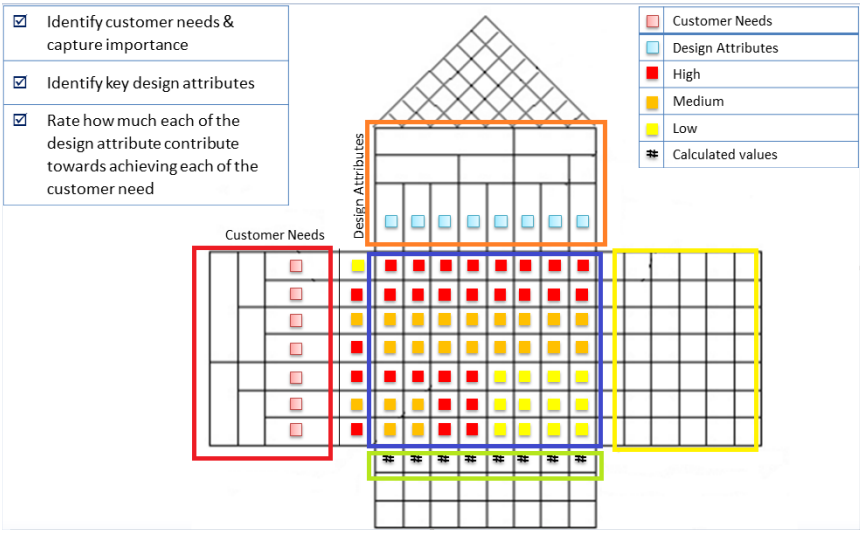

Figure 3.5: House of Quality Matrix

provide us with a mechanism to find the most influential parameter from each domain. Figure 3.5 shows the House of Quality implementation. There are "how" and "what" in each House of Quality matrix. Each subsequent House of Quality implementation use "how" of previous as "what" of current. Usually organizations deploy House of Quality in a cascaded manner. This enables them to use the customer input until the delivery phase.

House of Quality is a matrix which resembles a house. We used a reduced version of House of Quality matrix as we were not interested in the comparison between competitors. Figure 3.5 shows the major components of the House of Quality where each component is highlighted using different colour rectangles. For ease of understanding we can divide the House of Quality matrix into 6 smaller parts. The Top of the matrix describes the "hows" it is shown by using an orange coloured rectangle. The left most part of HoQ, housed in a red coloured rectangle, and this is the "whats" part. This part is used for acquiring user feedback regarding their priorities of needs/requirements. The central section, housed by a blue rectangle, are the calculated weights as per priorities given by the users and priority given by the experts. The bottom part, shown within a green box, contains the ranks as the sum of weights. Apart from these 4 there are 2 more parts. The yellow rectangle shows the comparison of our product or service with other products and services available in the market. The top triangle shows the correlation matrix which contains correlations between "hows". Either, it's positive or negative correlation, or its weak or strong correlation. Usually correlations are ignored for calculating rankings as they generally affect the priorities for the "Hows". Usually House of Quality is used in a cascaded style so hows from the first HoQ matrix becomes whats for the next HoQ matrix. This enables an organization to reflect customer

demand into each phase of product manufacturing or service delivery.

Statistical Significance Analysis

3.4 Model for MOS Prediction

Our aim was to develop a model to predict video MOS or classification of MOS into known groups. For this we looked around for available statistical methods. The following discussion highlights the comparisons we made and the final method we selected for modeling.

3.4.1 Predictive Models

Model development for prediction purposes involves statistical methods which enable us to predict a future event. Usually for such modeling we either work in the area of statistics or data mining for forecasting trends and probabilities. We have a set of known values, after analyzing this set of known values we want to predict the unknown data or unforeseen event. This model could be a simple linear equation, a complex regression analysis, artificial intelligence based technique, decision tree or a rule base method. We were interested in regression analysis area and discriminant analysis for prediction and classification purposes.

3.4.2 Regression Model

By a regression model we can predict variations in a dependent variable from the independent variables or variables [92] [93]. In a regression model a coefficient for each independent variable is calculated. An equation comprising of all the independent variables and a constant term enable us to predict the dependent variables variations. Simple regression makes use of least square regression for calculating best fit values for the slope and intercept. An effort is made to minimize the error between observed and predicted values. More complex models, like logistic regression and discriminant analysis, require complex calculations, but the underlying principle stays the same [94].

3.4.3 Logistic Regression

Logistic regression can be used for estimating the relationship between one or more independent variables and a single dependent variable. There are three known types of logistic regression. Binary logistic regression, ordinal logistic regression and nominal logistic regression. When the dependent variable is of type binary, i.e. it can take one of two categories, it can be modeled by using a binary logistic regression. If the dependent variable is categorical and comprises of more than two categories, and if

these categories are ordered, ordinal logistic regression can be used. For the case where there are more than two categories but these categories are not ordered we can use nominal logistic regression [95]. In logistic regression we calculate a p-value, which represents the probability of the response variable, and regression coefficients for each of the independent variables as well as the intercept. All these values are calculated using maximum likelihood estimation, in an iterative technique. Logistic regression can be described as the log odds of the ratio of P and Q. Where P stands for probability of success and Q stands for probability of failure.

3.4.4 Maximum Likelihood Estimations

Maximum likelihood estimation is a statistical estimation method that is designed to maximize the likelihood that the observed values of the dependent variable may be predicted from the independent variables. We estimate the likelihood of coefficient or the cutoff value as well as the coefficient for each independent variable. Estimating logistic regression parameters requires a maximizing of the log likelihood function. By changing coefficient values we find the value which maximizes the value of the likelihood function.

Our reason for selecting ordinal logistic regression was the type of data we had available from our experimentation. We had independent variables which were categorical. The output or dependent variable was also categorical. We have taken the mean of all the volunteer ratings but in essence this mean explains nothing. If 1 is worst and 5 is excellent then what is 3.5 on a MOS score? Hence we decided to keep the dependent variable which were ordered categories. We perform the conversion of MOS to categorical data. We explained the conversion in detail within the relevant section.

3.5 Discriminant Analysis

Discriminant analysis is another technique for building predictive models. It attempts to simultaneously analyze the difference between two or more discrete values of a response variable with respect to one or more predictor variables, which is similar to the goal in logistic regression. The method discovers the discriminating power of each predictor variable for determining different values in the response variable.

Discriminant Analysis Assumptions

All analysis of variance (ANOVA) assumptions are applicable to discriminant analysis.

- The predictor variables should not be highly correlated with each other, nor should a variable be a function of another variable. Otherwise, the matrix will

not have a unique discriminant solution. Discriminant analysis is highly sensitive to outliers.

- Class are determined on the basis of dependent variables which must be categorical.
- Unequal sample sizes are acceptable, but the sample size of the smallest discrete value needs to exceed the number of predictor variables.

Though we were interested in discriminant analysis for classification or prediction but the assumption that the independent variables should be continuous restricted us from deploying discriminant analysis.

3.5.1 Logistic vs. Discriminant

The purpose of discriminant analysis is to find a means to use the data to discriminate between classes. Therefore, the proper assessment of a discriminant procedure for a particular data set is not how well the data fit the assumptions, but how well the procedure works on a validation data set.

1. Discriminant analysis should only be used with continuous independent variables. Logistic regression can work with continuous as well as categorical independent variables.
2. Logistic regression assumptions are more relaxed then discriminant analysis. Discriminant analysis needs to satisfy the assumptions of multivariate normality and equality of co-variance.

3.6 Validation

In order to check or validate our model we acquired additional data for experiment 4 i.e. the experiment including influential parameters from all three domains. By using Taguchi design we were only required to test only 18 conditions. Instead of testing 18 conditions we tested 36 conditions. Used half of the data for model generation whereas the remaining 18 conditions were used for model validation. Validation is discussed in Chapter 8.

3.7 Ethics

For our research there was a need to interview volunteers and to conduct video testing experiments as well. For this we applied for low risk notification from the Massey University Ethics Committee which was approved. During our research, the code of ethical conduct for research involving human participants was implemented.

3.8 Summary and Conclusion

This chapter discussed the detailed methodology used to handle the research questions. We explained step by step the methods used, and presented reasons for why certain parallel methods were not preferred. The exact implementation details as per each step of methodology will be included in upcoming relevant chapters.

Chapter 4

Content Domain Influential Parameters

4.1 Introduction

In our research and analysis into the content domain, we have explored the effects of content domain parameters on perceptual quality or quality of experience (QoE). In particular, we identified content domain parameters that are configured to compress video to meet various requirements. Investigations were made to identify the order of influence of these parameters on QoE. These effects were introduced by varying parameter values between low, middle and high values. The Video on Demand (VoD) service of Internet protocol television (IPTV) was emulated within our testbed. The testbed that was developed is capable of emulating effects of compression artefacts. It has already been realized that compression parameter selection and optimization will affect output quality. We conducted subjective tests for video quality assessment. Observers were required to rate perceptual quality of each video. These subjective tests were carried out firstly, for the reason that it's been reported that psychological effects can be introduced due to the design and setup of experiments as well. We wanted to reduce these effects and collect true perceptive quality scores for the videos. Secondly, the number of parameters studied in earlier work was limited [96],[97], [98], [99] [100]. Many of the parametric models published in the literature use bit rate as a parameter. We wanted to investigate a number of parameters from the compression domain and observe the suitability of bit rate as the most influential modeling parameter. Our results indicate that there are notable relationships between different parameters and there is a cumulative effect on overall quality.

Table 4.1: Control Factors and Levels.

Control Factors and Levels				
Control Parameters	Labels	Level 1	Level 2	Level 3
Bit Rate	A	200	800	1200
B-Frames	B	0	2	4
Quantizer	C	12	26	51
Partition Decision	D	5	3	1

4.2 Design of Experiment for content domain

To effectively design the experiment we need to identify the parameters which will be included in this experiment.

4.2.1 Parameters

In Taguchi DoE method parameters considered for an experiment are of two types. Control parameters and noise parameters. Control parameters are those parameters which we want to control to find out their effect whereas noise factors are those factors which we do not want to control. So we will be looking at effect of control factors under the influence of noise factors. The following sections explains these two types of factors for content domain experiment.

Control Factors

The control factors which were analysed during this experiment were: bit rate, B-frames (bi-directionally predicted frames), quantization parameter (QP), and partition decision. Table 4.1 shows the control factor and levels for each parameter. Parameter B-frame specifies the maximum number of consecutive B-frames to be achieved. In H.264, B-frames allow significant PSNR gains. It speeds up the second pass and may also speed up a single pass encode if the adaptive B-frame decision is turned off. In that case, the optimal value for configuring B-frames within H.264 compression is usually no more than 1. Otherwise, high-motion scenes can suffer. With the adaptive B-frame decision turned on, higher values can be used for option B-frame. The Encoder rarely chooses to use more than 3 or 4 B-frames. We kept the adaptive B-frame option turned on and selected values of 0, 2 and 4 for the experiment. For the partition decision, the values used were 1, 3 and 5. The algorithm that is responsible for making the partition decision can be configured for best quality or fastest processing. Hence the trade-off between quality and speed is achieved by selecting a value between 1 and 5. A value of 1 is configured for the fastest partition decision and a value of 5 for ensuring the

best quality. In H.264, quantization is controlled by the QP, which ranges between 0 and 51. An equivalent quantizer step size can be calculated for each QP. Step size approximately doubles for every increase of 6 in QP.

Noise Factors

Noise factors considered for this experiment were motion, complexity and location (indoor or outdoor scene). Table 4.2 shows the combination of noise factors. Taguchi DoE requires that each condition i.e. a specific combination of all 4 control parameters should be repeated for all noise factors. We had 3 noise factors so we were looking at 23 combinations of noise. It was required that we repeat the experiment for each condition with 8 noise combinations. Two cases were ignored i.e. when the motion and complexity both were high and both were low, and location was not accounted for. We ignored location in both of these cases because once motion and complexity are both high or both are low; location will not affect the quality. Figure 4.1 shows the possible combinations after ignoring these two cases. HHI and HHO effectively became HHI whereas LLI and LLO became LLI.

Table 4.2: Noise Factors and Levels.

Noise Factors and Levels		
Noise Parameters	Level 1	Level 2
Motion	High	Low
Complexity	High	Low
Location	High	Low

4.3 Experimental Setup

Content domain experiment required a test bed and video sequences. The following sections talk about the test bed, video sequences, test layout, user task and any customization that was required for content domain experiment.

4.3.1 Test Bed

Computers used for this experiment were Intel core i5 CPU machines running at 3.60 GHz with 4 GB RAM. Running 64-bit Windows 7 enterprise with Service Pack 1 installed. These machines were using integrated Intel HD graphics. Monitors used were ViewSonic VS 13239 LED 1080p Full HD.

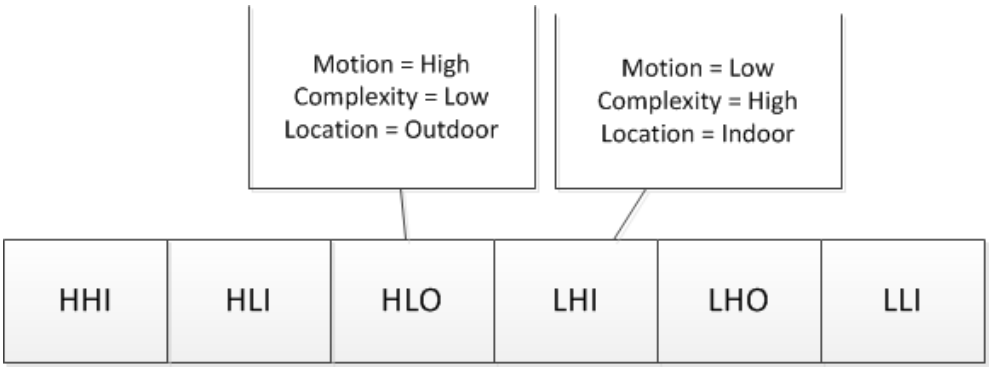

Figure 4.1: Combination of noise factors

Table 4.3: Array with control factors.

L9 Array with control factors				
Exp. No	A	B	C	D
1	200	0	12	5
2	200	2	26	3
3	200	4	51	1
4	800	0	26	1
5	800	2	51	5
6	800	4	12	3
7	1200	0	51	3
8	1200	2	12	1
9	1200	4	26	5

4.3.2 Selection of Test Sequences

We extracted benchmark HD quality videos from Blu-ray movies supporting H.264 compression and full HD resolution of 1920 * 1080. The content was selected from different movies encompassing many different genres. The degraded videos were created from these benchmark videos by configuring H.264 compression with different parameter settings. Altogether, 54 clips were used for the experiment and there were 6 categories of content type. For each condition as per Table 4.3, every observer was shown a group of 6 videos. This was for fulfilling the requirement of Taguchi DoE so that the experiment is repeated for each noise factor.

4.3.3 Test Layout

After selecting an orthogonal array L9 containing 4 columns, one for each of 4 control factors, the motion, complexity and location were considered as noise factors. Table 4.3 shows the L9 array with control factors. We randomized the whole experiment

process so that all the videos were presented to the viewers randomly. A group of 16 observers volunteered to participate in the experiment. They were screened to confirm that they had no prior experience in video compression or production. Each observer was shown 54 clips where each clip was 10 seconds long. The hardware playing these videos was set up in such a way that the volunteers were not required to move from their seat. The distance and height from the screens were adjusted accordingly. Larger screens were placed further away from the viewers than smaller screens. For the test setup, the [87] recommendations were followed closely. Undistorted/benchmark video was played followed by 3 seconds of delay and then the distorted video was played. Approximately 5 seconds were given for user feedback selection. Feedback was given by assigning a number between 1 and 5 where 1 represented the lowest quality and 5 being the highest quality. The users were required to select a level of quality, based on their perception, between excellent, good, fair, poor and bad. As per their selection, a relative score was recorded. Approximately 30 seconds was required for completing one assessment. After 6 assessments a break of 3 minutes was given. This completed one segment. After 6 segments, instead of 3 minutes, a break of 5 minutes was allowed. The whole experiment was completed within approximately 45 minutes. Whereas initially 5 minutes were spent in explaining the experiment procedure whilst later on 10 minutes were consumed in filling in the feedback form.

4.3.4 User Task

Observers were required to assess 54 videos in approximately 45 minutes. Their task was to assess each degraded video in comparison to the original/undistorted video. They were only briefed about the experimental process and were not trained about artefacts and the 5 level Likert scale used for this experiment. They were asked to enter their feedback about the perceived quality. At the end of the experiment they were asked to fill in a feedback form, containing questions about the experimental process and artefacts presented in the experiment. The findings will be discussed in Section 4.4.

4.3.5 Customization

From earlier work, [74] [75] [76] we realized that to record the true perception of an observer, we needed to reduce the effects of boredom as well as the memory effect and recency effects. In each assessment, playing a benchmark/undistorted video ensured that for each assessment the user had a recent reference in mind to compare the quality and thus minimized the memory effect on users. The experiment was designed to present conditions randomly hence users had no perception or expectation of the level of quality for the next assessment based on the current assessment, [101]. This reduced

the effect of either step-wise increasing or step-wise decreasing quality levels. Without randomization the user would be scoring on expectations of quality rather than a true perception of quality. A few reasons that could have contributed towards introducing boredom in a user are: length of the experiment, length of the selected video and the content displayed. In order to reduce boredom we made an effort to keep the time required for the experiment to less than an hour. In addition, we made an effort to keep the experiment interesting and exciting so that the boredom effect does not come into play. Forgiveness effect was reduced by keeping the video length to 10 seconds and playing the content randomly. Seferidis introduced this term of "Forgiveness effect" which related to the phenomena where user tend to forgive impaired video when it is followed by a substantial period of high quality or unimpaired video [74] [102]. This ensured that the observers were only rating the immediate quality rather than suffering from the forgiveness effect. For this experiment we denied the user any control over the flow of the experiment and they were not given any functionality to select the genre of video.

4.4 Results and Analysis

There was a requirement to establish the fact that the MOS scores acquired had a general consensus among the observers. Also to identify the most influential parameters and to check the statistical significance of the results. The following methods were used to establish these facts:

Pairwise Correlation

To verify user feedback was based on perceived quality rather than randomly assigned values we calculated the correlation between each pair of users.

Fleiss´ Kappa

This technique is frequently utilized in psychological subjective experiments for calculating concordance among all the experiment participants. Hence we decided to use this to verify the fact that the users were given feedback as per the quality of video shown to them.

Signal to Noise Ratio Analysis

This technique was used to identify the most influential parameters having an effect on quality.

4.4.1 Pairwise Correlation

The first analysis that was performed was to validate the feedback we recorded from users. The feedback can only be considered valid if it is not random. Our aim was

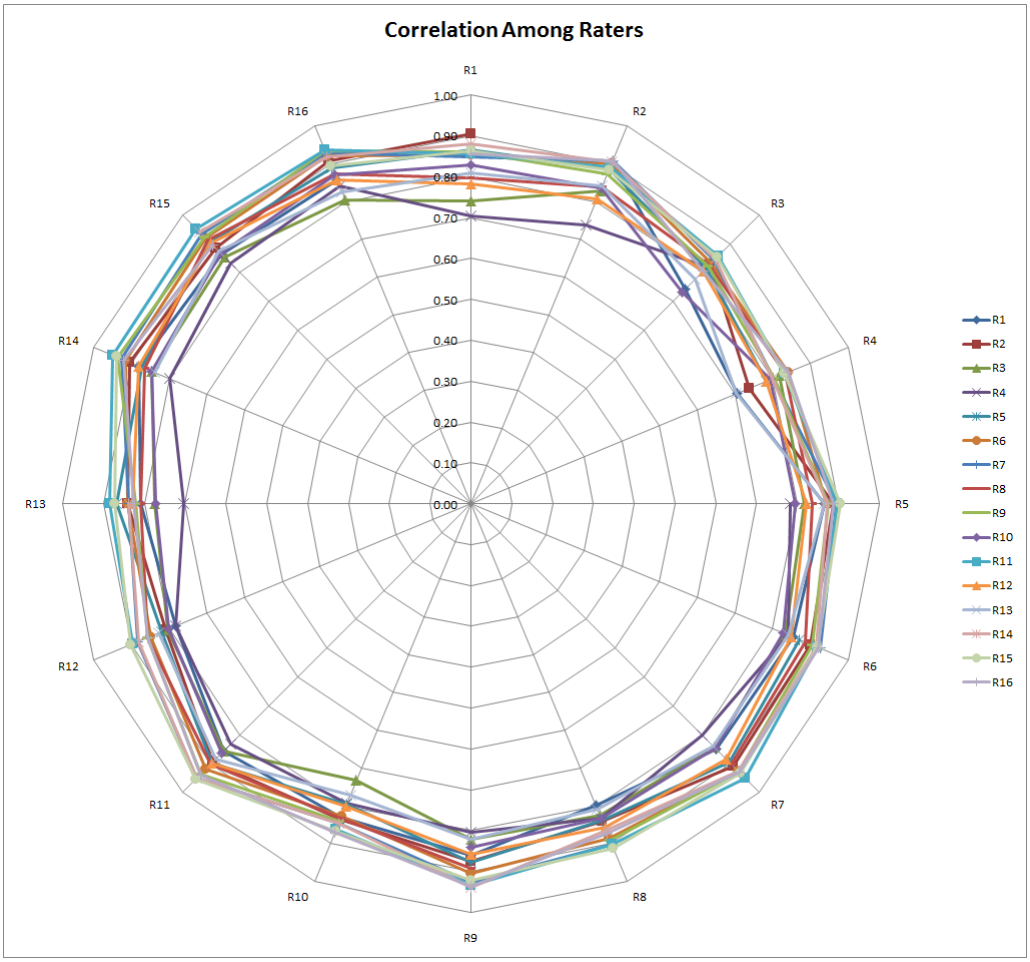

Figure 4.2: Correlation among raters

to capture perceptive quality feedback and all the observers were assessing the same media so that the scores should not be random. For this purpose, we calculated the correlation between observers. Hence each observer's score was checked against every other observer for each video. The results show that 90.8% pairwise correlation was higher than 0.8 whereas 0.7 was the lowest correlation value. Figure 4.2 shows the plot of correlations between pairs of observers. Although correlation among pairs of observers was high it's not enough to say that there was considerable agreement among observers generally.

4.4.2 Fleiss´ Kappa

Another statistical measure called Fleiss´ Kappa was employed for assessing the concordance (agreement) between observers when assigning categorical ratings [89]. Fleiss´ Kappa is commonly used in psychological studies for the said reason. It enables us to calculate concordance among any number of observers for a fixed number of items. The Kappa,

Table 4.4: Content Effect Table.

Effect Table				
Level	A	B	C	D
1	20.36	22.57	25.29	22.23
2	21.35	20.76	26.87	19.32
3	20.13	18.50	9.67	20.29
Δ	1.22	4.07	17.20	2.92
Rank	4	2	1	3

k, is defined by Equation 3.1. The value of k for 16 observers was 0.675 with standard error of 0.0083 within a 95% confidence interval of 0.650 to 0.685. The factor $1 - \bar{P}_e$ gives the degree of agreement that is attainable above chance, and the factor $\bar{P} - \bar{P}_e$ gives the degree of agreement actually achieved above chance. Where k=1 is absolute agreement and k=0 means no agreement at all.

This value of k=0.675 means that there was substantial agreement between observers, the table proposed by [90] was used for interpretation of the results. Kappa can be affected by prevalence but still provides more information than raw proportion of agreement.

Table 4.5: MOS scores per condition per trial.

MOS scores per condition per trial							
Exp. No	T1	T2	T3	T4	T5	T6	Mean
1	4.64	4.36	4.64	4.71	4.86	4.79	4.67
2	4.43	4.71	4.71	4.79	4.86	4.21	4.62
3	1.14	1.29	1.14	1.43	2.79	2.64	1.74
4	4.86	4.71	4.93	4.50	4.93	4.86	4.80
5	1.43	2.64	2.86	2.86	2.93	3.36	2.68
6	5.00	4.29	4.50	4.93	4.21	4.86	4.63
7	2.43	3.93	1.93	2.71	4.50	4.07	3.26
8	4.86	4.43	4.36	4.93	4.64	4.29	4.58
9	4.79	4.36	4.64	4.71	4.86	4.29	4.61

4.4.3 Signal to Noise Ratio Analysis

The observers' score was recorded and the mean was calculated for all users on a per video basis. Later, each category of noise factor was combined to produce Table 4.5. It shows conditions in the rows and noise factors in the columns. The mean for all the noise factors combined was also calculated for further analysis. We want to increase

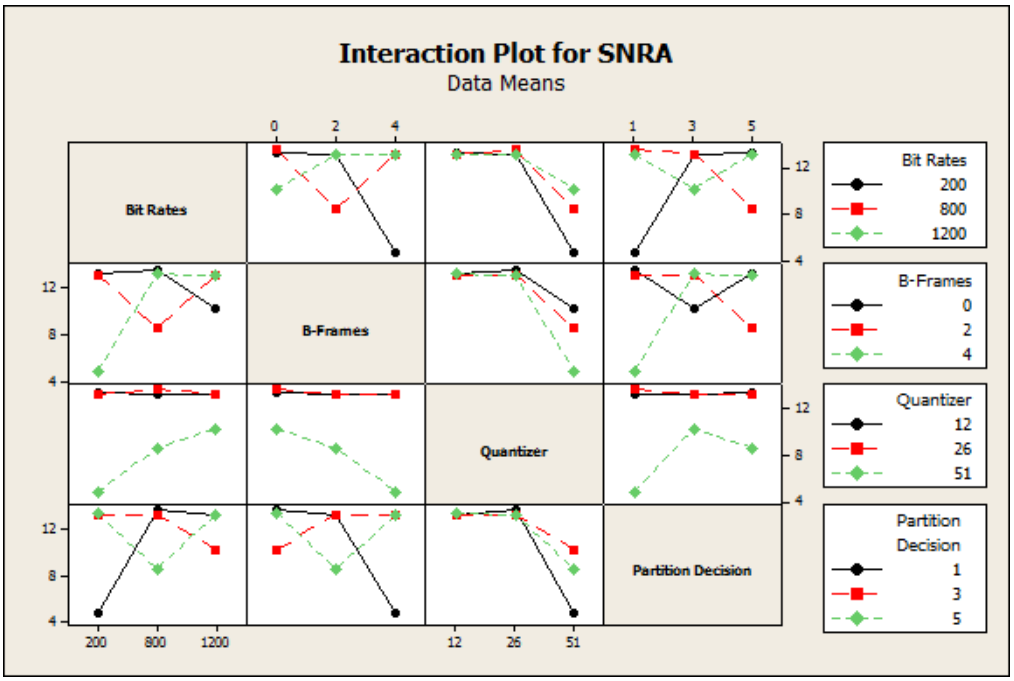

Figure 4.3: Interaction Plot

the perceived quality whenever possible, hence the aim of the analysis is to find out parameters that affect the perceived quality. We have computed the signal to noise ratio (S/N) for each condition of the target and created a response chart, Cct on the perceived quality. We used Equation 4.1 for calculating the S/N.

$$S/N = -10\log(\frac{1}{n}\sum_{i}\frac{1}{y_i^2}) \tag{4.1}$$

Once the SN ratio is calculated for each trial i.e. for each noise factor type, we came up with the response chart shown in Table 4.6. Response chart shows the levels for each parameter and respective SNR. In order to calculate the effect of a parameter we wanted to find out the range of its effect. So, from Table 4.6, we generated the effect table i.e. Table 4.4. Table 4.6 also shows calculation of the effect of SNR for parameter C where its level was 2. To populate the effect table for level two of parameter C i.e. the quantizer, we took the mean of all three SNR values pertinent to level 2. This became the level 2 value for the quantizer. Once all the values were calculated we calculated Δ which was the difference between the maximum SNR and the minimum SNR as shown in Table 4.4. This gave us the range of effect for each parameter. The bigger the value, the higher the effect of that parameter on quality. We ranked the parameters as per their range of effect. From Table 4.4 we can identify that the quantizer parameter had the biggest value of Δ whereas parameter B-frames came second while parameter bit rates were the least effective. Figure 4.3 i.e. also called

Table 4.6: Response Table.

Response Table				
A	B	C	D	SN
1	1	1	1	28.62
1	2	2	2	25.47
1	3	3	3	6.99
2	1	2	3	29.25
2	2	3	1	12.18
2	3	1	2	22.62
3	1	3	2	9.86
3	2	1	3	24.63
3	3	2	1	25.90

the interaction plot shows the effect of the parameter quantizer in relation to the other parameters. We want to highlight the fact that bit rate has been considered to be the most commonly used parameter for developing parametric statistical models for video quality [96],[97], [98], [99] and [103]. Apart from parametric models, the parameter bit rate has been regularly used as a comparison parameter in the latest techniques being developed for estimating QoE [104]. The interaction plot Figure 4.3 graphically represents the interactions between all the parameters. We found that due to the fact that bit rate is mainly controlled by the quantization parameter, the quantizer value increases when the bit rate is reduced. The same was reported by [105]. In addition to that, we also found that, at lower bit rates, increasing B-frames over all reduces the perceived quality whereas at higher bit rates, higher B-frame values result in better perceived quality. At a lower bit rate, as the partition decision quality choice parameter increases, quality gradually increases. At higher bit rates we see a varied response which is counter intuitive and we need to further investigate these variations. With an increased value of the QP, quality goes down irrespective of B-frame level selection. The Quantizer relationship with the partition decision parameter is such that on low and medium QP values, the perceived quality stays high but at a high QP value i.e. 51, quality is reduced drastically and the partition decision parameter selection has little or no effect on overall perceived quality. From Table 4.5, we calculated the variance and found that for conditions 1,2,4,6,8 and 9 the variance was low and it shows that noise factors of motion, complexity and location did not affect these conditions as much. These conditions were all rated higher with a MOS score of 4 which is almost always considered as good or excellent quality by observers generally. In cases of condition 3, 5 and 7 the variance is high and the conditions were highly affected by the noise factors i.e. motion, complexity and location. The reason seems to be that the selection

of highest value of QP introduces high loss compression. Hence, in order to compress HD videos, and still achieve a high perceptive MOS, high QP values should always be avoided. Figure 4.4 shows the pie chart of effect for each parameter. We also calculated the PSNR value for each video frame by frame and used the mode value as the representative value for each video. We used the mode because we wanted to select the most frequent PSNR value to represent PSNR score. Generally, PSNR scores were found to be related to the MOS scores. In Figure 4.5 the differences were due to the weakness of PSNR against structural artefacts [31]. We also repeated the same analysis for noise factors and found out that motion was the most influential noise factor. In order to improve quality we need to consider the level of motion within the content.

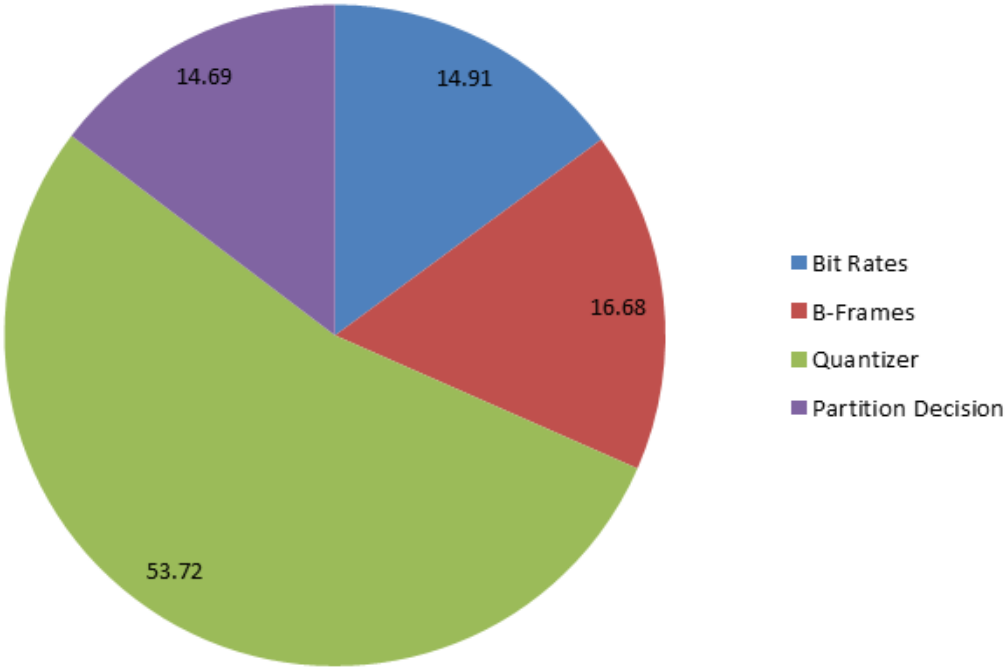

Figure 4.4: Pie Chart of influential parameters of content domain

Optimal Parameter Configuration

SNR analysis for content domain identified Quantization parameter as the most influential parameter. We need to find the optimal configuration for the content domain parameters in order to improve SNR. For this purpose we analyze the main effect plot of Signal to Noise Ratio (SNR) and identify the level of each parameter which enhance the SNR. Figure 4.6 show the main effect plot for SNR with highlighted parameter levels. Red circles identify the settings which will ensure high SNR. Combination of these parameter levels will become the optimal setting for content domain. In order to

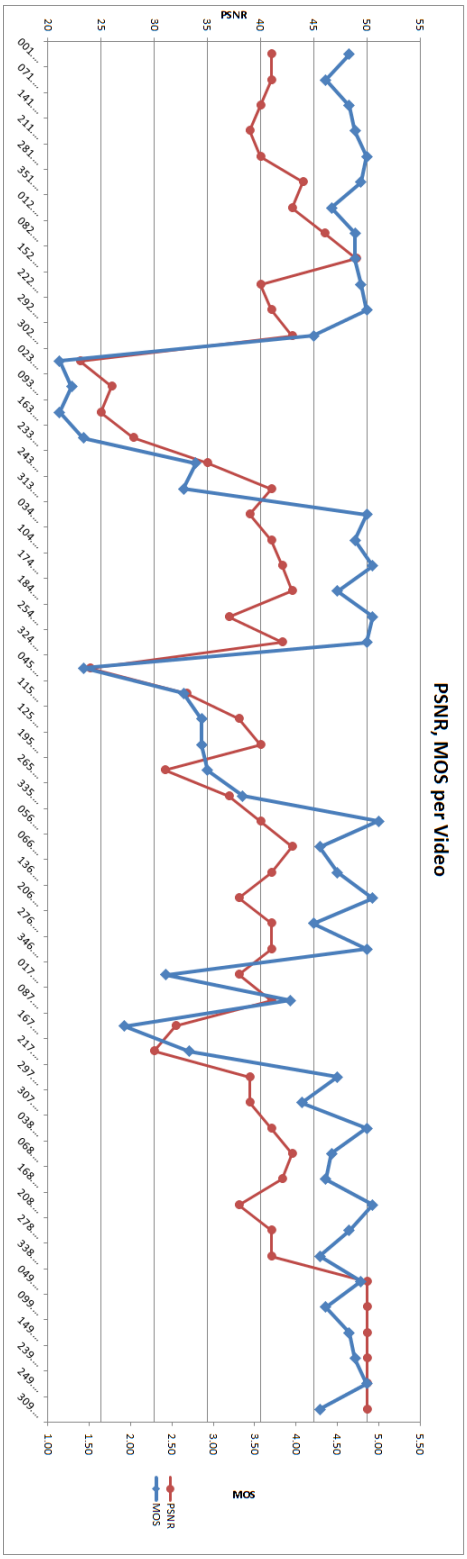

Figure 4.5: Combination of noise factors

Table 4.7: Predicted SNR values for optimal configuration

Predicted Values				
	Configuration	PSNR	PMEAN	PSTDE
1	A 3, B 1, C 2, D 2	15.86	5.00	0.43
2	A 2, B 1, C 2 , D 3	13.61	4.79	0.17
3	A 1, B 1, C 1, D 1	13.36	4.67	0.17

verify our conclusion predicted the SNR values on the basis of these configurations and some other variations for comparison. Our selected setting should produce the largest SNR value. Table 4.7 highlight the optimal setting as well as show the predicted values for different combinations. In Table 4.7 PSNR means predicted SNR, PMEAN is predicted mean and PSTDE is predicted standard deviation. Configuration highlighted is the optimal configuration identified from the Figure 4.6. We can see that the predicted SNR is the highest for this configuration. Predictions for generating Table 4.7 were done by using prediction option within DOE menu of Minitab16.

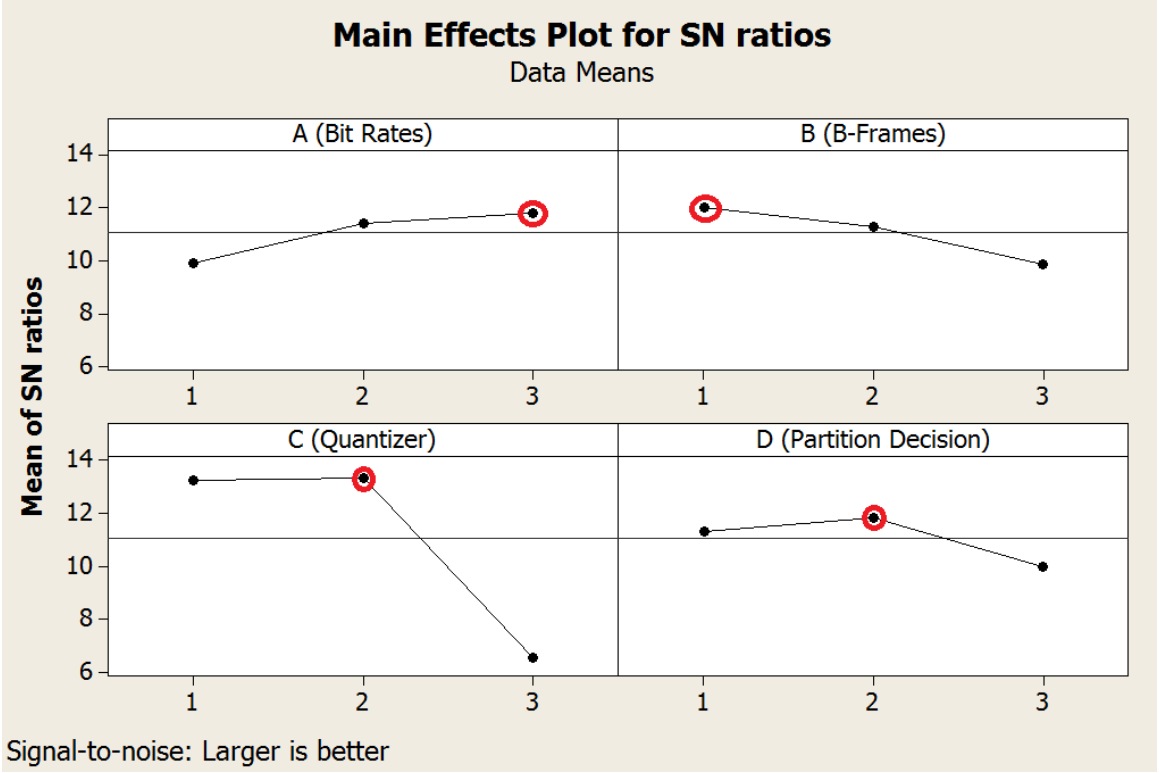

Figure 4.6: Optimal parameter configuration for content domain

4.4.4 Quality Function Deployment

Quality Function Deployment (QFD) is a method to transform user demands into quantitative parameters. It also enable quality deployment of these selected quantitative parameters and it also provide the methods for achieving quality. One of the important tools regularly used in QFD deployment is House of Quality (HoQ). HoQ is a matrix which resembles a house where different sections are used for defining relationship between customer needs and service provider's capabilities. In order to use HoQ we identified user expectations from feedback forms filled in by volunteers at the end of subjective experiments. The feedback was converted into expectations shown in Table 4.8. These demands were then presented to an independent group of 16 volunteers. They were asked to view these videos and then rate them on the basis of their importance. A Mode value was considered for rater responses. The video selection for this experiment was done on the basis of the MOS for each video. We were interested in acceptable videos and maximum quality videos. All those videos which consistently got an "excellent" rating from all raters were considered for the maximum quality videos. Videos without a single score of less than fair but no single score higher than good were placed in the acceptable group. Raters were asked to rate each expectation/demand on a Likert scale of 1-5. For HoQ analysis we used the same four parameters i.e. bitrate, b-frame, quantization and partition decision. The customer demands were the same as mentioned in Table 4.8. The same feedback score was used for the analysis. An expert was asked to provide technical feedback with reference to customer demands. Table 4.8 shows the customer demands, the Importance rating given by volunteers and corresponding expert feedback. Weights were calculated using customer feedback and expert feedback. These weights were calculated by using Equation 4.2. The results are shown in Table 4.9. According to HoQ, the parameters of packet loss, jitter and packet reorder are equally influential. Parameter importance was plotted using the pie chart shown in Figure 4.7.

The Likert scale used represents the most important demand represented by "5" and "1" representing the least important demand. We collected technical characteristics data from domain experts.

$$W_j = \sum_{i=1}^{r} M_{ij} \times I_i \tag{4.2}$$
$$i = 1....r, j = 1....c$$

4.4.5 Comparison of Taguchi Method and HoQ Matrix

These two methods investigated the problem from their different perspectives. The Taguchi method uses a loss function to calculate the signal to noise ratio and rank the

Table 4.8: HoQ with raters and experts feedback.

HoQ matrix showing customer demands and experts feedback					
User Expectations/Demands	Importance Rating	Bitrate	B-Frame	Quantizer (QP)	Partition Decision
Smooth Video (No Jitter)	5	1	0	1	0
Video with No Blockiness (Not pixelated)	3	3	1	9	0
Video with No blurring	3	3	1	9	0
Synchronization of voice and video	4	1	0	0	0
Meets acceptable quality (Overall)	4	1	1	9	1
Time to load before playback	3	1	1	1	0
Acceptable Video Quality for high motion (Camera panning or fast moving content)	5	1	3	0	9
Max Video Quality for low motion (Head and shoulder)(Fixed camera)	2	3	0	9	0
Acceptable Video Quality for Complex scenes (Many object/lots of colours/Details)	5	3	3	9	0
Acceptable Video Quality for simple scenes (Few object/few colours/Not much details)	3	0	0	3	0
Continuity of service	5	1	0	0	0

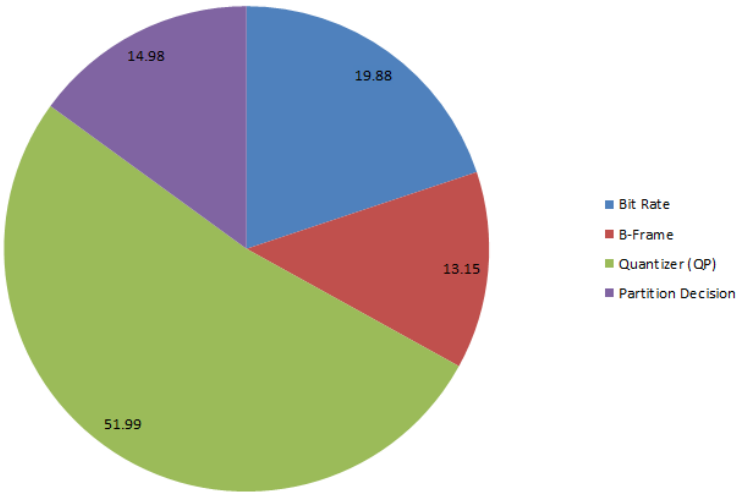

Figure 4.7: Pie Chart of influential factors using HoQ matrix

Table 4.9: Calculated Weights for HOQ.

Ranking of Parameter by Calculated Weights				
	Bitrate	B-Frame	Quantizer	Partition Decision
	2	4	1	3
W_i	65	43	170	49

parameters on the basis of range of effect. QFD considers input from raters as well as technical feedback from the experts. Experts were asked to rate each demand against a technical parameter using a 1, 3 and 9 scale. A rating of 1 meant no association between the technical parameter and demand, 3 meant a weak association and 9 meant a strong association. Weights are calculated using Equation 4.2 and, based on these weighted importance scores, the parameters were ranked. Weighted scores are shown in Table 4.9. Once the weights were calculated, we needed to check if the results were statistically significant. For this purpose permutation sampling was used. From Table 4.10 we can see that only the parameter of Quantizer is significantly different from all other parameters.

There is an insignificant difference between jitter and packet loss and packet reorder. We obtained multiple comparison tables of p-values of technical characteristic differences, p-values smaller than 0.05 show a significant difference at the 5% level of significance. The results show that quantizer parameter is statistically the most significant parameter and the other three are statistically not significant in their effect. The Taguchi method and HoQ method both obtained similar results. This fact is evident

Table 4.10: P-value table using permutation sampling.

	B-Frame	Partition Decision	Bitrate	Quantizer
B-Frame	NA	0.5	0.4	0.01
Partition Decision	NA	NA	0.5	0.02
Bitrate	NA	NA	NA	0.3
Quantizer	NA	NA	NA	NA

by comparing the pie charts presented in Figure 4.4 and Figure 4.7.

The results obtained by two independent methods using different approaches have provided us with confidence in our results. Both these methods highlight the fact that the parameter quantizer is the most influential parameter where as bitrate, b-frame and partition decision are relatively less influential on QoE and that these three are similar in their effect on QoE.

4.4.6 Feedback from observers

At the end of our experiment we wanted to collect information from users about their impressions concerning the experiment. This would help us in conducting further experiments and avoiding or reducing unwanted psychological effects. They were asked to fill in a questionnaire which was directed towards finding better ways of organizing the subsequent experiments. The observers become disconnected from the experiment process due to the type of content selected and/or due to the extended length of the experiment. Some work is needed to be done in this regard, to evaluate the attention span of adult observers with regards to HD video assessment. There is a need to provide observers with a choice of video genre selection. From the feedback we learned that almost all the participants would like to have more control on the selection of videos. This fact could also improve the interest of a user in the overall experiment and would ensure consistent scores based on perception of quality. From our results we also concluded that repetition of selected video could also be a source of boredom. Though novice observers were not able to name the artefacts, they were able to identify the effects of these artefacts. We also wanted to know if the observers understood the explanation given at the start of the experiment. It could be the case the observers were assessing some other aspect of the video and misunderstood the notion of video quality.

4.5 Summary and Conclusion

In this experiment, we analyzed the influence of content domain parameters; namely, bit rate, quantization (QP), partition decision and B-frame on QoE of an IPTV like service. We also investigated the effect of uncontrolled parameters namely motion, complexity and location. Quantization was found to be the most influential control parameter. On the other hand, motion was the most influential noise parameter. By avoiding low quantization configuration within H.264 high perceptual quality for HD video can be guaranteed for any class of video content. Low quantization affects the output quality more than any other parameter. High perceptual quality of HD video containing high motion content can be achieved if low quantization is avoided in H.264 compression. Traditionally, the Taguchi method has been used for model/parameter optimization in the engineering discipline. We used the Taguchi method for finding the parameters which can be used for optimization of video quality. The Taguchi method enabled us to reduce boredom effects by reducing the total experimental time using orthogonality associated with the method. We identified a relationship between the parameters of Quantization with B-frame and the partition decision. We also were able to confirm the findings of [106] about the relationship between bit rate and quantization.

Chapter 5

Network Domain Influential Parameters

5.1 Introduction

In the network domain we analyzed the effects of network parameters on perceptual quality or Quality of Experience (QoE). Here we identified appropriate network domain parameters that can be simulated to introduce controlled artefacts. Investigations were made to identify the order of effect of these parameters on QoE. These effects were introduced by varying parameter values between low, medium and high values. The Video on Demand (VoD) service of Internet Protocol Television (IPTV) was emulated within our testbed. We used Netem which is part of the iproute2 package of tools now available in most current Linux distributions. It provides network emulation functionality for testing protocols by emulating the properties of wide area networks. It's already been realized that network conditions such as end to end delay, jitter, packet loss and packet reorder will affect output quality. We conducted subjective tests for video quality assessment. Observers were required to rate the perceptual quality of each video. In Chapter 4, we tested the content domain parameters and identified parameters having first and second order effects on perceptual quality [107]. For the current experiment, we wanted to identify the order of effect caused by network domain parameters. For this experiment, we considered the impact of psychological effects and utilized the techniques to overcome these effects as described in Chapter 3 Section 3.3.1. A number of parameters were studied from the network domain and were used for modeling. In order to develop better models we need to investigate the order of effect of these parameters and use the most influential ones for modeling. We shall establish through our experiments that the packet reorder parameter was the most influential parameter, although, it was found that there was not much of a difference in impact between packet reorder, jitter and packet loss. We analyzed the results using

Table 5.1: Control Factors and Levels.

Control Factors and Levels				
Control Parameters	Labels	Level 1	Level 2	Level 3
Packet Loss	A	0.20%	1.00%	1.80%
Delay	B	20ms	100ms	250ms
Jitter	C	10ms	15ms	40ms
Packet Re-order	D	5%	7%	15%

both Taguchi SNR analysis and House of Quality (HoQ) from the QFD method.

5.2 Design of Experiment for network domain

For this experiment, the DoE procedure using the Taguchi approach was implemented to study the effects of multiple network parameters impacting on the QoE. In order to find the relationship between parameters within the network domain and overall quality, there was a need to identify the most influential parameters. By employing Taguchi DoE we studied the combinations of network parameters (control factors) and rater feedback in terms of Mean Opinion Scores (MOS) under the influence of several noise factors. Orthogonal arrays are the smallest possible matrix of combinations of the control parameters. This technique is good for minimizing repetitions, cost and time. An orthogonal array is selected on the basis of the number of parameters and levels per parameter. After setting up the data according to Taguchi recommended arrays, test runs were performed randomly. Taguchi prescribes three loss functions i.e. smaller the best, larger the better and nominal the best. We used the *larger the better* equation loss function to calculate the signal to noise ratio. Signal to noise ratio was used to analyze results and to determine the influence of individual factors and their order of effect.

5.2.1 Parameters

The parameters considered for this experiment were from network domain. The following discussion explain the selection of control factors from network domain.

Control Factors

The control factors that were analyzed during this experiment were: packet loss, delay, jitter and packet reorder. Table 5.1 shows these parameters and their levels. In [108] frame rate, bitrate, bandwidth and packet error rate were used for predicting perceptual video quality. Our approach involves looking at each domain independently to identify

within-domain and across-domain parameter effects. Hence, for our experiments, we tested bitrate within the content domain. Packet loss, packet reorder, delay and jitter were studied within the network domain. The parameter packet loss has been under study for finding its relationship with perceptive video quality in a number of studies, [109] [110] [111][112][113][46][114]. In Boulos et al. [112] an attempt was made to relate network parameter packet loss distribution and picture loss percentage to the effects on perceptual quality. All these above mentioned studies were able to show that content type was introducing a multiplier effect on packet loss. We also note that this variation is not accounted for in most current models. These models would suggest finding fitting variables from the available data for improving the model performance. There was a need to look for a method which could suggest a variation-resistant parameter selection. This method needs to look into the effects of various levels of a parameter's extreme values and to suggest combinations of parameters and their levels to be avoided to contain the content dependent variations. The Taguchi method performs all of the above tasks.

In Dai et al. [111] the impact of different packet loss frequencies and inter arrival times within a specific short period was evaluated. A logistic model was also presented for predicting packet loss based video degradation. In their future work authors mentioned that there was a need for an objective model that looks at the combined effect of the content domain parameters. Packet loss has been studied and analyzed regularly in the network domain. It's been considered to be the main culprit for introducing quality degradation [115]. The values selected for packet loss in this experiment were 0.2%, 1.0% and 1.8%. We selected 0.2% as its effect on quality is almost negligible. 1.8% packet loss is reported to degrade video quality to an unacceptable level. In the paper by Boulos et al. [112] they considered packet losses between 0.1% and 1.6%. Packet reordering was also considered as one of the contributors to degradation of video quality. Research done by [116] considered techniques to mitigate the effects of packet reordering. Authors considered 5% random packets were delayed by a specified amount of time. For our experiment, we considered 5%, 7% and 15% packet reordering. Packet reordering around 15% degrades video quality to an unacceptable level. Temporal jitter was also considered as a parameter affecting video quality [117]. The paper by Claypool et al. [118] concluded that jitter was degrading video quality as much as packet loss. Two levels of packet loss i.e. 8% and 22% were also considered. For their work the authors used two levels of jitter i.e. low jitter and 3 times the low level of jitter was considered to be high jitter. We considered delay between 20ms and 250ms. It's known that a value of delay around 200ms and above reduces the quality of video to an unacceptable level. Multiple parameters were discussed in these papers [119] [120] [121]. In [122] packet loss, packet reorder and bit rate were considered. The researchers selected

three parameters where one parameter was from content domain and the other two were from network domain. We have already discussed in the conclusions of Chapter 4 that the parameter bit rate was found to have the least influence on perceived video quality or QoE. In this work we are attempting to find the most influential parameters from the network domain. Results of this experiment will enlighten us as to whether the parameters packet loss and packet reorder are good parameters for developing a better model.

Noise Factors

Noise factors considered for this experiment were motion, complexity and location (indoor or outdoor scene). Table 5.2 shows the combination of noise factors.

Table 5.2: Noise Factors and Levels.

Noise Factors and Levels		
Noise Parameters	Level 1	Level 2
Motion	High	Low
Complexity	High	Low
Location	High	Low

5.3 Experimental Setup

This experiment was designed to identify the influential parameters from network domain. For this the following test-bed was constructed to facilitate the experiment.

5.3.1 Setup

For our test bench, video clips for the experiment and the test layout generally followed the discussion presented in Section 3.3. Computers used for this experiment were Intel core i5 CPU machines running at 3.60 GHz with 4 GB RAM. Running 64 bit windows 7 Enterprise with Service Pack 1 installed. These machines were using integrated Intel HD graphics. Monitors used were ViewSonic VS 13239 LED 1080p Full HD. This streaming server was running an Intel Core i7-2600 CPU @ 3.40 GHz under the Ubuntu 12.04 operating system. We extracted test sequences following the method discussed in Section 3.3. Table 5.3 shows the L9 array with control factors.

5.3.2 Test Layout

Test layout for this experiment was similar to the one described in Section 4.3.3.

Table 5.3: L9 Array with control factors.

L9 Array with control factors				
Exp. No	A	B	C	D
1	0.20%	20ms	10ms	5%
2	0.20%	100ms	15ms	7%
3	0.20%	250ms	40ms	15%
4	1.00%	20ms	15ms	15%
5	1.00%	100ms	40ms	5%
6	1.00%	250ms	10ms	7%
7	1.80%	20ms	40ms	7%
8	1.80%	100ms	10ms	15%
9	1.80%	250ms	15ms	5%

5.3.3 User Task

User task for this experiment was similar to the one described in Section 4.3.4.

5.3.4 Customization

For this experiment we used Netem to simulate network anomalies [123] [124]. Our streaming server was running on Ubuntu 12.04 LTS. We used the network traffic shaping capability of Netem. An application was developed in Python which used the command line tool called tc for automation of the testing process. For traffic shaping, Netem uses traffic classification, policy rules, queue disciplines and quality of service (QoS). The tc utility communicates with the kernel via the netlink socket interface. A queuing layer exists between the network device and the protocol output. The default queuing discipline is a simple FIFO packet queue. Combinations of packet loss, delay, jitter and packet reorder were simulated. We showed the original undistorted video and then displayed the distorted video. Conditions were toggled using the Python application. During each assessment, conditions were updated immediately, once the first video finished. This ensured that the combinations could be enforced and traffic was stable as per requirement before the evaluation of the distorted video.

5.4 Results and Analysis

As already discussed in Section 4.4 there is a need to establish the validity of user feedback. A feedback is valid if its non-random as well as there is general consensus among the observers. The following methods were used to establish these facts:

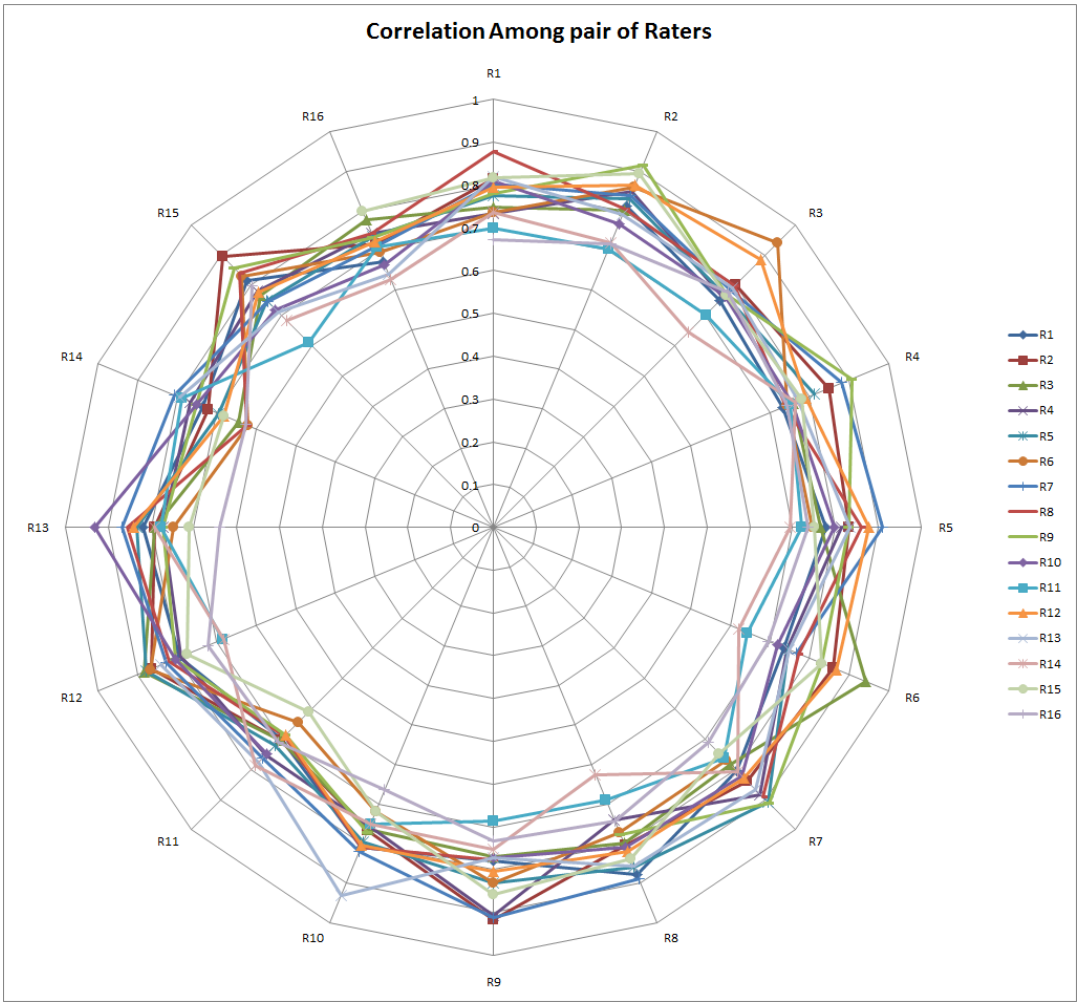

Figure 5.1: Correlation among raters

5.4.1 Pairwise Correlation and Fleiss´ Kappa

The feedback in a subjective experiment can be considered valid if the majority of raters agree on the output. We analyzed raters' feedback for correlation between individual raters. Each observer's score was checked against every other observer for each video. The results show that 96.3% of pairwise correlations were higher than 0.7, whereas 0.62 was the lowest correlation value. Figure 5.1 shows the plot of correlations between pairs of observers. In addition to correlation we also utilized Fleiss´ Kappa for assessing the concordance (agreement) between observers when assigning categorical ratings [89]. It enables us to calculate concordance among any number of observers for a fixed number of items. Fleiss´ Kappa k for 16 observers was 0.432. We used Equation 3.1 for calculating kappa i.e. k. Table A.1 shows the values calculated using this equation. This value of k obtained in this way signifies that there was moderate agreement between observers. A table proposed by [90] was used for interpretation of the results.

Table 5.4: MOS scores per condition per trial.

MOS scores per condition per trial							
Exp. No	T1	T2	T3	T4	T5	T6	Mean
1	4.18	4.36	3.82	4.71	3.64	4.36	4.18
2	3.73	3.27	3.82	3.09	3.64	2.82	3.39
3	1.00	1.55	1.00	1.27	1.36	1.55	1.29
4	2.45	2.73	2.73	2.36	1.00	2.18	2.24
5	1.00	1.00	2.18	2.82	2.45	1.91	1.89
6	2.45	2.45	2.00	3.18	1.36	2.82	2.38
7	1.55	1.64	1.00	1.73	2.64	2.82	1.89
8	1.00	1.91	1.91	1.18	1.64	1.00	1.44
9	2.18	2.45	1.82	1.64	2.09	2.09	2.05

5.4.2 Signal to Noise Ratio Analysis

The observer's score was recorded and a mean was calculated for all raters on a per video basis. We wanted to increase the perceived quality whenever possible. Hence the aim of the analysis was to find parameters that affected the perceived quality. We have computed the signal to noise ratio (SNR) for each condition and created a response chart, and determined the parameters that have the highest and lowest effect on the perceived quality. We used Equation 5.1 for calculating the SNR. Once the SNR is calculated for each trial i.e. each noise factor type, we derived Table 5.4. It shows conditions in the rows and noise factors in the columns. The mean for all the noise factors combined was also calculated for further analysis. We wanted to increase the perceived quality whenever possible; hence the aim of the analysis was to establish the parameters that affected the perceived quality. Response chart in Figure 5.5 shows the levels for each parameter and respective SNR. In order to calculate the effect of a parameter we wanted to determine the range of its effect. So, from Table 5.5 we generated the effect table i.e. Table 5.6. Table 5.5 also shows calculations for the effect of SNR for parameter C where its level was 3. To populate the effect table for level 3 of parameter C i.e. the jitter, we took the mean of all three SNR values pertinent to level 3. This became the level 3 value for the jitter. Once all the values were calculated we calculated Δ the difference between the maximum SNR and the minimum SNR as shown in Table 5.6. This gave us the range of effect for each parameter. The bigger the value, the higher the effect of that parameter on quality. We ranked the parameters as per their range of effect. From Table 5.6 we can identify that the packet reorder parameter had the largest value of Δ whereas the parameters of jitter and packet loss also had an almost equal effect on quality. Parameter delay displayed the least effect on perceived quality. For Figure 5.3 shows the effect of the parameter quantizer in

Table 5.5: Response Chart

Response Table				
A	B	C	D	SN
1	1	1	1	20.50
1	2	2	2	18.63
1	3	3	3	14.33
2	1	2	3	10.77
2	2	3	1	7.87
2	3	1	2	11.40
3	1	3	2	8.60
3	2	1	3	10.39
3	3	2	1	17.07

relation to other parameters. In the network domain, the parameter of packet loss is the most widely investigated parameter [115][117][118][125][126]. Though from the results we can clearly see that packet reorder and jitter are as important as packet loss. For developing better models we need to consider all these parameters. Our results endorse the finding of [108] as well i.e. that the effect of network parameters on quality is more than the effect of content parameters [107]. The changes in quality due to different combinations of parameters within the network domain reduced the perceptual quality to lower levels compared to the content domain parameters. Figure 3 and Figure 4 in [118] show the relationship of packet loss and jitter to perceptual quality. The interaction plot shown in Figure 5.2 also shows the same behaviour. In addition, this interaction plot shows the relationship between packet reorder and all other parameters. The packet loss relationship with all other parameters is almost similar. When packet loss is at its minimum value the perceptual video quality is good. As the value of packet loss is increased the perceptual quality decreases immediately and drastically.

$$SNR = -10\log(\frac{1}{n}\sum_{i}\frac{1}{y_i^2}) \tag{5.1}$$

At the maximum value of packet loss, the quality is really poor. Jitter and packet reorder at their maximum value also exhibit the same behaviour. Hence this interaction plot shows that perceptual video quality is affected by jitter, packet reorder and packet loss equally. For developing a model to predict perceptual quality we need to consider these three parameters at the very least. From our results, we were able to confirm the findings of [118] and [122]. Apart from these results we also noticed that the amount of data that was required for the experiments conducted by other researchers

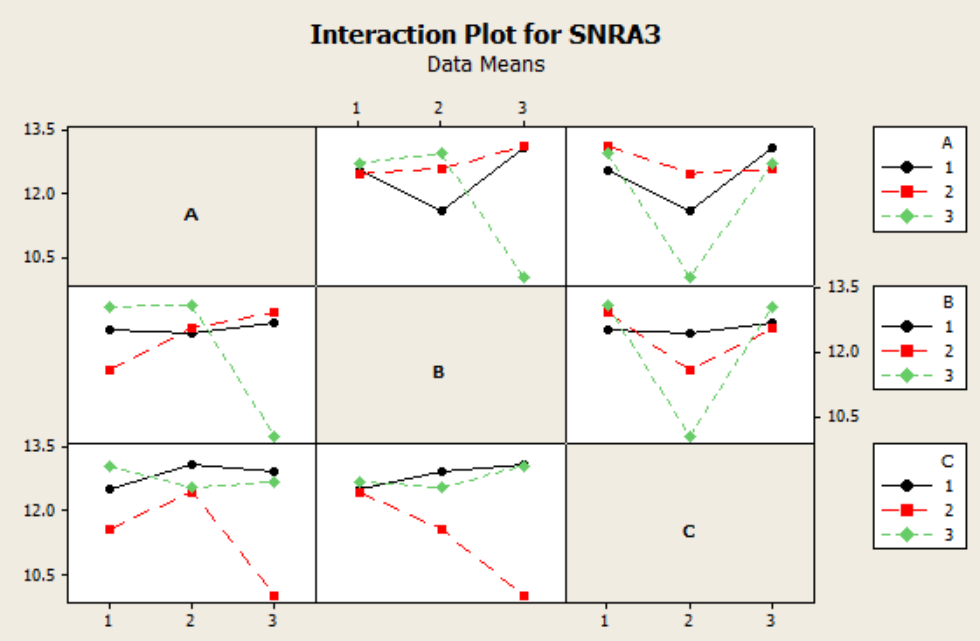

Figure 5.2: Interaction Plot

Table 5.6: Effect Table

Effect Table				
Level	A	B	C	D
1	8.18	7.16	7.00	7.22
2	5.00	5.33	7.21	6.98
3	4.05	4.75	3.03	3.04
Δ	4.13	2.41	4.17	4.18
Rank	3	4	2	1

was huge. A large number of participants was required and a large number of videos were processed during their experiments. We were able to confirm their results and were able to shed light on a few other relationships within a much shorter period of time, at reduced cost and requiring fewer volunteers. We were also able to mitigate any psychological effects by keeping the experimental run time low. Hence Taguchi DoE can be used with confidence for designing subjective experiments used for assessing perceptual video quality and the generated results are seen to be reliable.

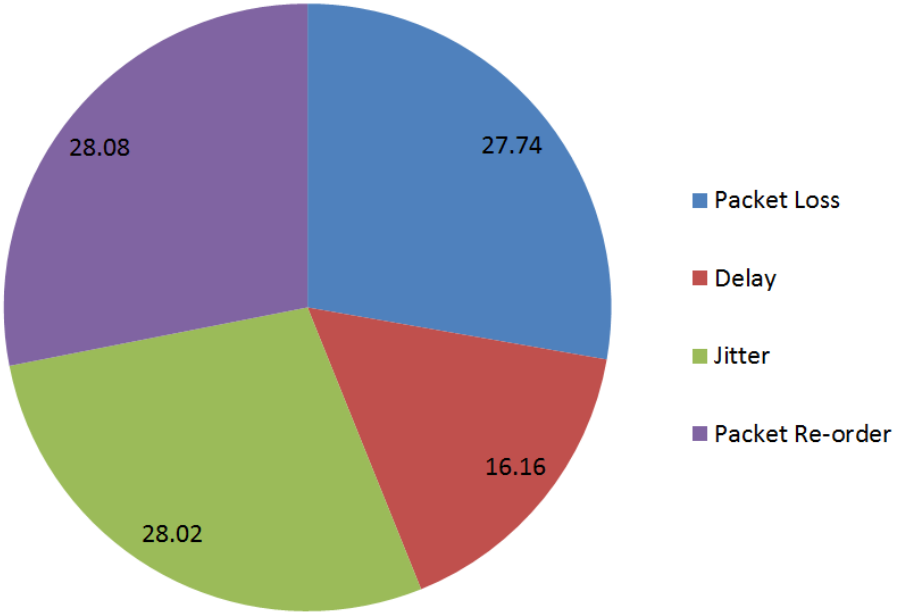

Figure 5.3: Pie Chart of influential factors

Optimal Parameter Configuration

SNR analysis for network domain identified three parameters having substantial effect on QoE. We need to find the optimal configuration of all the three parameters in order to improve SNR. For this purpose we analyze the main effect plot of Signal to Noise Ratio (SNR) and identify the level of each parameter which enhance the SNR. Figure 5.4 show the main effect plot for SNR with highlighted parameter levels. Red circles identify the settings which will ensure high SNR. Combination of these parameter levels will become the optimal setting for network domain. In order to verify our conclusion we predicted the SNR values on the basis of these configurations and some other variations for comparison. Our selected setting should produce the largest SNR value. Table 5.7 highlight the optimal setting as well as show the predicted values for different combinations. In Table 5.7 PSNR means predicted SNR, PMEAN is predicted mean and PSTDE is predicted standard deviation. Configuration highlighted is the

Table 5.7: Predicted SNR values for optimal configuration

Predicted Values				
	Configuration	PSNR	PMEAN	PSTDE
1	A 1, B 1, C 2, D 1	12.53	4.07	0.34
2	A 1, B 1, C 1 , D 1	12.32	4.17	0.39
3	A 2, B 1, C 2, D 1	9.35	3.29	0.68

optimal configuration identified from the Figure 5.4. We can see that the predicted SNR is the highest for this configuration. Predictions for generating Table 5.7 were done by using prediction option within DOE menu of Minitab16.

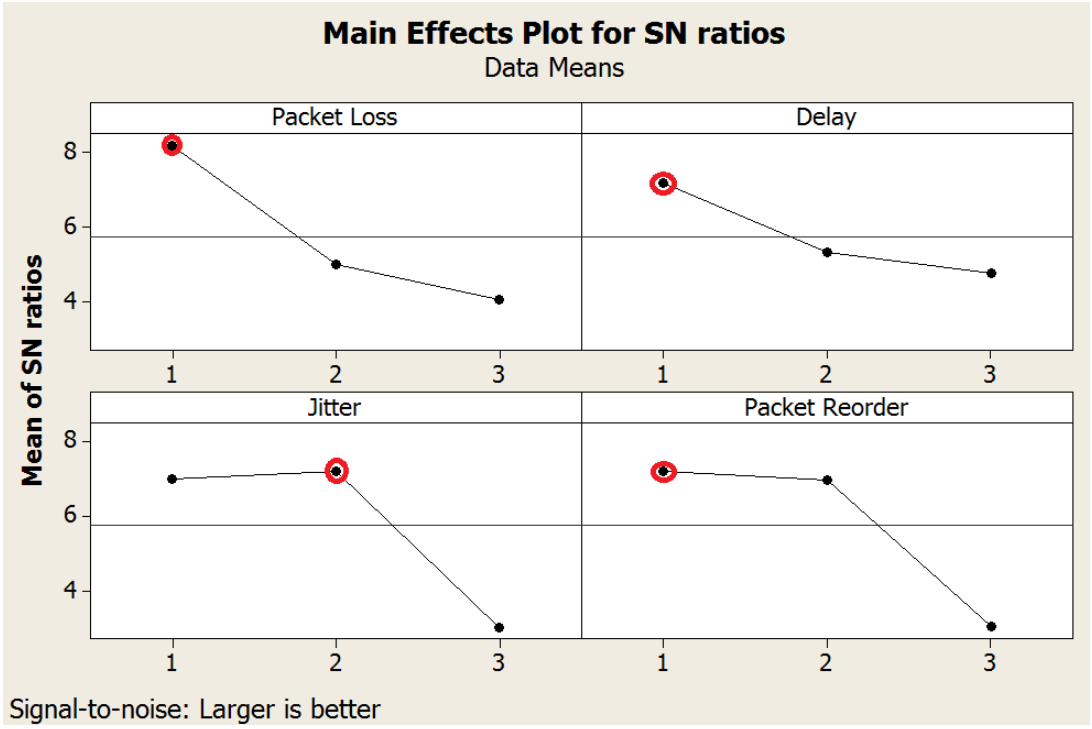

Figure 5.4: Optimal parameter configuration for network domain

5.4.3 Quality Function Deployment

In order to use HoQ we identified user expectations from feedback forms filled in by volunteers at the end of earlier subjective experiments. The feedback was converted into expectations shown in Table 5.8. These demands were then presented to an independent group of 16 volunteers. They were asked to view these videos and then rate them on the basis of their importance. A Mode value was considered for rater responses. The video selection for this experiment was done on the basis of the MOS for each video. We were

interested in acceptable videos and maximum quality videos. All those videos which consistently got an "excellent" rating from all raters were considered for the maximum quality videos. Videos without a single score of less than fair but no single score higher than good were placed in the acceptable group. Raters were asked to rate each expectation/demand on a Likert scale of 1-5. For HoQ analysis we used the same four parameters i.e. packet loss, delay, jitter and packet reorder. The customer demands were the same as mentioned in Table 5.8. The same feedback score was used for the analysis. An expert was asked to provide technical feedback with reference to customer demands. Table 5.8 shows the customer demands, the Importance rating given by volunteers and corresponding expert feedback. Weights were calculated using customer feedback and expert feedback. These weights were calculated by using Equation 5.2. The results are shown in Table 5.9. According to HoQ, the parameters of packet loss, jitter and packet reorder are equally influential. Parameter importance was plotted using the pie chart shown in Figure 5.5.

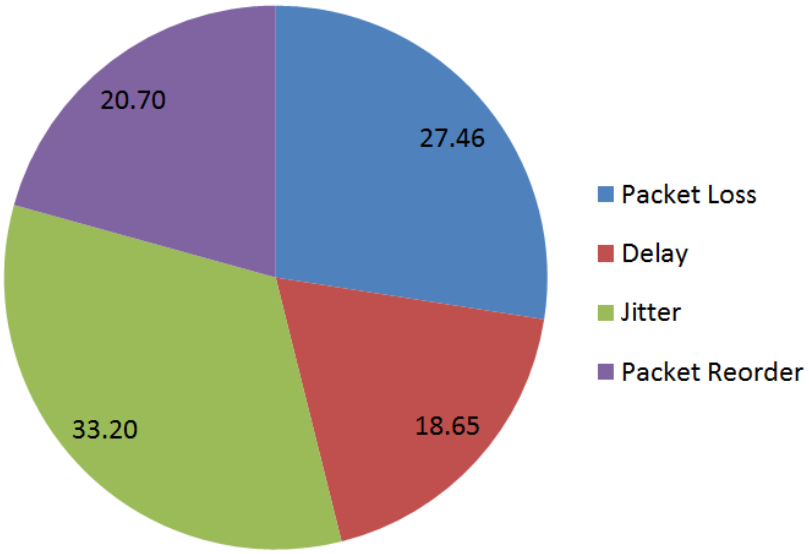

Figure 5.5: Pie Chart of influential factors using HoQ matrix

The Likert scale used represents the most important demand represented by "5" and "1" representing the least important demand. We collected technical characteristics data from domain experts.

$$W_j = \sum_{i=1}^{r} M_{ij} \times I_i \tag{5.2}$$
$$i = 1....r, j = 1....c$$

Table 5.8: HoQ with raters and experts feedback.

HoQ matrix showing customer demands and experts feedback					
User Expectations/Demands	Importance Rating	Packet Loss	Delay	Jitter	Packet Reorder
Smooth Video (No Jitter)	5	3	0	9	1
Video with No Blockiness (Not pixelated)	3	1	1	1	3
Video with No blurring	3	1	1	1	3
Synchronization of voice and video	4	3	1	1	1
Meets acceptable quality (Overall)	4	3	3	3	3
Time to load before playback	3	3	3	9	9
Acceptable Video Quality for high motion (Camera panning or fast moving content)	5	9	3	3	3
Max Video Quality for low motion (Head and shoulder)(Fixed camera)	2	3	3	3	1
Acceptable Video Quality for Complex scenes (Many objects/lots of colours/Details)	5	3	3	3	3
Acceptable Video Quality for simple scenes (Few object/few colours/Not much details)	3	3	3	9	1
Continuity of service	5	1	3	1	0

Table 5.9: Calculated Weights for HOQ.

Ranking of Parameter by Calculated Weights				
	Packet Loss	Delay	Jitter	Packet Reorder
	2	4	1	3
W_i	134	91	162	101

5.4.4 Comparison of Taguchi Method and HoQ Matrix

These two methods investigated the problem from their different perspectives. The Taguchi method uses a loss function to calculate the signal to noise ratio and rank the parameters on the basis of range of effect. QFD considers input from raters as well as technical feedback from the experts. Experts were asked to rate each demand against a technical parameter using a 1, 3 and 9 scale. A rating of 1 meant no association between the technical parameter and demand, 3 meant a weak association and 9 meant a strong association. Weights are calculated using Equation 5.2 and, based on these weighted importance scores, the parameters were ranked. Weighted scores are shown in Table 5.9. Once the weights were calculated, we needed to check if the results were statistically significant. For this purpose permutation sampling was used. From Table 5.10 we can see that only the parameter of jitter is significantly different from the delay parameter.

There is an insignificant difference between jitter and packet loss and packet reorder. We obtained multiple comparison tables of p-values of technical characteristic differences, p-values smaller than 0.05 show a significant difference at the 5% level of significance. The results show that packet loss, jitter and packet reorder are statistically not different from each other. The Taguchi method and HoQ method both obtained similar results. This fact is evident by comparing the pie charts presented in Figure 5.3 and Figure 5.5.

The results obtained by two independent methods using different approaches have provided us with confidence in our results. Both these methods highlight the fact that the parameter packet loss, jitter and packet reorder are statistically equally influential parameters.

5.5 Summary and Conclusion

In this experiment, we analyzed the influence of network domain parameters viz packet loss, delay, jitter and packet reorder on QoE of IPTV like service. The effect of noise factors like motion, complexity and location were also considered. Network domain parameters affect video quality more than content domain parameters. HoQ method

Table 5.10: P-value table using permutation sampling.

	Delay	Packet Reorder	Packet Loss	Jitter
Delay	NA	0.7471	0.1764	0.0256
Packet Reorder	NA	NA	0.2939	0.057
Packet Loss	NA	NA	NA	0.374
Jitter	NA	NA	NA	NA

and Taguchi method provided similar insight into the parameter ranking problem. HoQ method was used here to verify the results obtained from the Taguchi method. Packet reorder, jitter and packet loss are equally affecting perceived video quality. High packet loss, packet reordering and jitter must be avoided to provide an acceptable level of video quality. We draw two conclusions after comparing results of Taguchi method and HoQ method. Firstly: jitter, packet loss and packet reorder are similar in affecting perceptual video quality. Secondly: getting similar results from two independent methods indicates that the true perception of video quality was captured in subjective experiments. We need to develop models that can objectively predict video quality. These models need to concentrate on these influential parameters for better predictions.

Chapter 6

CPP Domain Influential Parameters

6.1 Introduction

In this component of our research we analyzed the effects of the CPP domain parameters on QoE. We identified CPP domain parameters that affect video quality. Investigations were made to identify the order of influence of these parameters on QoE. These effects were applied by varying parameters in a range between low, middle and high values. This experiment was a bit different from the two previous experiments. In this experiment, the CPP parameters that were chosen were concerned with hardware. We selected processing power, display size and buffer size as our control parameters. It has already been realized that individual hardware selection can affect video quality. The amount of available memory, buffer size, the medium of transmission, display size and technology, all affect video quality. CPP parameter selection and optimization will affect perceived output quality. We conducted subjective tests for video quality assessment. Observers were required to rate the perceptual quality of each video. We considered the ITU P.910 recommendations for conducting these experiments. Literature reviewed from the CPP domain usually involved finding the effects of a single parameter on video quality. Most parametric models developed for video quality measurements ignored the need to consider a parameter from the CPP domain. We included three parameters from the CPP domain and determined the suitability of these parameters to become one of the model parameters. Our results indicate that QoE is affected by selection of the buffer size - more than any other parameter from the CPP domain. Selecting a buffer size that is the equivalent of 300 ms will adversely affect the perceived video quality under the influence of varied content type and complexity.

Table 6.1: Control Factors and Levels.

Control Factors and Levels				
Control Parameters	Labels	Level 1	Level 2	Level 3
Display Size	A	Mobile 4.8”	Tablet 10.1”	Flat Screen 24”
Processing Power	B	1 GHz	2 GHz	3 GHz
Buffer Size	C	10 ms	300 ms	500 ms

6.2 Design of Experiment for CPP domain

This experiment was designed for identifying influential parameters from CPP domain. The parameters selection was difficult as most parameters where either a hardware or was configured within a hardware. For this reason, certain combinations were either very difficult to achieve. The following discussion highlights the justification for parameter selection.

6.2.1 Parameters

The following section discuss in detail the selection of control parameters from CPP domain. We continue to consider the same noise factors.

Control Factors

The control factors which were analyzed during this experiment were: display size, processing power and buffer size. Table 6.1 shows the control factor and levels for each parameter.

The first parameter considered for this experiment was display size. We considered a flat screen 24″, a Tablet 10.1″ and mobile screen of 4.8″. These three are the widely used form factors now-adays. Though the technology used for these screens are the state of the art display technologies at the present stage of the study, we wanted to investigate the effect of different form factors on QoE. The second parameter considered was processing power. Available processing power enables complex decoding and promises seamless video play. For a typical set top box (STB) a manufacturer aims to provide the cheapest possible devices with minimum power requirements. This forces them to introduce a minimal device that can work with the required codecs. STBs also make use of a GPU to provide better video performance. A third parameter under investigation was buffer size. Buffer size plays an important role in handling transmission artefacts. Configuring buffer size appropriately ensures higher perceived video quality. We considered buffer sizes equivalent to 10 ms, 300 ms and 500 ms. We wanted to investigate an appropriate combination of these factors to find out which

parameter is the most influential for video quality.

Noise Factors

Noise factors considered for this experiment were motion, complexity and location (indoor or outdoor scene). Table 6.2 shows the combination of noise factors. Taguchi design requires that each condition i.e. a specific combination of all 3 control parameters should be repeated for all noise factors. We had 3 noise factors so we were looking at 2^3 combinations of noise. It was required that we repeat the experiment for each condition with 8 noise combinations. Two cases were ignored i.e. when the motion and complexity both were high and both were low, and location was not accounted for. We ignored location in both of these cases because once motion and complexity are both high or both are low; location will not affect the quality. Figure 4.1 shows the possible combinations after ignoring these two cases. HHI and HHO effectively became HHI whereas LLI and LLO became LLI.

Table 6.2: Noise Factors and Levels.

Noise Factors and Levels		
Noise Parameters	Level 1	Level 2
Motion	High	Low
Complexity	High	Low
Location	High	Low

6.2.2 Test Bed

Setup for this experiment was quite unique as we employed a varied set of equipment. We used desktop computers, tablets and mobile phones for this experiment. Computers used for this experiment were Intel core i5 CPU machines running at 3.60 GHz with 4 GB RAM. Running 64-bit Windows 7 Enterprise with Service Pack 1 installed. These machines were using integrated Intel ourHD graphics. Monitors used were ViewSonic VS 13239 LED 1080p Full HD. The tablet computer was an Asus Tablet TF700T supporting 10.1 1920x1200 resolution Full HD incorporating the Android operating system. This tablet was powered by an NVIDIA Tegra 3 Quad-core CPU 1.6 GHz, whereas for graphics it was using a 12-core ULP GeForce GPU. A Samsung Galaxy S-III powered by a 1.4 GHz quad-core Cortex-A9 1.2 GHz was used as the mobile form factor. It supported a Mali-400 MP4 for the GPU, and incorporated a 1 GHz internal memory. The display was a 4.8″ HD Super AMOLED 1280x720. In addition to this hardware we also used an Android based application called Splashtop Personal ©. This application was used for screen mirroring and provided us with the capability to set

Table 6.3: Array with control factors.

L9 Array with control factors			
Exp. No	A	B	C
1	Mobile	1.0 GHz	10 ms
2	Mobile	2.0 GHz	300 ms
3	Mobile	3.0 GHz	500 ms
4	Tablet	1.0 GHz	300 ms
5	Tablet	2.0 GHz	500 ms
6	Tablet	3.0 GHz	10 ms
7	Flat Screen	1.0 GHz	500 ms
8	Flat Screen	2.0 GHz	10 ms
9	Flat Screen	3.0 GHz	300 ms

values for the processor and buffer size parameters centrally. For customer premises processing we assumed that the local network was wired and was sufficient for HD video streaming. Ideally we wanted to keep all the communication over a wired medium. For this experiment we were required to test different screen sizes, processing power and buffer size. Complexity in the experimental setup was due to the presence of multiple operating systems and multiple pre-configured machines. By using standard Taguchi design, we came up with different combinations of display size, processing power and buffer size. Most of these combinations were not possible with the devices available. We decided to use a software based solution i.e. to use Splashtop Personal ©. Our streaming server was running Ubuntu 12.04 LTS (Precise Pangolin). We installed an Ubuntu application for Splashtop Personal © and client side applications for the Android Tablets and mobiles. Each tablet and mobile phone was able to recognize the streaming server, available over the local network, and was able to connect. After connection we were able to see the mirrored display over the tablets and mobile phones. We were able to configure the processing power within Ubuntu using the CPULIMIT package. The buffer size was changed by using VLC player.

6.2.3 Selection of Test Sequences

The test sequences were extracted from HD quality videos of Blu-ray movies. These were then compressed using H.264 with a full HD resolution of 1920 * 1080. The content was selected from different movies encompassing many different genres. For this work, the degradation under study was customer premises processing. No artefacts were introduced in the video and they were kept at the benchmark quality. Altogether, 54 clips were used for the experiment and there were 6 categories of content type. For each condition as per Table 6.3, every observer was shown all 54 clips in groups of 6 videos.

Each group had one video of each category. This was for fulfilling the requirement of the Taguchi design so that the experiment would be repeated for each noise factor.

6.2.4 Test Layout

Test layout for this experiment was similar to the one described in Section 4.3.4.

6.2.5 User Task

User task for this experiment was similar to the one described in Section 4.3.4.

6.2.6 Customization

We mentioned the customization in the experimental setup earlier. For our experiment, we also were required to customize the experimental process. Randomness of experimental design is essential for protecting against unknown variable effects. We needed to make special arrangements for this experiment to make the experimentation process randomized. As mentioned in the setup section, we placed the devices around the volunteers in a manner that ensured that they were not required to leave their seats and were only required to change direction towards the currently used display. Even these changes were not too frequent. Ample break time was provided to help the volunteers settle and feel comfortable with the environment after each change of direction.

Customer premises processing involves a large number of factors/parameters that have an effect on the perceived quality. The processing done on customer premises is supposedly far less than processing done during creation of content or compression. This enables thin devices having minimal capabilities and power requirements [127]. Two of the most important jobs performed at the customer premises are decoding and playing the content. Decoding is not very processor intensive and usually takes only 1/4 of any encoding processing power requirements. Even then, having a slower processor might affect the perceived video quality [128]. Also, for most mobile devices there is a dynamic control, triggered on excessive heat generation, for CPU/GPU. This dynamic control can affect the amount of processing power that is available for the decoding process [129][130].

Display size and technology are also important factors having an effect on perceived visual quality. There is a wide range of displays sizes and technologies in use today. Tablets and mobiles are full multimedia enabled and usually support full HD screens. Due to technological advances, display quality is improving rapidly. These new displays provide better viewing angles of up to 178 °, sharp display and better viewing experiences - even in outdoor situations. These displays offer higher resolutions, frame refresh rates and are able to handle high complexity scenes. It's been reported by researchers

that these different display technologies introduce somewhat different artefacts [52]. Some work was done to find a relationship between different display technologies especially the old CRT monitors vs newer LCD monitors, [53] [54]. Investigations were done to find distortion effects due to display size [55]. The researchers stated that distortion is the predominant factor when comparing HDTV as compared to SDTV. It's been reported that low distortion improves quality perception but high distortion reduced quality perception in the case of large size displays. Pechard et al. [56] in his work concluded that SAMVIQ [131] is more accurate for higher resolutions. Larger displays were also studied for identification of artefacts affecting them [57]. Our aim in this experiment was to identify which form factor enables better perceived video quality. It will help us identify which technology is the preferred technology in the context of perceived video quality. As far as the receiver buffer size is concerned it plays an important role in video playability according to [132]. The work done by [133] introduced receiver buffer for smoothing out video output. They commented that by introducing such a buffer they made a trade-off between short-term improvement and long-term smoothing of quality.

6.3 Results and Analysis

In an effort to establish the validity of the user feedback we deployed the following methods to verify the non-randomness of feedback and general consensus among observers regarding the video quality as discussed in detail in the Section 4.4. The following methods were used to establish these facts:

6.3.1 Pairwise Correlation

The first analysis that we performed was to validate the feedback recorded from users. The feedback can only be considered valid if it is not random. Our aim was to capture perceptive quality feedback and, as all the observers were assessing the same media, then the scores should not be random. For this purpose, we calculated the correlation between pairs of observers, i.e. each observer's score was checked against every other observer for each video. The results showed that for 85.8% the pairwise correlation was higher than 0.5 and 50% were over 0.6, the lowest correlation was 0.43. Figure 6.1 shows the plot of correlations between pairs of observers. Correlation among pairs of observers was on the lower side and we wanted to investigate what was the level of concordance among the group.

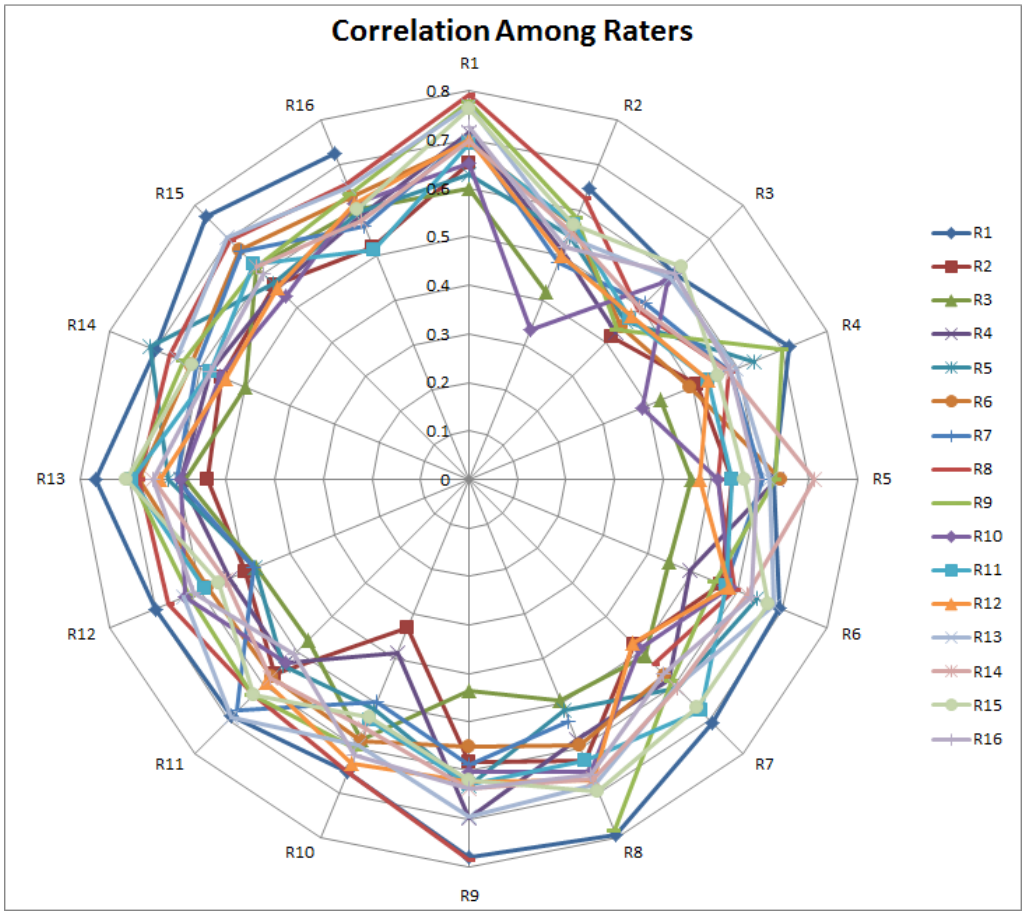

Figure 6.1: Correlation among raters

6.3.2 Fleiss´ Kappa

Fleiss´ Kappa was used for finding out concordance (agreement) between groups of observers [89]. It enables us to calculate concordance among any number of observers for a fixed number of items. We used Equation 3.1 for calculating kappa i.e. k. Table A.2 shows the values calculated for the equation. The value of k for 16 observers obtained was 0.220 which indicates that there was fair agreement among the group of observers. A table proposed by [90] was used for interpretation of the results. The factor $1 - \bar{P}_e$ gives the degree of agreement that is attainable above chance, and the factor $\bar{P} - \bar{P}_e$ gives the degree of agreement actually achieved above chance. Where k=1 is absolute agreement and k=0 means no agreement at all.

6.3.3 Signal to noise ratio analysis

The observers' score was recorded and the mean was calculated for all users on a per video basis. Later, each category of noise factor was combined to create Table 6.5. It

Table 6.4: Content Effect Table.

Effect Table			
Level	A (Display)	B (Processor)	C (Buffer)
1	12.41	12.57	12.86
2	12.71	12.36	11.37
3	11.90	12.08	12.79
Δ	0.82	0.50	1.49
Rank	2	3	1

Table 6.5: MOS scores per condition per trial.

MOS scores per condition per trial							
Exp. No	T1	T2	T3	T4	T5	T6	Mean
1	3.13	4.36	4.69	4.71	4.38	4.19	4.24
2	2.31	3.31	4.13	4.19	4.44	4.44	3.80
3	4.25	4.75	4.25	4.69	4.44	4.69	4.51
4	4.44	4.50	3.06	4.13	4.69	4.38	4.20
5	4.63	4.31	4.44	4.69	4.19	3.31	4.26
6	4.56	4.63	4.38	4.50	4.69	4.38	4.52
7	4.63	4.50	4.44	4.13	3.69	4.56	4.32
8	4.00	4.69	4.50	4.31	4.44	4.81	4.46
9	1.56	3.38	2.81	2.94	4.19	4.13	3.17

shows conditions in the rows and noise factors in the columns. The mean for all the noise factors combined was also calculated for further analysis. We want to increase the perceived quality whenever possible; hence the aim of the analysis is to find out the parameters that affect the perceived quality. We have computed the signal to noise ratio (SNR) for each condition by using the *larger the better formula* Equation 4.1 and created a response chart, and determined the parameters that have the highest and lowest effect on the perceived quality.

Once the SNR ratio is calculated for each trial i.e. for each noise factor type, we came up with the response chart shown in Table 6.6. Response chart shows the levels for each parameter and respective SNR. In order to calculate the effect of a parameter we wanted to find out the range of its effect. So, from Table 6.6, we generated the effect table i.e. Table 6.4. Table 6.6 also shows calculation of the effect of SNR for parameter C where its level was 2. To populate the effect table for level two of parameter C i.e. the buffer, we took the mean of all three SNR values pertinent to level 2. This became the level 2 value for the buffer. Once all the values were calculated we calculated Δ which was the difference between the maximum SNR and the minimum SNR as shown

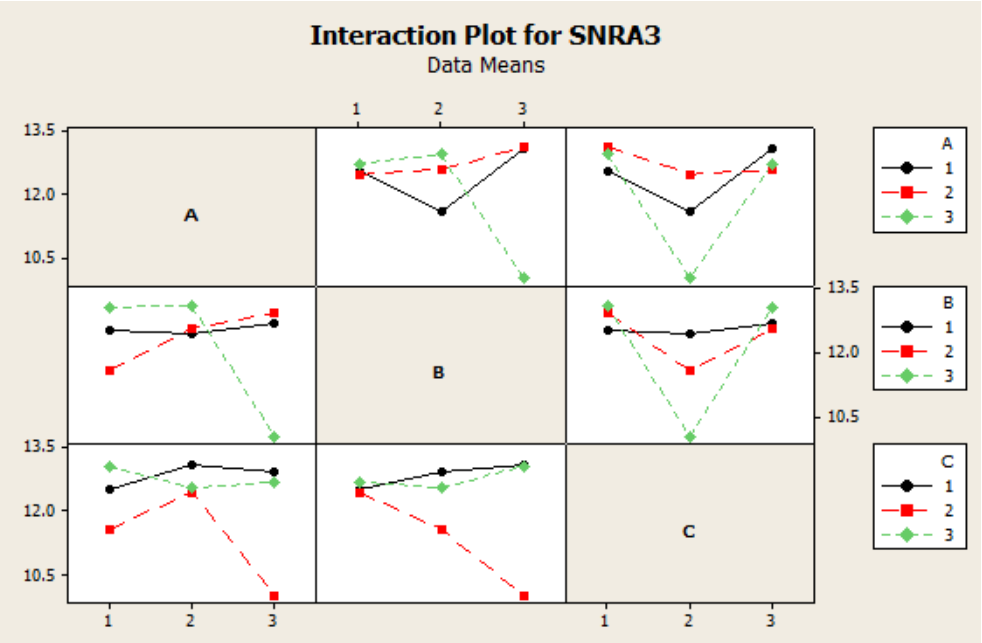

Figure 6.2: Interaction Plot

in Table 6.4. This gave us the range of effect for each parameter where we found that the larger the value, the greater the effect of that parameter on quality. We ranked the parameters as per their range of effect.

From Table 6.4 we can identify that the buffer parameter had the biggest value of Δ whereas the display parameter came second while the processor parameter was the least effective. An interaction plot was generated to show the effects of the buffer size parameter in relation to other parameters Figure 6.2. As we already discussed in the previous chapters that most models for video quality included parameters from content domain and network domain. Customer premises domain was neglected in these multi-parameter models. Though studies were done in isolation to find out the effect of few of these parameters on the end to end quality. Our experiment shed some light on the parameters from the CPP domain and highlights the importance of these parameters for inclusion in developing video quality models. Graphical representation of interaction between all three parameters is shown in Figure 6.2.

From the interaction diagram we can see that by selecting different levels for the parameter buffer, the SNR value decreased. The plot of interaction between the parameter processor and the parameter buffer we can see that for level two of buffer i.e. 300 ms the SNR score decreases. The same is true for the plot between the display and the buffer parameters. Usually 300 ms is the default value within the VLC. The perceptive video quality is very low for conditions where the value for the parameter buffer is 300 ms. In the interaction plot between the display and the processor parameters we can

Table 6.6: Response Table.

Response Table			
A	B	C	SN
1	1	1	12.55
1	2	2	11.60
1	3	3	13.08
2	1	2	12.47
2	2	3	12.57
2	3	1	13.10
3	1	3	12.71
3	2	1	12.93
3	3	2	10.05

see that using a flat screen 24″ reduces the perceptive video quality. This could be due to the fact that large displays are susceptible to enhancing the artefacts more than the smaller displays.

The best SNR score was for condition 6 which used a 10.1″ Tablet with a 3 GHz processor and 10 ms of buffer. The worst condition was condition 9 which used a 24″ LED monitor with a 3 GHz processor and 300 ms of buffer. Most other conditions were almost similar and produced SNR scores quite close to each other.

From these results we can infer that tablets with their advanced display technology and better form factor capture better perceptive video quality. At the same time, buffer size selection is important for perceptive video quality. Processing power which was represented by the processor parameter was not affecting QoE as even the minimal processor in our study was good enough to decode and play the video.

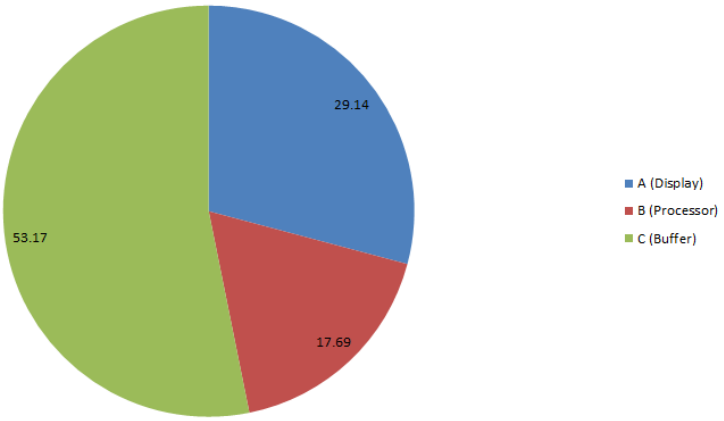

Figure 6.3: Pie Chart of influential factors

Table 6.7: Predicted SNR values for optimal configuration

Predicted Values				
	Configuration	PSNR	PMEAN	PSTDE
1	A 2, B 1, C 1	13.72	4.65	0.25
2	A 2, B 1, C 2	11.44	3.97	0.72
3	A 2, B 1, C 3	13.65	4.62	0.27

Optimal Parameter Configuration

From the SNR analysis we now know that buffer size is the most influential parameter. We want to find out the optimal configuration of all the parameters which can improve SNR. For this purpose we analyze the main effect plot of Signal to Noise Ratio (SNR) and identify the level of each parameter which enhance the SNR. Figure 6.4 show the main effect plot for SNR with highlighted parameter levels. Red circles identify the settings which will ensure high SNR. Combination of these parameter levels will become the optimal setting for CPP domain. In order to verify our conclusion we make use of the predict the SNR values on the basis of these configurations and some other variations. Our selected setting should produce the largest SNR value. Table 6.7 highlight the optimal setting as well as show the predicted values for different combinations. In Table 6.7 PSNR means predicted SNR, PMEAN is predicted mean and PSTDE is predicted standard deviation. Configuration highlighted is the optimal configuration identified from the Figure 6.4. We can see that the predicted SNR is the highest for this configuration. Predictions for generating Table 6.7 were done by using prediction option within DOE menu of Minitab16.

6.3.4 Quality Function Deployment

For HoQ analysis we used the same three parameters i.e. display, processor and buffer. The customer demands were the same as mentioned in the Table 6.8. The same feedback score was used for the analysis. An expert was asked to provide technical feedback with reference to customer demands. Rows of Table 6.8 shows the customer demands and importance rating (I) given by volunteers, whereas the columns contain the corresponding expert feedback. Weights were calculated using customer feedback and expert feedback. These weights were calculated by using Equation 6.1. The results are shown in Table 6.9. According to HoQ, the buffer parameter was identified as the most important parameter, whereas the display parameter was the second most important parameter. The processor parameter came in as the least important parameter. Parameter importance was plotted using the pie chart shown in Figure 6.5.

Table 6.8: HoQ with raters and experts feedback.

HoQ matrix showing customer demands and experts feedback				
User Expectations/Demands	Importance Rating	Display	Processor	Buffer
Smooth Video (No Jitter)	5	3	0	9
Video with No Blockiness (Not pixelated)	3	1	1	1
Video with No blurring	3	1	1	1
Synchronization of voice and video	4	3	1	1
Meets acceptable quality (Overall)	4	3	3	3
Time to load before playback	3	3	3	9
Acceptable Video Quality for high motion (Camera panning or fast moving content)	5	9	3	3
Max Video Quality for low motion (Head and shoulder)(Fixed camera)	2	3	3	3
Acceptable Video Quality for Complex scenes (Many objects/lots of colors/Details)	5	3	3	3
Acceptable Video Quality for simple scenes (Few object/few colors/Not much details)	3	3	3	9
Continuity of service	5	1	3	1

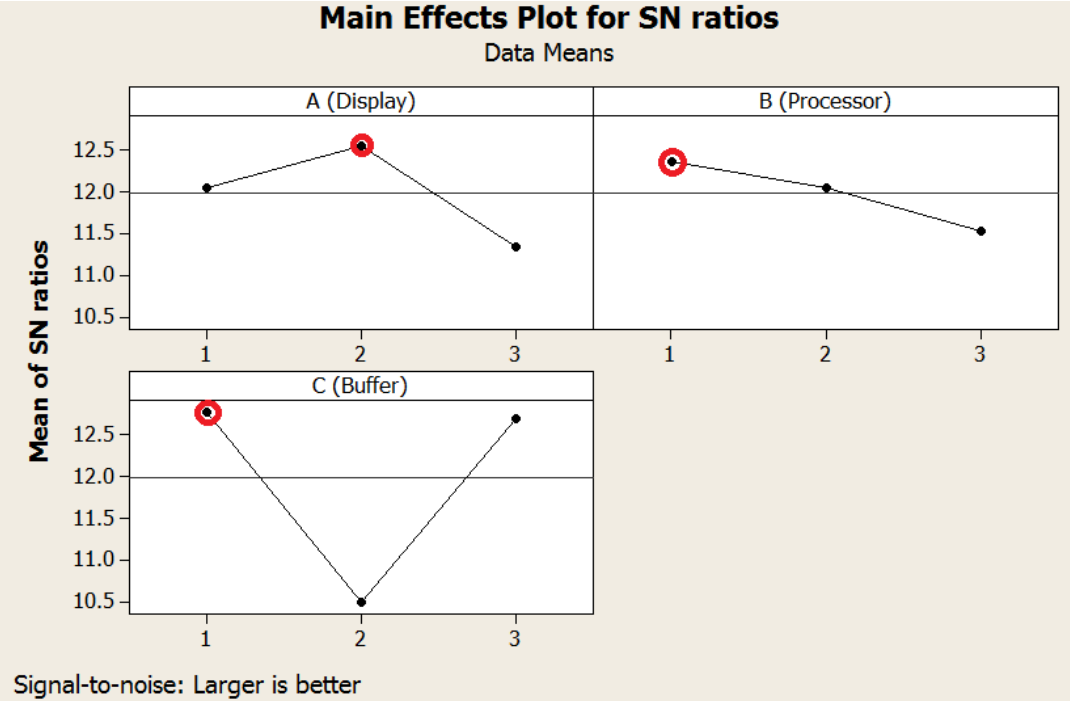

Figure 6.4: Optimal parameter configuration for CPP domain

Table 6.9: Calculated Weights for HOQ.

Ranking of Parameter by Calculated Weights			
	Buffer	Display	Processor
	1	2	3
W_i	171	78	62

The Likert scale used represents the most important demand represented by "5" and "1" representing the least important demand. We collected technical characteristics data from domain experts.

$$W_j = \sum_{i=1}^{r} M_{ij} \times I_i \tag{6.1}$$
$$i = 1....r, j = 1....c$$

6.3.5 Comparison of Taguchi method and HoQ matrix

Once we compared the results from both methods it became obvious that they rank the parameters similarly. In both methods, the buffer parameter was determined as the most important parameter. The processor parameter emerged as the least important parameter. We also checked the statistical significance of the results generated by

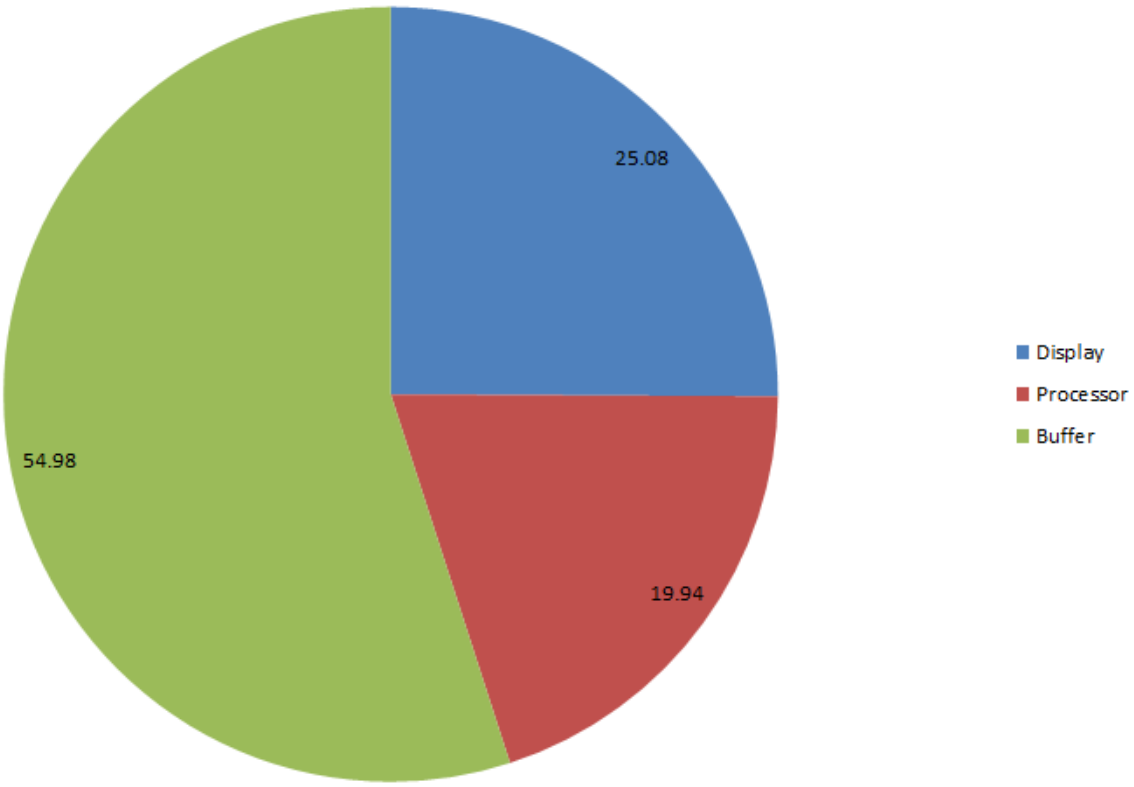

Figure 6.5: Pie Chart of influential factors using HoQ matrix

the HoQ matrix. For this purpose, we used permutation sampling [134]. Table 6.10 shows the statistical significance using p-values. Very small p-values at the 5% level of significance indicate that the buffer parameter is significantly different from the other two parameters. The p-values for the processor and display parameters suggest that there is not much of a difference between these two parameters. The difference between parameter display and the processor parameters are statistically insignificant.

The results obtained by two independent methods using different approaches have provided us with confidence in our results. Both these methods identify the buffer parameter as the most influential parameter affecting QoE. We intend to study this parameter further in developing an end to end QoE model in Chapter 7.

6.4 Summary and Conclusion

In this experiment, we analyzed the influence of customer premises processing domain parameters namely display, processor and buffer size on QoE. We emulated IPTV video on demand functionality for these experiments. These parameters were investigated under the influence of noise factors involving motion, complexity and location. We

Table 6.10: P-value table using permutation sampling.

	Display	Processor	Buffer
Processor	NA	0.7192	0.0078
Display	NA	NA	0.0326
Buffer	NA	NA	NA

identified buffer size as the most influential control parameter. Configuring the buffer parameter with the optimum settings will ensure that the effect of content variation will be the least. We also identified the optimum configurations for the customer premises processing domain. The other two parameters under investigation were the processor and the display parameters. They affected video quality equally and statistically there was not a significant difference between the affect they had on perceptive video quality. We were able to verify results generated by Taguchi method with the HoQ matrix. Both methods ranked the parameters in the same order.

Chapter 7

End to End Study of Influential Parameters

7.1 Introduction

In order to consolidate the research performed on the individual domains discussed in Chapters 4,5, and 6, we selected the most influentual parameters for each domain to develop an integrated view from end-to-end of the service delivery chain. In addition, we wanted to investigate the order of effect on QoE with respect to all these parameters. Once again, these effects were studied by varying parameters between low, middle and high values. We selected quantizer, packet reordering, and packet loss, jitter and buffer size as our control parameters.

From the discussion in previous chapters it's obvious that each of these parameters affects video quality. We conducted subjective experiments for video quality assessment. Observers were required to rate the perceptual quality of each video. We considered the ITU P.910 recommendations for conducting this experiment. From the literature that we reviewed we were unable to identify a model which incorporated parameters from all three domains. Most parametric models developed for video quality measurements ignored quantization from content, packet reordering from network domain and buffer size from the CPP domain. We included five parameters from across all three domains. Our results indicate that QoE is affected by the packet loss - more than any other parameter from all of the three domains. Quantization also affects QoE considerably. Buffer size was found to be the third most influential parameter. We were not able to identify any earlier QoE model in the literature that incorporated buffer size with any other parameters.

Table 7.1: Control Factors and Levels.

Control Factors and Levels				
Control Parameters	Labels	Level 1	Level 2	Level 3
Packet Loss	PL	0.20%	1.00%	1.80%
Quantizer	QUA	12	26	51
Jitter	JIT	10ms	15ms	40ms
Packet Reorder	PR	5%	7%	15%
Buffer Size	BS	10ms	300ms	500ms

7.2 Design of Experiment for end to end influence

This section discuss in detail the experiment which considered all the most influential parameters from all three domain. The test-bed needs to be able to handle parameters from content domain, network domain and CPP domain. The following section discuss the parameter selection.

7.2.1 Parameters

This experiment was designed to test the end to end effect of parameters on QoE. For this the control factors were the most influential parameters from the three domains.

Control Factors

The control factors which were analyzed during this experiment were: quantizer, packet loss, jitter, packet reorder and buffer size. Table 7.1 shows the control factor and levels for each parameter.

Noise Factors

Noise factors considered for this experiment were motion, complexity and location (indoor or outdoor scene). Table 7.2 shows the combination of noise factors. We had 3 noise factors so we considered 2^3 combinations of noise. It was required that we repeat the experiment for each condition with 8 noise combinations. Two cases were ignored i.e. when the motion and complexity both were high and both were low, and location was not accounted for. We ignored location in both of these cases because once motion and complexity are both high or both are low; location will not affect the quality. Figure 4.1 shows the possible combinations after ignoring these two cases. HHI and HHO effectively became HHI whereas LLI and LLO became LLI.

Table 7.2: Noise Factors and Levels.

Noise Factors and Levels		
Noise Parameters	Level 1	Level 2
Motion	High	Low
Complexity	High	Low
Location	High	Low

7.2.2 Test Bed

This experiment required a test bed where we could configure parameter settings for all three domains viz content, network and CPP. For this purpose we setup the streaming server over Ubuntu 12.04 LTS (Precise Pangolin) running on Intel i7 3.6 GHz with 4 GB RAM. The clients used for this experiment were Intel core i5 CPU machines running at 3.60 GHz with 4 GB RAM. Running 64-bit Windows 7 Enterprise with Service Pack 1 installed. These machines were using integrated Intel HD graphics. Monitors used were ViewSonic VS 13239 LED 1080p Full HD. Quantizer values where changed by compressing videos at different levels of quantization. Netem was used for controlling network behaviour. Buffer size values where changed on the VLC application we developed for playing video for the experiment. Taguchi design was used for setting up

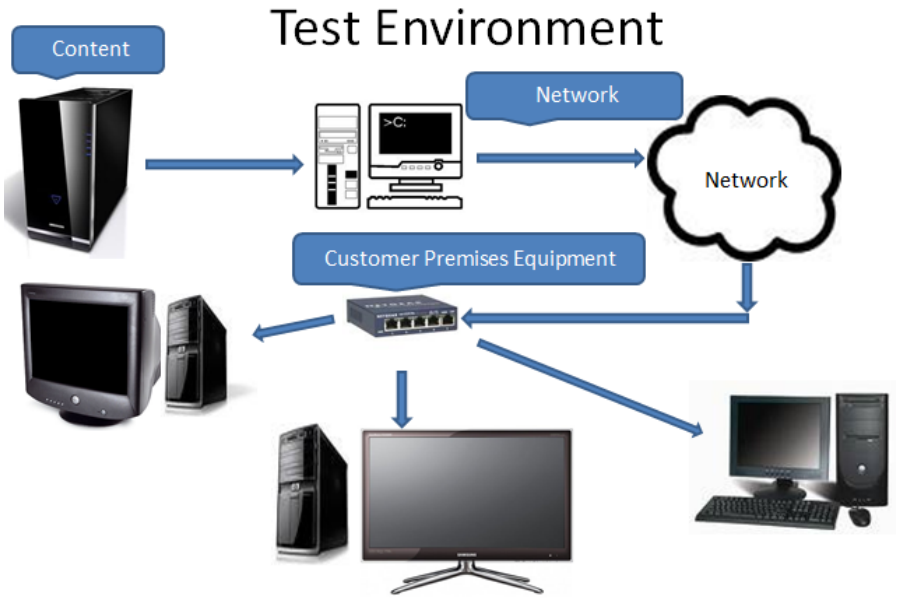

Figure 7.1: Test bed setup

the experiment. We selected the L18 array for 5 parameters each having 3 levels. In order to validate the model we generated 18 additional combinations and recorded the

Table 7.3: Array with control factors.

L18 Array with control factors					
Exp. No	PL	QUA	JIT	PR	BS
1	0.20%	12	10ms	5%	10ms
2	0.20%	12	15ms	7%	300ms
3	0.20%	12	40ms	15%	500ms
4	0.20%	26	10ms	5%	300ms
5	0.20%	26	15ms	7%	500ms
6	0.20%	26	40ms	15%	10ms
7	1.00%	51	10ms	7%	10ms
8	1.00%	51	15ms	15%	300ms
9	1.00%	51	40ms	5%	500ms
10	1.00%	12	10ms	15%	500ms
11	1.00%	12	15ms	5%	10ms
12	1.00%	12	40ms	7%	300ms
13	1.80%	26	10ms	7%	500ms
14	1.80%	26	15ms	15%	10ms
15	1.80%	26	40ms	5%	300ms
16	1.80%	51	10ms	15%	300ms
17	1.80%	51	15ms	5%	500ms
18	1.80%	51	40ms	7%	10ms

subjective response of volunteers. Table A.4 shows the factorial design for 5 parameters where each parameter has 3 levels. Column status displays the combinations used by Taguchi design (1-18) as well as the additional 18 combinations (A1-A18) for validation purposes.

7.2.3 Selection of Test Sequences

The test sequences were extracted from HD quality videos of Blu-ray movies. These were then compressed using H.264 with a full HD resolution of 1920 * 1080. The content was selected from different movies encompassing many different genres. For this work, the degradation in content was introduced by changing the quantizer values. Videos were compressed according to the combinations selected by L18 array and the additional 18 combinations for validation purpose. Altogether, 216 clips were used for the experiment and there were 6 categories of content type. For each condition as per Table 7.3, every observer was shown all 216 clips in groups of 6 videos. Each group had one video of each content type i.e. the noise factor combination. This was for fulfilling the requirement of the Taguchi design so that the experiment would be repeated for each noise factor.

7.2.4 Test Layout

Noise factors which were considered were motion, complexity and location. After selecting the orthogonal array L18 based on 3 control factors and 3 noise factors. Table 7.3 shows the L18 array with control factors. We randomized the whole experimental process, so that all the videos were presented randomly to the viewers and all the conditions were presented randomly as well. A group of 16 observers volunteered to participate in the experiment. They were screened to confirm that they had no prior experience in video compression or production. Each observer was shown 216 clips where each clip was 10 seconds long. The whole experiment was divided into two equal halves. Volunteers were provided with the option to either complete the whole task in one day or they could take a break of 1 day in between two halves of the experiment. Most volunteers preferred to complete the testing within the same day. There was a compulsory break of 1 hour in between the two halves when the volunteers opted for same day task completion. For each video evaluation, volunteers where provided with approximately 5 seconds for feedback selection. The users were required to select a level of quality, based on their perception, between excellent, good, fair, poor and bad. As per their selection, a relative score was recorded. Approximately 30 seconds was required for completing one assessment. From previous experiments we had learned that volunteers would like to take breaks only once they needed them. Otherwise compulsory breaks reduced their interest and become a source of frustration as they wanted to continue with the task. Hence, we provided volunteers with the option of taking a break after a segment of 6 assessments was complete. Almost all of the volunteers did not take a break in the first 30 minutes of the experiment. They were allowed 3 breaks each of 2 minutes. Not a single volunteer opted for more than 2 breaks. On average the whole experiment was completed within 60 minutes. 5 minutes were spent in explaining the experimental procedure to the participants.

7.2.5 User Task

Observers were required to assess 216 videos in approximately 60 minutes. For this experiment the task given to volunteers was to assess the video quality under compression levels, various network conditions and buffer size. They were briefed about the experimentation process. The undistorted videos were played before each distorted video was displayed. Their task was to assess video quality for each situation. Each situation was a combination of 5 different parameters. They were briefed about the experimental process and were not trained concerning artefacts. A 5 level Likert scale was used for this experiment. They were asked to enter their feedback about the perceived quality. The findings will be discussed in Section 7.3.

7.2.6 Customization

Randomness of experimental design is essential for protecting against unknown variable effects. The application we developed ensured that the each video was randomly selected and a reference undistorted video was displayed before each distorted video.

7.3 Results and Analysis

As per the discussion in Section 4.4 there was a need to establish the validity of the user feedback. We want to verify the non-randomness of feedback. Moreover, there was a need to check the general level of consensus among observers regarding the video quality. The following method were deployed for measuring non-randomness and level of consensus among observers:

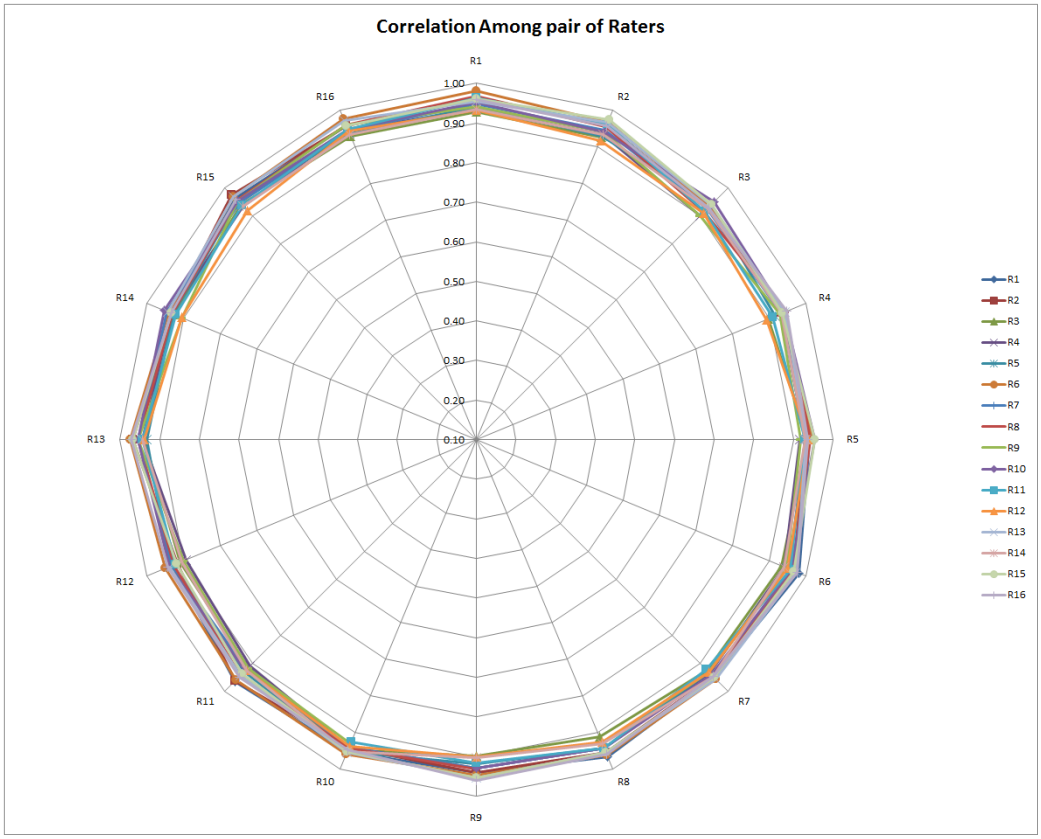

Figure 7.2: Correlation among raters

7.3.1 Pairwise Correlation

We wanted to confirm that the feedback we recorded from volunteers was actually assessing the video quality and was not random. For this purpose, we calculated the

correlation between pairs of observers, i.e. each observer's score was checked against every other observer for each video. The results showed that approximately 0.3% of the pairwise correlations were less than 0.95. Whereas a very high percentage i.e. 97.5% were over 0.95. The lowest correlation value was 0.89. Figure 7.2 shows the plot of correlations between pairs of observers. Correlation among pairs of observers was on the higher side and indicated that pair of raters generally agreed to the level of quality. Further investigation for the level of concordance among the group of raters is discussed in the next section.

7.3.2 Fleiss´ Kappa

Fleiss´ Kappa was used for finding out concordance (agreement) between groups of observers [89]. It enables us to calculate concordance among any number of observers for a fixed number of items. We used Equation 3.1 for calculating kappa i.e. k. Table A.3 shows the values calculated for the equation. The value of k for 16 observers was 0.742 which indicates that there was substantial agreement among the group of observers. A table proposed by [90] was used for interpretation of the results.

Table 7.4: Content Effect Table.

Effect Table					
Level	PL	QUA	JIT	PR	BS
1	7.31	5.02	4.05	4.05	3.47
2	2.19	4.88	2.78	2.92	4.16
3	0.93	0.51	3.59	3.45	2.79
Δ	6.38	4.51	1.27	1.13	1.37
Rank	1	2	4	5	3

7.3.3 Signal to Noise Ratio Analysis

The observers' score was recorded and the mean was calculated for all users on a per video basis. Later, each category of noise factor was combined to create Table 7.5. It shows conditions in the rows and noise factors in the columns. The mean for all the noise factors combined was also calculated for further analysis. We wanted to increase the perceived quality whenever possible; hence the aim of the analysis is to find out the parameters that affect the perceived quality. We have computed the signal to noise ratio (SNR) for each condition by using the *larger the better formula* Equation 4.1 and created a response chart, and determined the parameters that have the highest and lowest effect on the perceived quality.

Table 7.5: MOS scores per condition per trial.

MOS scores per condition per trial							
Exp No	T1	T2	T3	T4	T5	T6	Mean
1	2.19	3.94	2.00	3.00	1.19	3.81	2.67
2	2.00	3.00	2.13	3.88	2.13	1.13	2.38
3	1.13	2.94	3.88	2.06	3.88	3.13	3.02
4	4.00	4.06	4.00	4.00	4.88	3.31	4.06
5	2.13	3.00	1.94	1.00	2.94	2.00	1.98
6	4.13	2.94	3.94	2.19	2.13	2.81	2.38
7	1.00	1.06	1.06	1.13	1.00	1.00	1.04
8	1.00	1.13	1.00	1.00	1.00	1.13	1.04
9	1.00	1.19	1.06	1.00	1.25	1.00	1.08
10	1.81	2.06	1.00	2.13	1.13	1.19	1.48
11	2.00	1.13	2.13	2.25	2.00	1.31	1.85
12	2.19	2.00	2.00	2.00	2.19	1.19	1.79
13	1.00	1.13	1.81	1.06	2.00	1.13	1.40
14	1.06	1.13	2.00	1.00	1.19	1.06	1.08
15	1.13	1.13	1.13	1.06	1.13	1.00	1.06
16	1.88	1.06	1.13	1.13	1.06	1.13	1.10
17	1.06	1.06	1.06	1.06	1.06	1.06	1.06
18	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Once the SN ratio is calculated for each trial i.e. for each noise factor type, we obtained the response chart shown in Table 7.6. This response chart shows the levels for each parameter and respective SNR values. In order to calculate the effect of a parameter we wanted to find out the range of its effect. So, from Table 7.6, we generated the effect table i.e. Table 7.4. Table 7.6 also shows calculation of the effect of SNR for parameter PL i.e. packet loss where its level was 2. To populate the effect table for level two of parameter PL, we took the mean of all three SNR values pertinent to level 2. This became the level 2 value for the buffer. Once all the values were calculated we calculated Δ which was the difference between the maximum SNR and the minimum SNR as shown in Table 7.4. This gave us the range of effect for each parameter where we found that the larger the value, the greater the effect of that parameter on quality. We ranked the parameters as per their range of effect.

From Table 7.4 we can identify that the parameter packet loss had the biggest value of Δ whereas the quantizer parameter came second while the packet reorder parameter was the least effective. An interaction plot was generated to show the effects of the parameter packet loss in relation to other parameters. The Figure 7.3 presents the interaction effect of all the parameters.

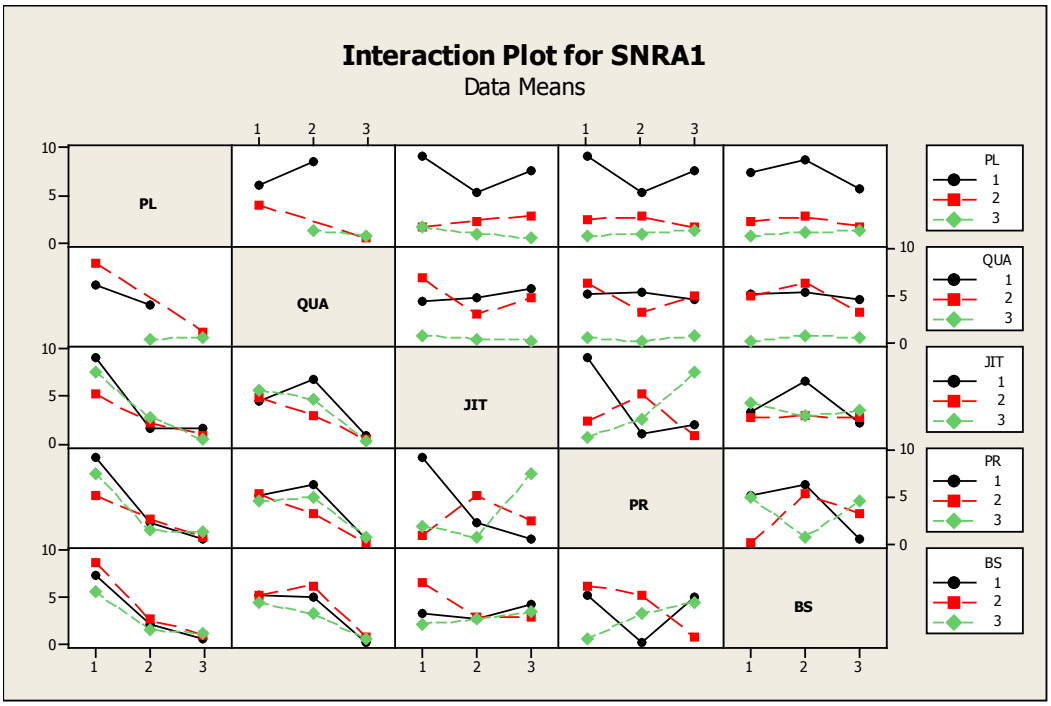

Figure 7.3: Interaction Plot

From the interaction diagram we can see that by selecting higher packet loss and lower quantization, the video quality fell to unacceptable levels. Interactions between the parameters packet loss and quantizer parameter with any other parameter took the output quality to unacceptable levels. The patterns are the same and hence indicate that these two parameters affect QoE more than the other three parameters. Level 2 and level 3 settings for the packet loss parameter reduced the output quality to unacceptable levels irrespective of the level selected for the other parameters. In the case of the quantizer parameter, once we select level 3, the output quality goes down to unacceptable levels irrespective of the levels of the other parameters. Interaction between buffer size and jitter indicate that selecting the highest value for buffer size we see a marginal increase in SNR.

The best SNR score was for condition 4 which introduced 0.2% of packet loss, quantizer value of 26, jitter of 10ms, 5% packet reordering and 300ms of buffer size.

From these results we can infer that higher packet loss and lower quantizations will affect QoE and especially if they are both set to their lowest level in parallel. Buffer size should be set to accommodate the application in use. Otherwise, large buffering will increase latency in play and less buffering will not be able to correct the jitter effects.

Table 7.6: Response Table.

Response Table					
PL	QUA	JIT	PR	BS	SN
1	1	1	1	1	6.28
1	1	2	2	2	5.58
1	1	3	3	3	6.39
1	2	1	1	2	11.97
1	2	2	2	3	4.85
1	2	3	3	1	8.77
2	3	1	2	1	0.33
2	3	2	3	2	0.31
2	3	3	1	3	0.59
2	1	1	3	3	2.66
2	1	2	1	1	4.18
2	1	3	2	2	5.05
3	2	1	2	3	1.74
3	2	2	3	1	1.23
3	2	3	1	2	0.75
3	3	1	3	2	1.32
3	3	2	1	3	0.53
3	3	3	2	1	0.00

7.3.4 Quality Function Deployment

For HoQ analysis we used the same five parameters i.e. Packet Loss, Quantizer, Jitter, Packet Reorder and Buffer size. The customer demands were the same as mentioned in Table 7.7. The same feedback score was used for the analysis. Expert feedback acquired for the previous experiments was used and the required fields were extracted for use in this experiment. Table 7.7 shows the customer demands, the importance rating given by volunteers and corresponding expert feedback. Weights were calculated using customer feedback and expert feedback. These weights were calculated by using Equation 7.1. The results are shown in Table 7.8. According to HoQ, the buffer size parameter was identified as the most important parameter, whereas the Quantizer parameter was the second most important parameter. Jitter, Packet Loss and Packet Reorder were 3^{rd},4^{th} and 5^{th} respectively. Parameter importance was plotted using the pie chart shown in Figure 7.5.

The Likert scale used represents the most important demand represented by 5 and 1 representing the least important demand. We collected technical characteristics data

Table 7.7: HoQ with raters and experts feedback.

HoQ matrix showing customer demands and experts feedback						
User Expectations/Demands	Importance Rating	PL	QUA	JIT	PR	BS
Smooth Video (No Jitter)	5	1	3	9	0	9
Video with No Blockiness (Not pixelated)	3	9	1	1	3	1
Video with No blurring	3	9	1	1	3	1
Synchronization of voice and video	4	0	3	1	9	9
Meets acceptable quality (Overall)	4	9	3	3	9	3
Time to load before playback	3	1	3	9	9	9
Acceptable Video Quality for high motion (Camera panning or fast moving content)	5	0	9	3	3	3
Max Video Quality for low motion (Head and shoulder)(Fixed camera)	2	9	3	3	0	3
Acceptable Video Quality for Complex scenes (Many object/lots of colors/Details)	5	9	3	3	0	3
Acceptable Video Quality for simple scenes (Few object/few colors/Not much details)	3	3	3	9	0	3
Continuity of service	5	0	1	1	0	0

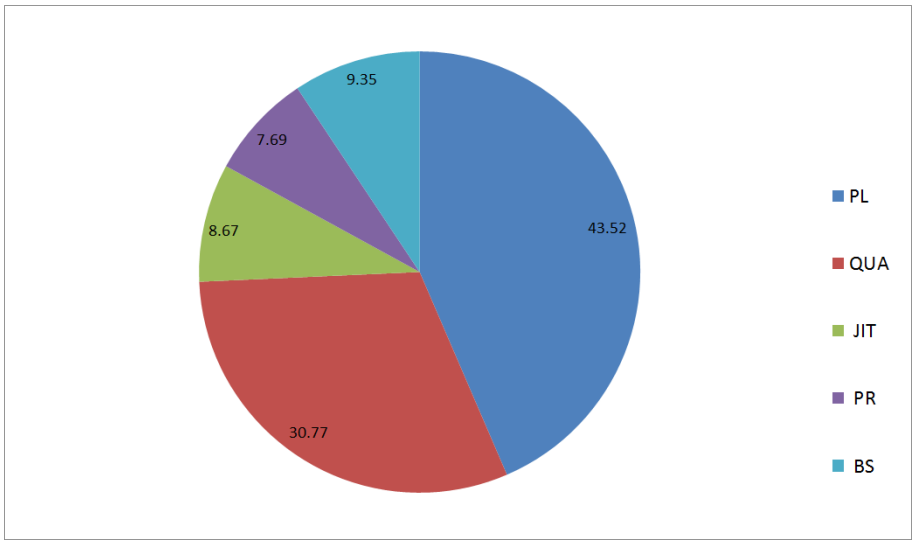

Figure 7.4: Pie Chart of influential factors

Table 7.8: Calculated Weights for HOQ.

Ranking of Parameter by Calculated Weights					
	Packet Loss	Quantizer	Jitter	Packet Reorder	Buffer Size
	4	2	3	5	1
W_i	134	170	162	132	171

from domain experts.

$$W_j = \sum_{i=1}^{r} M_{ij} \times I_i \tag{7.1}$$
$$i = 1....r, j = 1....c$$

7.3.5 Comparison of Taguchi Method and HoQ Matrix

For this experiment the ranking from the HoQ method was different when compared to the ranking achieved by SNR analysis. Quantizer and Packet Reorder were ranked similarly by both methods whereas Jitter was different by only one position. The major difference was in the ranking of Packet Loss and Buffer Size. We investigated the findings of the HoQ method further and checked the statistical significance of the results generated by the HoQ matrix. For this purpose, we used permutation sampling [134]. Table 7.9 shows the statistical significance using p-values. We found larger p-values at the 5% level of significance which indicated that there is not much of a difference between these parameters. These findings indicate that according to the HoQ method

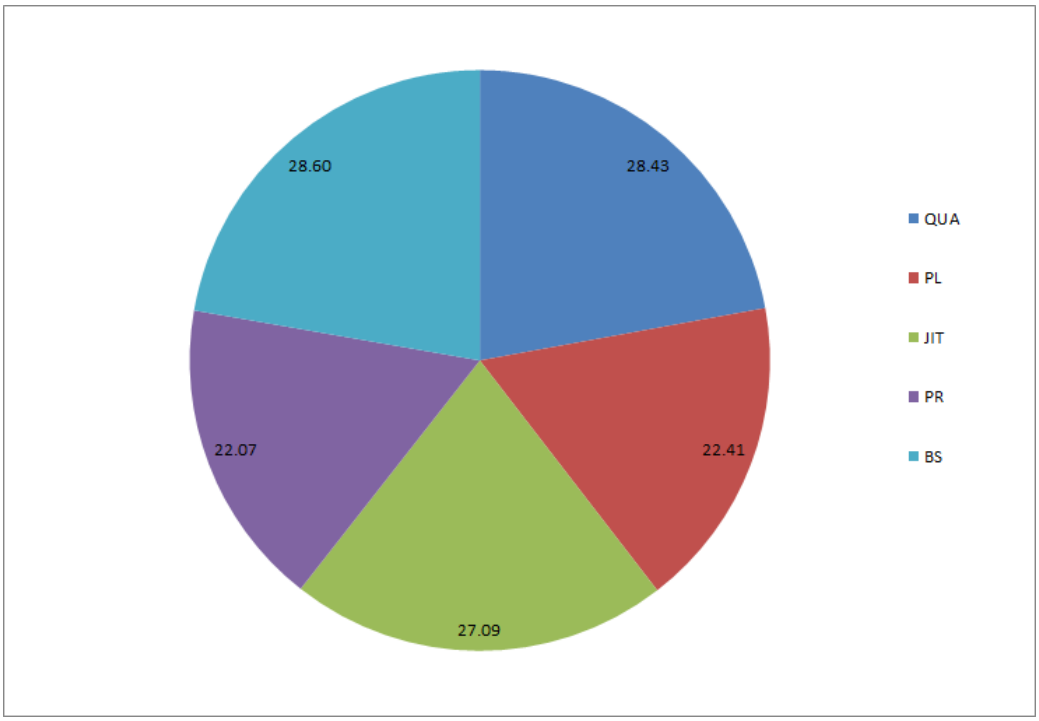

Figure 7.5: Pie Chart of influential factors using HoQ matrix

all these parameters are almost equally affecting QoE. While looking at the results obtained by SNR it's obvious that two parameters have more effect on QoE than all others combined. These unmatched results could be due to the inability of HoQ to assess proper ranking when most of the parameters involved were all important. We were able to identify that HoQ method works fine in normal conditions where few parameters are significantly important then others but once we use it recursively with parameters which were already identified as being influential in the previous runs of HoQ method, HoQ method fails to identify the difference between such influential parameters. The reason seems to be due to the weights assigned on the basis of the importance. Once all the parameters are assigned almost equal importance the difference between them become statistically insignificant. All the parameters included in this experiment were the most influential parameters from their respective domains. A further study is required to investigate this matter.

7.4 Summary and Conclusion

In this experiment, we analyzed the influence of parameters namely Packet Loss, Quantizer, Jitter, Packet Reorder and buffer size, from all three domains, on QoE. We emulated IPTV video on demand functionality for these experiments. These parameters

Table 7.9: P-value table using permutation sampling.

	Buffer Size	Quantizer	Jitter	Packet Loss	Packet Reorder
Buffer Size	NA	0.967	0.6422	0.5526	0.5428
Quantizer	NA	NA	0.6632	0.5826	0.5664
Jitter	NA	NA	NA	0.8972	0.8940
Packet Loss	NA	NA	NA	NA	0.9812
Packet Reorder	NA	NA	NA	NA	NA

were investigated under the influence of noise factors of motion, complexity and location. We identified packet loss as the most influential control parameter. Controlling the packet loss parameter over the network and keeping its value down should enable us in obtaining a better QoE score. In Chapter 5 we noted that network parameters affect quality more than parameters from the content domain. The ranking confirms this statement as the parameter packet loss is the most influential parameter and the other parameters from within the network domain also are seen to have some effect on QoE. The second most important parameter according to SNR analysis was the Quantizer. Quantizer (QP) values should be kept lower to ensure a better QoE score.

Chapter 8

Modeling & Discussion

8.1 Chapter Overview

This chapter discusses the model development and the discussion regarding the predictability of the model. We have developed an ordinal logistic regression model for predicting video MOS. A minimum model was also developed for low complexity implementations - although there is a compromise with regard to accuracy concerning the minimum model. An independent data set was used for model validation. The results show that the model was able to predict the variations in video MOS for the independent data set.

8.2 Model Development

We are interested in using statistical models to help us predict user QoE on the basis of objectively measurable parameters. These parameters are the ones identified as most influential parameters from all three domains of content, network and CPP. By employing a statistical model we shall be able to extract information from the data we captured from volunteers in the form of MOS. The model can help us in predicting trends and behaviour patterns. A predictive model relies on capturing relationships between explanatory variables and the predicted variables from past occurrences, and exploiting them to predict the unknown outcome. For developing such models we shall consider the data collected in Chapter 7.

8.2.1 Ordinal Logistic Regression

Ordinal logistic regression is a model suitable for ordinal dependent variables [135]. It belongs to the general family of logistic regression models. It can only be applied to data which satisfies the proportional odds assumption that the relationship between any two pairs of outcome groups is statistically the same.

There are several ordinal models, including the proportional odds, partial proportional odds, continuation-ratio and ordinal logistic models. We decided to use the ordinal logistic regression model i.e. on the basis of reasons discussed in Section 3.4. The data used for this activity was collected as described in Chapter 7. The data was rearranged as shown in Table A.6. In Table A.6 the column headed by "Mean" contains the MOS for each video under the influence of each control condition. Whereas the column "motion" contains values between 1 and 6. These are coded values to represent each noise factor. For the ordinal logistic regression model we used the column headed " Group1". This column was generated from the MOS values. Each MOS value i.e. column mean was transformed into respective group1 value by taking the ceil of each respective value of column mean.

Model OLR1

We used Minitab 16 for implementing ordinal logistic regression. For ordinal outcomes, we can use ordinal logistic regression. It relies on the cumulative logit and models the predicted probability of multiple outcomes. As we considered MOS to be on the following scale

1= Worst

2= Bad

3= Fair

4= Good

5= Excellent

Ordinal logit model has the form:

$$logit(p_1) \equiv log\frac{p_1}{1-p_1} = \alpha_1 + \beta' x$$

$$logit(p_1 + p_2) \equiv log\frac{p_1 + p_2}{1-p_1-p_2} = \alpha_2 + \beta' x$$

$$\vdots$$

$$logit(p_1 + p_2 + \ldots + p_k) \equiv log\frac{p_1 + p_2 + \ldots + p_k}{1-p_1-p_2-\ldots-p_k} = \alpha_k + \beta' x$$

$$\text{and } p_1 + p_2 + \ldots + p_{k+1} = 1$$

Each equation models the odds of being in the set of categories on the left versus the set of categories on the right. Ordinal logistic regression (OLR) provides only one set of coefficients for each independent variable.

Table 8.1: Response Information

	Pooled Categories	Compared to	Pooled Categories
θ_1	1		2 3 4 5
θ_2	1 2		3 4 5
θ_3	1 2 3		4 5
θ_4	1 2 3 4		5

Table 8.2 displays the number of missing observations and the number of observations that fall into each of the response categories. Table 8.3 displays the estimated coefficients, standard error of the coefficients, z-values and p-values. For the logit link function, we also see the odds ratio and a 95% confidence interval for the odds ratio.

- The values Const(1), Const(2), Const(3) and Const(4) are estimated intercepts for the logits of the cumulative probabilities of MOS score on a Likert scale as stated above.
- The coefficient of 3.284 for PL is the estimated change in the logit of the cumulative MOS level probability when PL at level 2. All other covariates being constant. Due to a small p-value we can say that PL has an effect upon MOS and overall QoE for video.
- The coefficient of 3.626 for PL at level 3 also explains the estimated change in the logit of the cumulative MOS level probability when PL is 3 compared to PL being 1, with other covariates constant. Again, due to a small value, we can say that PL at level 3 has got an effect upon MOS and overall QoE for video.
- The coefficient of 0.269 for QUA at level 2 came out to be insignificant as the p-value is higher than 0.05. This was also reflected in our analysis in previous chapters using SNR. Quantization at level 2 i.e. at value 26, still produced results either better than acceptable or at least acceptable.
- The coefficient of 2.314 for level 3 of QUA and a p-value lower than 0.05 indicate that QUA at level 3 affect MOS.
- JIT at level 2 is not significant at α-level of 0.05 but we should still consider it significant as the p-value is close to 0.05.
- The coefficient of 1.688 for PR at level 2 with a low p-value indicate that it has got some effect on MOS for video quality.
- P-values for PL and QUA is less than 0.05, indicating that there is sufficient evidence that the coefficients are not zero using an α -level of 0.05. The p-values

for JIT, PR and BS are higher than 0.05 which indicates that there is no evidence to suggest these parameters have an effect on QoE.

- We realize that for odds ratio, as we select a higher level of parameter, a positive beta (coefficient) means higher odds of a lower ordered category (MOS). We can see from Table 8.3 that PL and QUA reported higher odds ratio values which indicate that higher levels of these parameters will increase the odds of producing lower quality videos.

Table 8.2: Response Information

Variable	Value	Count
Group1	1	21
	2	53
	3	20
	4	11
	5	3
	Total	108

The last log-likelihood from the maximum likelihood iterations along with the statistic G is as follows: Log-likelihood = -94.242 G=94.961, DF=11, p-value =0.001. These statistics test the null hypothesis. As the p-value is less than 0.05 at least one of the coefficients associated with the predictors is different from zero.

The goodness of fit test displays Pearson and deviance goodness of fit test. The values are shown in Table 8.4. The higher p values indicate that there is insufficient evidence to claim that the model does not fit the data adequately. We will accept the null hypothesis of an adequate fit. We also consider the measure of association, results are displayed in Table 8.5. It displays the number and percentage of concordant, discordant and tied pairs as well as correlation statistics of Somers' D, Goodman-Kruskal Gamma and Kendall's Tau-a.

- The table of concordant, discordant and tied pairs is calculated by pairing the observations with different response values. A pair is concordant if a video having a low MOS score has a higher probability of having a low MOS score, discordant if the opposite is true, and tied if the probabilities are equal. From Table 8.5 we can see that 87.5% of pairs are concordant and 12.4% are discordant where as 0.1% are tied. These can be considered as comparative measure of prediction. This high concordance means that the model's 88% of predictions were correct.

- Somers' D, Goodman-Kruskal Gamma and Kendall's Tau-a are summaries of the table of concordant and discordant pairs. These measures lie between 0 and 1

Table 8.3: Ordinal Logistic Regression Table

						95% CI	
Predictor	Coef	SE Coef	Z	P	Odds Ratio	Lower	upper
Const(1)	-6.90533	1.07590	-6.42	0.000			
Const(2)	-2.79587	0.874108	-3.20	0.001			
Const(3)	-0.541823	0.764549	-0.71	0.479			
Const(4)	1.79280	0.882045	2.03	0.042			
PL							
2	3.28387	0.676342	4.86	0.000	26.68	7.09	100.43
3	3.62573	0.688742	5.26	0.000	37.55	9.74	144.84
QUA							
2	0.268846	0.543443	0.49	0.621	1.31	0.45	3.80
3	2.31386	0.661688	3.50	0.000	10.11	2.76	36.99
JIT							
2	0.950614	0.505848	1.88	0.060	2.59	0.96	6.97
3	0.825769	0.505342	1.63	0.102	2.28	0.85	6.15
PR							
2	1.68756	0.523659	3.22	0.001	5.41	1.94	15.09
3	0.426229	0.490785	0.87	0.385	1.53	0.59	4.01
BS							
2	-0.918479	0.504111	-1.82	0.068	0.40	0.15	1.07
3	-0.342031	0.495216	-0.69	0.490	0.71	0.27	1.87
motion	0.0362188	0.117017	0.31	0.757	1.04	0.82	1.30

where larger values indicate that the model has a better predictive ability.Value for Kendall's Tau-a is around 0.51 where as for the other two statistics its 0.75.

Now we fit the intercepts and the coefficient values to generate the model equations. In ordinal logistic regression, the event of interest is observing a particular score or less. For the rating of video quality, the following equations model the odds:

$$\begin{aligned}
\theta_1 &= -6.91 + 3.284 \times PL_2 + 3.626 \times PL_3 + 2.314 \times QUA_3 + 1.688 \times PR_2 \\
\theta_2 &= -2.796 + 3.284 \times PL_2 + 3.626 \times PL_3 + 2.314 \times QUA_3 + 1.688 \times PR_2 \\
\theta_3 &= -0.542 + 3.284 \times PL_2 + 3.626 \times PL_3 + 2.314 \times QUA_3 + 1.688 \times PR_2 \\
\theta_4 &= 1.792 + 3.284 \times PL_2 + 3.626 \times PL_3 + 2.314 \times QUA_3 + 1.688 \times PR_2
\end{aligned} \tag{8.1}$$

To understand these above mentioned equations we need to understand the basic assumption of Ordinal Logistic Regression (OLR). When we fit an OLR we assume that the relationship between the independent variables and the logits are the same for all the logits. That means that the results are a set of parallel lines or planes-one for

Table 8.4: Goodness of Fit Tests

Method	chi-square	DF	P
Pearson	319.779	417	1.000
Deviance	188.484	417	1.000

each category of the outcome variable. In our case one equation for each category of MOS. By using these above mentioned equations we were able to predict the category which was most likely the result should belong to for each condition. The MOS score, categorical grouping of mean as shown in Table A.6 column group1. Group1 column was created by converting any MOS value less than or equal to one to category 1. Any MOS value higher than 1 but lower than 2 were assigned to category 2. MOS value higher than 2 but less than 3 were assigned to category 3. MOS value higher than 3 but less than 4 were considered in category 4 where as anything above 4 was considered in category 5. Line plot shown in Figure 8.1 of group1 data, MOS and values predicted by Model OLR1 show that predictions were accurate.

Table 8.5: Measures of Association

Pairs	Number	Percent	Summary Measures	
Concordant	3449	87.5	Somers' D	0.75
Discordant	759	12.4	Goodman-Kruskal Gamma	0.75
Ties	134	0.1	Kendall's Tau-a	0.51
Total	3942	100.0		

Model OLR2

After analyzing results of Model OLR1 i.e. the complete model we found that the model was able to predict approximately 88% of results correctly. At the same time we found out that either a parameter with all its levels was statistically insignificant or one of the level of a certain parameter was statistically insignificant. There was a need to revisit our model. The purpose was to find a model with minimum number of parameters. For this very reason we removed parameter JIT, parameter BS and parameter motion. As these three were having higher p-values. Which indicates that these are statistically insignificant. The model developed after employing only parameter PL, parameter QUA and parameter PR showed decrease in concordance. Which indicates that there is a tradeoff between complexity and accuracy. The results of reduced model called Model OLR2 are shown in the Table 8.6.

Once again we discuss the log-likelihood from the maximum likelihood iterations

Table 8.6: Ordinal Logistic Regression Table for Model OLR2

						95% CI	
Predictor	Coef	SE Coef	Z	P	Odds Ratio	Lower	upper
Const(1)	-6.44657	0.881757	-7.31	0.000			
Const(2)	-2.35593	0.594080	-3.97	0.000			
Const(3)	-0.247848	0.512879	-0.48	0.629			
Const(4)	1.81532	0.699184	2.60	0.009			
PL							
2	3.18576	0.657306	4.85	0.000	24.19	6.67	87.71
3	3.45501	0.674401	5.12	0.000	31.66	8.44	118.72
QUA							
2	0.273674	0.530118	0.52	0.606	1.31	0.47	3.72
3	2.25888	0.656462	3.44	0.001	9.57	2.64	34.66
PR							
2	1.64270	0.508150	3.23	0.001	5.17	1.91	13.99
3	0.452476	0.479223	0.94	0.345	1.57	0.61	4.02

along with the statistic G. The last value was as follows: Log-likelihood = -98.231 G=86.983, DF=6, p-value =0.001. This statistics tests the null hypothesis. As the p-value is less than 0.05 at least one of the coefficients associated with the predictors is different from zero. The goodness of fit tests displays Pearson and deviance goodness

Table 8.7: Goodness of Fit Tests for Model OLR2

Method	chi-square	DF	P
Pearson	61.5550	62	0.492
Deviance	60.9115	62	0.515

of fit tests. The values are shown in Table 8.7. The higher p-values indicate that there is insufficient evidence to claim that the model does not fit the data adequately. We will accept the null hypothesis of an adequate fit.

Table 8.8: Measures of Association for Model OLR2

Pairs	Number	Percent	Summary Measures	
Concordant	3347	84.9	Somers' D	0.73
Discordant	465	11.8	Goodman-Kruskal Gamma	0.76
Ties	130	3.3	Kendall's Tau-a	0.50
Total	3942	100.0		

From the Table 8.8 the only obvious difference between the results of Model OLR1 and Model OLR2 is the lower concordance value for OLR2. A plot shown in Figure 8.3 of MOS, categorical data from Group1 and prediction by Model OLR2 show that generally the prediction were close to the MOS. A comparative plot of predictions from Model OLR1 and Model OLR2 identify the area where Model OLR2 was not able to predict as well as Model OLR1 Figure 8.4 highlight the difference.

8.2.2 Model Validation

We needed to validate our model's predictability against actual data. For model validation purposes we collected additional data from volunteers. We collected additional data for those combinations which were not proposed by the L-18 orthogonal array. Our models i.e. Model OLR1 and Model OLR2 both were developed by only considering the data points gathered by the L-18 orthogonal array. The column status of Table A.4 provides us with information about which combinations were selected for model development and which combinations were selected for model validation. Hence the values which were used for model development were not used for model validation. Using these two independent groups of data helped us in model validation. Better results for model also justify the use of the Taguchi method for identifying the most influential parameters by only studying a fractional factorial set of combinations. From Minitab 16 we calculated the expected probabilities for each additional combination. On the basis of the highest probability, a category was selected for each combination or configuration of parameter. The validation plot in Figure 8.4 shows the predicted categories by Model OLR1 and Model OLR2, MOS and group1 categories. We can see that the model predictions for both Models were close to the actual data. Model OLR1 performed better than Model OLR2 as it was able to consistently predict categories accurately. For the 108 cases when we predicted the category for QoE, only for 3 instances the predicted categories were off by two categories. For 33 cases out of 108, which is 31% the predictions, were only a single category off the MoS score. Whereas 67% of the predicted QoE scores were on target. The plot reflects the good predictability of the Model OLR1 for unknown data sets. These statistics provide us insight that the model was able to predict QoE with good accuracy even for unknown data set.

8.2.3 Model Deployment

The general idea for this model deployment was to incorporate it within the set-top-box to act like any other meter works. It should be able to generate QoE score which could be sent back to the service provider. This single score will not compromise user privacy and will only provide information to the service provider regarding predicted

QoE. Frequency of such score could be discussed with user when signing SLA. Such a deployment will provide our model access to the raw data stream which arrives at the user end. From this raw stream the level of quantization can be identified. For packet loss and packet order we need to add some information at the source end while we packetize the compressed video. This additional information could be about number of packets per Group of Picture (GoP). This will help identify packets lost or packets which came in different order. Differential delay of incoming data can be identified with reference to the system clock. Buffer size can be identified from the set-top-box configuration. In addition this model can be optimized by setting the correct buffer size according to the jitter and packet reorder. This will help improve the effectiveness of our model. This additional work of configurable set-top-box is part of our future work.

8.3 Summary

In this chapter we presented a model developed by using ordinal logistic regression. For model development we utilized the data we acquired through experiment discussed in Chapter 7. There were independent variables which were categorical as all the parameters studied were having three level i.e. low, high and medium. The dependent variable was also categorical because we recorded the feedback from volunteers using an ordered scale of worst, bad, fair, good and excellent. From user/volunteers perspective the scales are equidistant. In the light of these feature we opted to use Ordinal Logistic Regression for developing a model. Ordinal Logistic Regression can be implemented in cases where the dependent variable is categorical and ordered. Another important assumption for implementing OLR is that the categories are of equal interval. Also ordinal logistic regression lowered the constraints of generalized logistic regression by allowing categorical independent variable or dummy categorical variables. The result of ordinal logistic regression show that parameter PL and parameter QUA affected quality more than any other parameter. A model which comprises of all the influential parameters was able to predict video MOS with upto 88% accuracy. We also experimented with a model comprising of parameter PL, QUA and PR. This minimum model was able to predict with an accuracy of 84.5%. Though the difference between two models were less but plots indicate that complete model was able to predict the variations better than the reduced model.

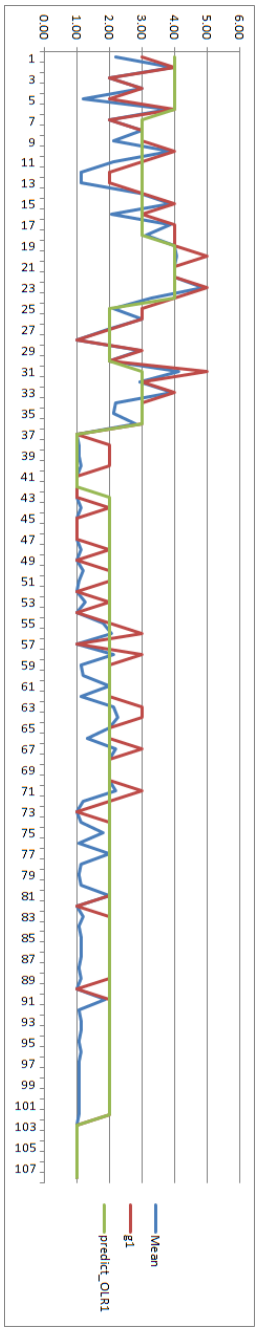

Figure 8.1: MOS, categorical data from Group1 and prediction by Model OLR1

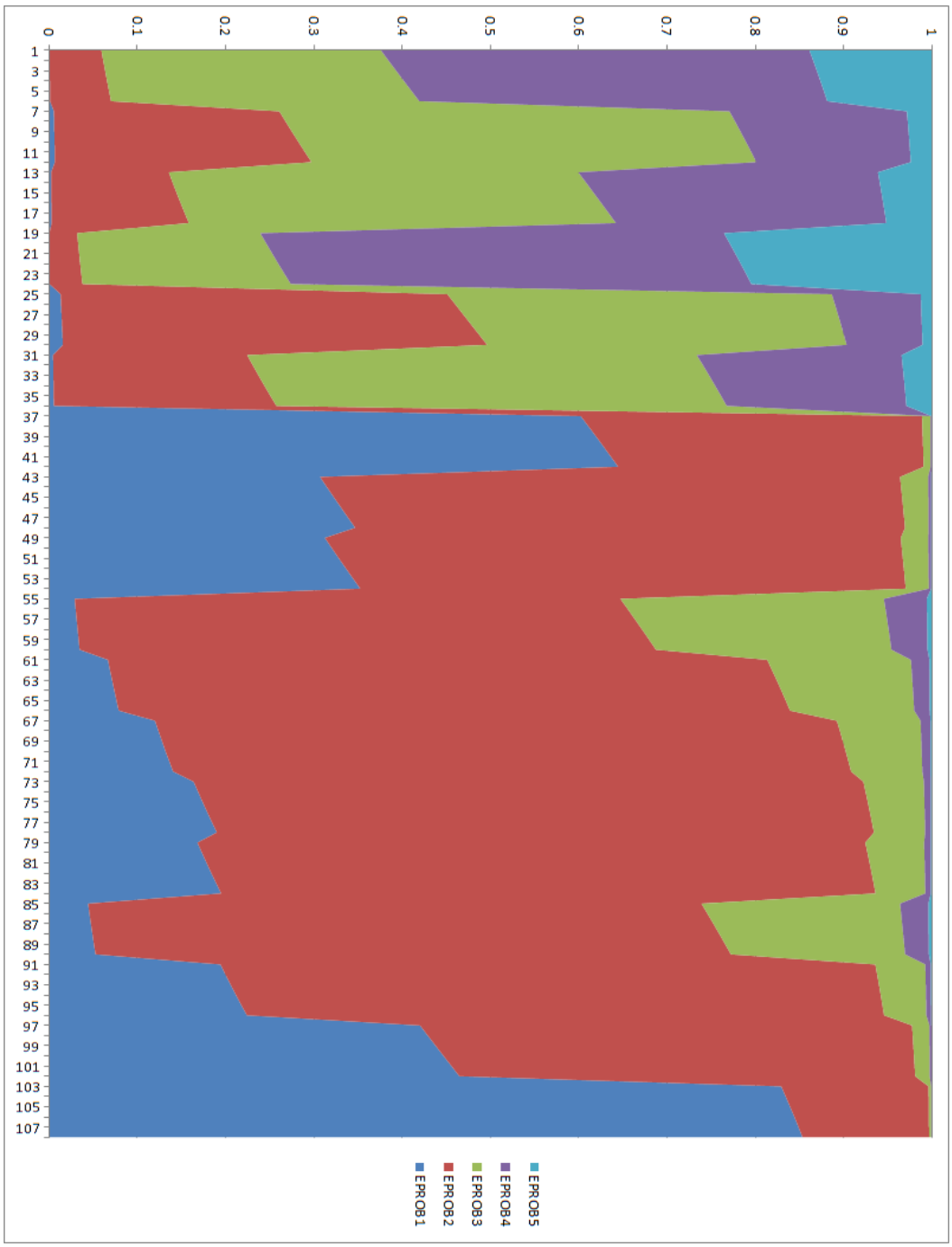

Figure 8.2: Cumulative probability plot for each video

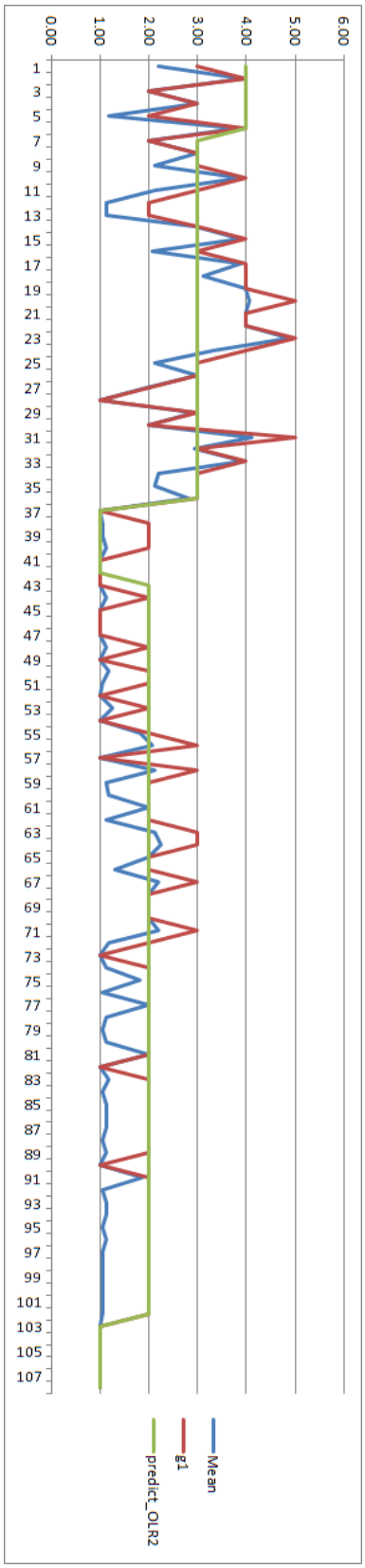

Figure 8.3: MOS, categorical data from Group1 and prediction by Model OLR2

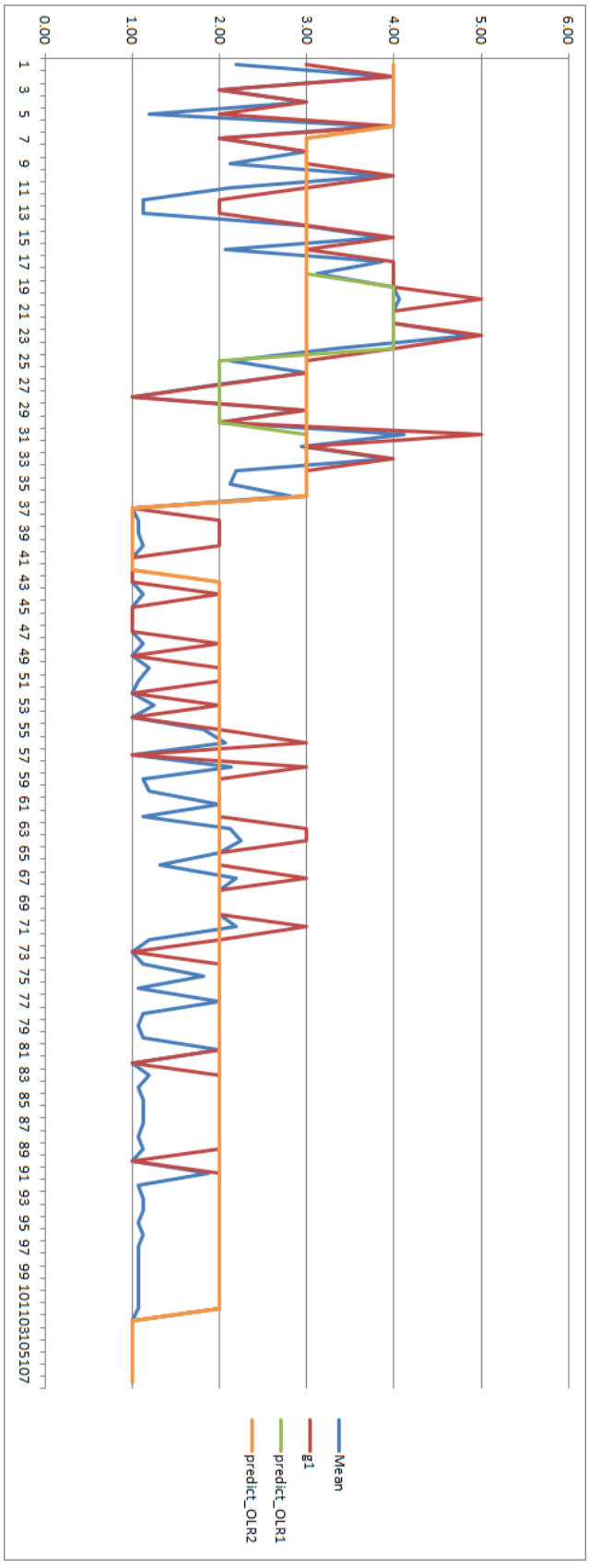

Figure 8.4: MOS, categorical data from Group1 and prediction by Model OLR1 and Model OLR2

Chapter 9

Conclusions & Future Work

9.1 Chapter Overview and Summary

The research described in this thesis set out to investigate the notion of end-end Quality of Experience (QoE) across three domains that link customers with providers of video or IPTV of video which is a fast emerging technology at the present time. The research was focussed on creating a robust mathematical model for QoE in this area. An extensive literature review was conducted into the issues and factors surrounding video QoE and the following research problems were identified from that study:

- Identifying most influential parameters from each of three critical domains which link users to the IPTV or video source and that affected user Quality of Experience
- Quality of Experience involves obtaining users' subjective judgements on quality perception and hence there is typically a requirement to conduct user surveys to obtain data that reflects these perceptions. Many of these experiments may be costly in time and funding, hence there is a significant need to reduce the experimentation time and yet maintain accuracy for the results
- Linking the most influential parameters from each of the identified domains into a comprehensive model that can be used to accurately predict subjective judgements (MOS) of video whilst based on objectively measurable quantities. Developing a prediction model, based on influential parameters from the three domains that could be used by service providers in order to determine the QoE of their customers using the video service and potentially improve their service offereings by appropriate adjustment of manageable parameters in the domains.

From literature that we reviewed, we were not able to identify any model which included a combined set of parameters from the content domain, network domain and customer premises processing domain and yet all three domains are traversed by video

services and parameter settings within those domains have a significant impact on the QoE experienced by users of the service. The existing models generally picked up only a limited number of parameters for use in modeling this QoE. We wanted to scientifically identify the model parameters from all three domains. For this we conducted targeted experiments for each of the individual three domains in order to extract appropriate parameters and identify the most influential parameters from among them. Moreover, the literature helped to identify the fact that many video quality experiments suffered from psychological effects such as boredom and memory. We utilized Taguchi robust design method for solving these two issues. Firstly, to reduce boredom and memory effects by reducing the experimentation time. Secondly, by reducing the number of combinations proposed by factorial designs. We also used correlation among raters as an indicator of non-randomness of the subjective feedback that we collected. Fleiss´ Kappa for testing the group concordance was also employed for the validation of data collected. These steps ensured that we collected reliable data rather than just random feedback.

In Chapter 1 we briefly introduced the concepts of QoE. Latest research directions and identification of areas which need further research. We also summarized the research contributions and the thesis outline. Chapter 2 is an overview of existing QoE research. The importance of parameter configuration and their effect on QoE. In addition we also looked into the current research done specifically in the area of video quality. A discussion of the important features of earlier work and their advantages and disadvantages was included. Chapter 3 provided a comprehensive overview of the systematic, theoretical analysis of the methods applied in this research. We also discussed the methods we preferred to use and comparisons with other parallel methods which we eliminated from consideration. In Chapter 4 we conducted experiments for identifying the most influential parameter from the content domain. We also proposed the optimum configuration for achieving better user QoE. Optimization was achieved by minimizing the loss function. Method proposed by Taguchi made use of a logarithmic function which was the ratio of mean performance to variation in mean performance due to known but uncontrolled variables. Higher values of SNR indicate the optimum levels of control factors. Hence, optimum performance can be delivered under the influence of noise factors. SNR represents sensitivity to variability and is a required measure for optimizing the robustness of a process or product. We also identified the configuration which should be avoided at all times in order to ensure acceptable quality service delivery. In Chapter 5 and Chapter 6 we considered the network and CPP domains individually also to capture the defining parameters. In Chapter 7 we took a global view and examined the most influential parameter across all three domains and ranked the remaining parameters in order of their influence on QoE. In Chapter 8 we developed an ordinal logistic model for predicting MoS for video QoE. The discussion

also included the analysis of the model and its performance or ability of prediction. Lastly, we conclude our work in this chapter.

9.2 Summary of Research Outcome

We have proposed a framework for answering the research questions. This framework includes the classification of influencing factors into three domains i.e. content, network and Customer Premises Processing (CPP). We conducted experiments for identifying influential parameters from each individual domain. These most influential parameters were later combined together to investigate their combined effect on end to end video QoE.

We have concluded that an appropriate QoE prediction model should be develop from the most influential parameters extracted from each of the three identified domains. We were also able to propose novel changes in the way these subjective experiments can be conducted for reducing psychological effects and validating the feedback-/subjective data we collected from volunteers. We were also able to make use of two independent methods for identifying and confirming the most influential parameters. These methods generated comparable results and this was very satisfying. We were not able to find a similar case study which found that Taguchi signal to noise ratio analysis and House of Quality methods were comparable methods. Taguchi method not only identified the most influential parameters from a domain it also identified the optimal configurations. These optimal configurations promise acceptable level of video QoE under all types of content variation.

We developed an ordinal logistic regression model for predicting video QoE. Ordinal logistic regression helps us in modeling data where independent and dependent variables are categorical. We proposed a reduced model based on statistically significant parameters. This reduce model can be used for predicting video QoE where there is a trade-off between complexity and accuracy. Model results were validated by collecting extra data points for the sake of model validation. The model performed very well with the validation data set and proved its worth in predicting video quality.

9.3 Conclusions

The basic aims of this thesis were to identify the parameters which affect quality in video services such as IPTV and to further understand the relationships between these parameters and their order of effect on overall video QoE. To find out the configuration(s) which could ensure acceptable quality and lastly to develop a predictive model which could predict MOS for video QoE based on parameters from the three domains

that influence this quality. Later we need to validate the results of the models we would develop.

At the start of this research we set the following aims and objectives for this research work.

- To develop a video database for improving subjective quality assessments against boredom effects.
- To validate subjective assessment feedback against randomness.
- To find justification for including parameters for model development
- To identify optimal configurations for successful and continuous service delivery.
- To identify configurations which should be avoided for successful and continuous service delivery.
- To find out the most influential parameters from the three domains.
- To identify parallel methods for identifying the most influential parameters (validation).
- To develop an end to end model for predicting video QoE.
- To validate the results generated by the prediction model.

The thesis aims were based on the current state-of-the-art research at the time of writing. We identified these grey areas by an in depth literature review and identified the need to research these issues and to find answers which can broaden our knowledge horizons. First aim was to conduct subjective video quality assessment experiments for generating the benchmark MOS scores for the research work. To conduct a successful subjective video quality assessment experiment there were two basic requirements. Firstly, we needed to reduce the psychological effects such as boredom, memory effect and forgiveness effect. Secondly, there was a need to evaluate the validity of the collected feedback. In order to reduce the psychological effect we looked at methods used in psychological studies. We used pairwise correlation and Fleiss´ Kappa methods for validating the feedback against randomness. As a random feedback means that the users were not rating the actual service and due to a certain psychological state they randomly provided the feedback. The main reason for random feedback is boredom. To eliminate this factor we developed video dataset from latest movies [136]. This database contained clips which were interesting and belonged to different genre which kept most of the volunteers interested in the experiment. To remove the memory effect and forgiveness effect we introduced small length video clips i.e of 10 seconds as proposed by

SAMVIQ. Results of pairwise correlation and Fleiss´ Kappa for all the experiments help us to conclude that volunteers were rating the videos as the concordance score was always satisfactory for all the experiments. Another source of introduction of boredom is the length of video and length of the overall experiment. In a case where there is a need to conduct an experiment for number of parameters the number of combinations become huge thus requiring longer time to finish the experiment. In some cases multiple session are required to complete the experiment. If the experiment become tedious and very long it introduces boredom for the volunteers. For this reason there was a need to reduce the experimentation time. We used Taguchi method for reducing the number of parameter combination that should be tested to generate significant results. Taguchi method not only reduced the time required to conduct the experiments it also provided us a method to analyze the data and rank the parameter on the basis of its influence on service quality. By adopting the above methods we were able to reduce the boredom effect to negligible levels. Moreover, the informal discussion at the end of each subjective experiment identified that all the volunteers were interested in the experiment due to the use of latest movie clips. We reviewed the ITU standards P-910 for conducting subjective experiments for capturing video quality. This and other ITU standards explain in detail the requirements for conducting subjective video quality assessment experiments but generally are quiet about reducing psychological effects. There is a need to incorporate the design of experimental techniques into these standards.

Another gap that we identified from literature review was relating to non availability of scientific justification regarding parameter inclusion in QoE models. We wanted to investigate the reasons for selecting or rejecting a certain parameter for model development. For this purpose we needed a certain method which could rank the parameters on the basis of its influence on QoE. Taguchi method provided us with a method to rank the parameters on the basis of their influence on QoE. In addition, this method helped us in reducing the time required to complete an experiment. We utilized the L9 orthogonal array for the first three experiments and L18 for the fourth experiment. These fractional factorial designs helped us in conducting a minimum number of experiments that were representative of a full factorial design. In Chapter 4 content domain parameters including the Bit Rate, the B-Frame, the Quantizer and the Partition Decision were tested. The aim of the experiment was to identify the most influential parameter from the content domain. Each of these parameters having 3 levels each, were included in the experiment. The videos were compressed using the combinations proposed by the L9 array of the Taguchi robust design method. The result and analysis help us to conclude that Quantization is the most important parameter from this domain. From literature review its evident that mostly bitrate is being used as a model parameter for predicting video QoE. Bitrate could be the easiest to be measured but it is certainly

not the most influential parameter from the content domain.

In Chapter 5 we studied the parameters packet loss, jitter, delay and packet reorder from the network domain. The experiment was conducted on a testbed which could emulate multiple network configurations and anomalies. The main objective of the experiment was to identify the most influential parameter from network domain. All of these parameters where having 3 levels each. The videos used for this experiment where compressed with HD and no content based anomalies were introduced. The noise factors considered in this experiment where the different classes of video content. These classes were the combination of level of motion, level of complexity and location. The network configurations were generated by the combinations proposed by L9 array of Taguchi robust design method. The feedback from the volunteers was studied and the SNR analysis helped us in identifying that packet loss, jitter and packet reorder were significantly influential in affecting QoE and that these three were statistically similar in effect. In the literature we reviewed we were not able to identify any model that made use of these three parameters collectively for developing network based QoE model. This relationship among these three parameter identified the importance of scientific investigation into parameter selection for model development.

In Chapter 6 we included three parameters i.e. parameter display size, parameter processing power and parameter buffer size in this experiment. The objective of this experiment was to investigate the best candidate parameters for model development from the CPP domain. It has has been realized that selection of hardware as well as software for processing at the customer end affects video quality. We concluded that buffer size is the most influential parameter from CPP domain. The other two parameters under investigation were the processor and the display parameters. They affected video quality equally and statistically there was not a statistically significant difference between the affect they had on perceptive video quality or QoE. We were not able to identify any model, from the work we reviewed, which incorporated CPP domain parameters with content or network domain parameters. Our experiment identified the importance of the parameter buffer size and the affect it introduce on video QoE. We conclude that buffer size should be included in model predicting video QoE.

Once we conducted experiment with all three individual domains we wanted to conduct an end to end experiment where the effect of all the most influential parameters from the three domains could be studied. In Chapter 7 we included all the most influential parameters from the three domain and tried to identify the most influential parameters from among the three domains. SNR analysis help us identify packet loss as the most influential parameter. Controlling the packet loss parameter over the network and keeping its value down should enable us in obtaining better MOS for video QoE. We also concluded that network parameters affect quality more than parameters

from content and CPP domains. The ranking achieved by SNR analysis confirms this statement as parameter packet loss is the most influential parameter and the other parameters from network domain also have some effect on QoE. The second most important parameter according to SNR analysis was Quantizer, it's values should be kept lower to ensure better MOS score for video QoE.

From our results and analysis we were able to conclude that by ranking parameters using a quality focus method helped us in building justification for inclusion and rejection of parameters. Also, understanding the relationship among these parameters is very important for successful continuous service delivery. During the process we were able to relate to the findings of earlier work done by other researchers and also discovered novel relationship among these parameters.

Another aim of the thesis was to find the optimum parameter configuration for QoE. Taguchi method enable us to identify the optimum parameter configuration by the use of quality loss function. This ensures that the design is robust and SNR can be used for optimization as well. For the SNR calculation, Taguchi considered average quality characteristics, standard deviation and a target quality value, where the standard deviation is caused by noise variables. We used the Taguchi SNR method for finding the optimum parameter configuration for each domain. It also help us avoid the setting which will reduce the QoE. We identified the optimum parameter configuration of content domain parameters as discussed in detail in Section 4.4.3. From network domain we studied packet loss, delay, jitter and packet re-order. On the basis of our results and analysis we concluded that packet loss, jitter and packet re-order were equally important and were having significant influence on QoE. We also found the optimum settings of network domain parameters for service delivery as discussed in Section 5.4.2. For CPP domain we studied display size, processing power and buffer size. Our results identified buffer size as the most influential parameter from CPP domain. We also identified the optimum parameter setting for the CPP domain parameters as discussed in Section 6.3.3. From the literature that we reviewed we were not able to find optimum parameter configurations for HD video service.

Next aim of the thesis was to validate our findings. A parallel method, which is used in the social sciences and quality analysis domain, and known as QFD was employed in parallel for finding out the most influential parameters. Separate questionnaires were used for capturing the user's perception. The results of the QFD method were compared with the results of the SNR analysis. QFD made use of HoQ tool/method and helped us in ranking the parameters on the basis of their importance from user perspective and the capability of the service provider. The results of HoQ method showed ranking similar to SNR method. For content domain it selected Quantization as the most important parameter. From network domain it identified that packet loss,

jitter and packet reorder are influential and similar in effect. For the CPP domain it identified buffer size as the most influential parameter. Taguchi SNR method and HoQ method both work with different inputs but generated the similar results helped us become confident on our findings and conclusions. When we attempted to use HoQ method for the end to end domain we found that the results were not significant at all and this method was not able to differentiate between any of the parameters included in the end to end experiment. The reason behind this failure is the fact that the HoQ approach made use of weights for importance calculation and once all the parameters are assigned equal weights the method was not able to differentiate between them. We were not able to identify any case study based comparison between Taguchi SNR method and HoQ method.

The last two aims that were listed for this research work were related to developing a predictive model and validation of results. In Chapter 8 we developed two models based on Ordinal Logistic Regression (OLR). OLR is the most suitable method for modeling MOS for video QoE in our case because we had categorical independent variables and a categorical dependent variable. The complete model based on all the five most influential parameter were able to successfully predict MOS for video QoE upto almost 88%. Though the value seems to be in the higher 80s the model's predictability is more than that. The model was able to predict the video quality with higher accuracy if we considered good and excellent as one group whereas bad and worst as another group. Which is logical, as mostly users are concerned if the service offered to them is good, acceptable or bad. In addition to the complete model we also looked at a reduced model which only incorporated statistically significant parameters. This model included the parameter packet loss, parameter quantization and parameter packet reorder. The predictability of this model was around 84% which is less than the complete model. The results acquired from the reduced model were plotted against the original values collected from users. Generally the predictions were quite close to the actual data but the model was not able to follow the changes in quality every time. It could provide us acceptable prediction results with less complexity as this model was developed using three parameters. The result of OLR show that parameter packet loss and parameter quantization affected quality more than any other parameter. Though the difference between two models were less but plots indicate that the complete model was able to predict the variations better than the reduced model.

9.3.1 Novel Contributions

This research has contributed to our existing domain knowledge in a number of new and novel ways:

1. We investigated a larger set of parameters and relevant for identifying their effect

on QoE and inter-parameter relationship

2. We were able to identify parameter quantization and parameter buffer size to be the most influential parameters of the content and CPP domains respectively. Whereas parameter packet loss, parameters packet reorder and parameter jitter were found to be equally important within the network domain.

3. We were able to identify the effectiveness of the Taguchi DoE approach in conducting our subjective experiments. Such utilization of Taguchi DoE significantly reduced the time and cost of experimentation without compromising experiment's evaluation capability.

4. We were able to utilize and (independent) HoQ method, in parallel to the Taguchi DoE method, for identifying order of effect of parameters and validating our Taguchi method findings.

5. We developed a model based on 5 parameters for predicting end to end video QoE, which has been demonstrated, gives good and accurate indications of this quality across the three key domains over which services are transmitted to users in the video IPTV environment.

9.4 Future Work

The thesis discussed the need for developing model based on parameters from the three domains of content, network and CPP. We were able to include 11 control parameters and 3 noise factors in 3 experiments. We identified the most influential parameters and then identified the order of effect of these most influential parameters on end to end MOS of video QoE.

We were able to identify the following weaknesses in our research

1. Basic selection of parameters were on the existing literature

2. Customer premises equipment experiments were restricted due to limitation of hardware combinations available for experimentation

3. Human psychology depends on multiple parameters and those were not studied in this work

4. Wireless communication is not an integral part of home network and we assumed all communication to be over wired medium

One of the identified weakness of this research was that instead of an exhaustive parameter research we made an effort to reduce the initial set of parameters to a

manageable number. For this reason we included the parameters which were repeatedly reported in the literature. A further study is required to investigate an even larger set of parameters. Especially for the CPP domain there is a need to investigate a larger set of parameters. This require development of special equipment which could be configured in combinations of the selected parameters. A special test bed needs to be developed to help enable the experiment execution.

Human psychology play an important role in evaluation of services being offered to them. For this study we reduced the parameters to the engineering domain and only discussed few psychological effects. A further study is required to incorporate important aspect of culture and impact of socio-economic condition. For this research we tried to develop interest, of volunteers/viewers, in the experimentation process. For this we used movie clips, which helped us in increasing the viewer's interest in the experiment. These clips were of 10 second length, which were very short. Further study is required to investigate the effects of using a longer video clip. In addition to the video length we also want to investigate the effect of varied genre of movies. Use of full length movie and option of selecting this full length movie from a pool of selected movies will simulate the actual scenario of an IPTV video on demand service deployment.

For the CPP domain we assumed the communication to be over the wired medium. There is a need to quantify the effect of wireless communication for end to end QoE. In addition we also want to quantify the end to end effect of parameters on interactive communication applications. Such an evaluation will require integration of further parameters which affect quality of an interactive application.

We also utilized HoQ method of QFD and we found that QFD failed to deliver the results, for analyzing the best parameter from the group of most influential parameters. There is need of further investigation of this fact and an improvement in HoQ is required to be proposed.

The outcome of our work can be utilized by content producers, content distributors, network provider or service providers and by the companies who are selling equipment for the CPP. We proposed the configurations which should be avoided in order to provide a minimal acceptable service as well as the configuration which will ensure at least good or excellent feedback from viewers. The content producers or distributors can use this knowledge in producing content which is better suited to constraints in the network and CPP domains. The knowledge of network domain parameters will help us in avoiding scenarios where the service provider will lose customer without knowing about it. If they can avoid the pitfalls they will be able to provide an acceptable service. The service provider will also be able to predict expected QoE and can fix the issue even before they become an issue. The model can be used by network planner for identifying minimum requirement within the system for a successful system deployment

which ensure good service quality.

Appendix A

Code Fragments and additional tables

A.1 Code Fragments

```
1
2  ####R-Programme
3  ####Program for significance test using permutation sampling####
4  y=c(3,0,9,0,1,1,1,3,1,1,1,3,3,1,1,9,3,3,3,9,3,3,9,+
5  9,9,3,3,3,3,3,3,0,3,3,3,0,3,3,9,0,1,3,1,0)
6  imp_rating=c(5,3,3,4,4,3,5,2,5,3,5)
7  R=11; C=4
8  hoq=matrix(y,nrow=R, ncol=C,byrow=TRUE)
9  hoq; imp_rating
10 tech_rating_mat=hoq*imp_rating
11 initial_tec_rating=colSums(tech_rating_mat)
12 initial_tec_rating
13 srt=sort(initial_tec_rating, decreasing = FALSE)
14 srt; initial_tec_rating=srt
15 d=NA
16 for(i in 1:C)
17   {
18   d[i]=initial_tec_rating[i]
19   }
20 replic=5000
21 permute=matrix(NA,nrow=replic,ncol=C)
22 diff=matrix(NA,nrow=replic,ncol=C)
23 for(i in 1:replic)
24   {
25   y=sample(y, replace=F)
26   #y=rpois(R*C,3)
27   hoq=matrix(y,nrow=R, ncol=C)
28   tech_rating_mat=hoq*imp_rating
```

```
  pseudo.t=colSums(tech_rating_mat)
  permute[i,]=pseudo.t
  }
p_value_mat=matrix(NA,nrow=C,ncol=C)
dif=NA;p_value=NA
diff=c(rep(NA,replic))
for(j in 1:C)
  {
  for (i in 1:C)
    {
    dif[i]=abs(d[j]-d[i])
    diff=abs(permute[,j]-permute[,i])
    if (i==j)
      {
      p_value[i]=NA
      }
    else
      {
      if(i<j)
        {
        p_value[i]=NA
        }
      else
        {
        p_value[i]=length(diff[(diff)>(dif[i])])/replic
        }
      }
  }
p_value_mat[j,]=p_value
}
p_value_mat
```

A.2 Tables

Table A.1: Calculation for Network Domain Fleiss´ Kappa

1	2	3	4	5	Pi
0	1	12	3	0	0.575
0	11	5	0	0	0.542
0	0	0	6	10	0.5
0	0	1	10	5	0.458

Continued on next page

Table A.1 – continued from previous page

1	2	3	4	5	Pi
0	0	0	9	7	0.475
0	0	4	8	4	0.333
0	0	0	6	10	0.5
9	5	2	0	0	0.392
0	0	0	7	9	0.475
0	0	0	6	10	0.5
0	0	0	5	11	0.542
0	0	6	10	0	0.5
0	1	9	6	0	0.425
0	0	0	8	8	0.467
0	0	0	4	12	0.6
0	0	10	6	0	0.5
0	0	0	8	8	0.467
0	0	0	11	5	0.542
0	0	0	9	7	0.475
0	0	1	8	7	0.408
0	0	1	3	12	0.575
0	4	11	1	0	0.508
0	0	1	12	3	0.575
0	0	2	8	6	0.367
0	0	0	9	7	0.475
0	0	0	8	8	0.467
0	1	13	2	0	0.658
0	0	2	10	4	0.433
0	0	0	5	11	0.542
0	0	0	8	8	0.467
0	0	0	11	5	0.542
0	0	1	4	11	0.508
0	0	2	10	4	0.433
0	0	1	11	4	0.508
0	0	0	5	11	0.542
0	3	11	2	0	0.492
0	0	0	9	7	0.475
0	0	0	13	3	0.675

Continued on next page

Table A.1 – continued from previous page

1	2	3	4	5	Pi
0	0	0	5	11	0.542
0	0	2	9	5	0.392
0	0	0	5	11	0.542
0	0	0	9	7	0.475
0	0	0	10	6	0.5
0	0	0	9	7	0.475
0	0	8	5	3	0.342
0	0	0	9	7	0.475
0	0	0	7	9	0.475
0	0	0	14	2	0.767
0	0	0	5	11	0.542
0	0	1	8	7	0.408
0	0	11	5	0	0.542
0	0	0	3	13	0.675
0	0	0	10	6	0.5
0	0	1	11	4	0.508
9	26	118	385	326	864
0.01	0.03	0.137	0.446	0.377	

Table A.2: Calculation for CPP Domain Fleiss´ Kappa

1	2	3	4	5	Pi
0	1	12	3	0	0.575
0	11	5	0	0	0.542
0	0	0	6	10	0.500
0	0	1	10	5	0.458
0	0	0	9	7	0.475
0	0	4	8	4	0.333
0	0	0	6	10	0.500
9	5	2	0	0	0.392
0	0	0	7	9	0.475
0	0	0	6	10	0.500
0	0	0	5	11	0.542
0	0	6	10	0	0.500

Continued on next page

Table A.2 – continued from previous page

1	2	3	4	5	Pi
0	1	9	6	0	0.425
0	0	0	8	8	0.467
0	0	0	4	12	0.600
0	0	10	6	0	0.500
0	0	0	8	8	0.467
0	0	0	11	5	0.542
0	0	0	9	7	0.475
0	0	1	8	7	0.408
0	0	1	3	12	0.575
0	4	11	1	0	0.508
0	0	1	12	3	0.575
0	0	2	8	6	0.367
0	0	0	9	7	0.475
0	0	0	8	8	0.467
0	1	13	2	0	0.658
0	0	2	10	4	0.433
0	0	0	5	11	0.542
0	0	0	8	8	0.467
0	0	0	11	5	0.542
0	0	1	4	11	0.508
0	0	2	10	4	0.433
0	0	1	11	4	0.508
0	0	0	5	11	0.542
0	3	11	2	0	0.492
0	0	0	9	7	0.475
0	0	0	13	3	0.675
0	0	0	5	11	0.542
0	0	2	9	5	0.392
0	0	0	5	11	0.542
0	0	0	9	7	0.475
0	0	0	10	6	0.500
0	0	0	9	7	0.475
0	0	8	5	3	0.342
0	0	0	9	7	0.475

Continued on next page

Table A.2 – continued from previous page

1	2	3	4	5	Pi
0	0	0	7	9	0.475
0	0	0	14	2	0.767
0	0	0	5	11	0.542
0	0	1	8	7	0.408
0	0	11	5	0	0.542
0	0	0	3	13	0.675
0	0	0	10	6	0.500
0	0	1	11	4	0.508
9	26	118	385	326	864
0.010	0.030	0.137	0.446	0.377	

Table A.3: Calculation for End to End Fleiss´ Kappa

1	2	3	4	5	Pi
0	13	3	0	0	0.68
0	0	2	13	1	0.66
0	16	0	0	0	1.00
0	0	16	0	0	1.00
13	3	0	0	0	0.68
0	0	3	13	0	0.68
0	16	0	0	0	1.00
0	1	14	1	0	0.76
0	14	2	0	0	0.77
0	0	2	14	0	0.77
0	14	2	0	0	0.77
14	2	0	0	0	0.77
14	2	0	0	0	0.77
0	2	13	1	0	0.66
0	0	2	14	0	0.77
0	15	1	0	0	0.88
0	0	2	14	0	0.77
0	0	14	2	0	0.77
0	0	0	16	0	1.00
0	0	2	11	3	0.49

Continued on next page

Table A.3 – continued from previous page

1	2	3	4	5	Pi
0	0	0	16	0	1.00
0	0	0	16	0	1.00
0	0	0	2	14	0.77
0	0	11	5	0	0.54
1	12	3	0	0	0.58
0	0	16	0	0	1.00
1	15	0	0	0	0.88
16	0	0	0	0	1.00
0	1	15	0	0	0.88
0	16	0	0	0	1.00
0	0	1	12	3	0.58
0	1	15	0	0	0.88
0	0	2	13	1	0.66
0	13	3	0	0	0.68
0	14	2	0	0	0.77
0	3	13	0	0	0.68
16	0	0	0	0	1.00
15	1	0	0	0	0.88
15	1	0	0	0	0.88
14	2	0	0	0	0.77
16	0	0	0	0	1.00
16	0	0	0	0	1.00
16	0	0	0	0	1.00
14	2	0	0	0	0.77
16	0	0	0	0	1.00
16	0	0	0	0	1.00
16	0	0	0	0	1.00
14	2	0	0	0	0.77
16	0	0	0	0	1.00
13	3	0	0	0	0.68
15	1	0	0	0	0.88
16	0	0	0	0	1.00
12	4	0	0	0	0.60
16	0	0	0	0	1.00

Continued on next page

Table A.3 – continued from previous page

1	2	3	4	5	Pi
3	13	0	0	0	0.68
1	13	2	0	0	0.66
16	0	0	0	0	1.00
0	14	2	0	0	0.77
14	2	0	0	0	0.77
13	3	0	0	0	0.68
0	16	0	0	0	1.00
14	2	0	0	0	0.77
0	14	2	0	0	0.77
0	12	4	0	0	0.60
0	16	0	0	0	1.00
13	1	2	0	0	0.66
0	13	3	0	0	0.68
0	16	0	0	0	1.00
0	16	0	0	0	1.00
0	16	0	0	0	1.00
0	13	3	0	0	0.68
13	3	0	0	0	0.68
16	0	0	0	0	1.00
14	2	0	0	0	0.77
3	13	0	0	0	0.68
15	1	0	0	0	0.88
0	16	0	0	0	1.00
14	2	0	0	0	0.77
15	1	0	0	0	0.88
14	2	0	0	0	0.77
0	16	0	0	0	1.00
16	0	0	0	0	1.00
13	3	0	0	0	0.68
15	1	0	0	0	0.88
14	2	0	0	0	0.77
14	2	0	0	0	0.77
14	2	0	0	0	0.77
15	1	0	0	0	0.88

Continued on next page

Table A.3 – continued from previous page

1	2	3	4	5	Pi
14	2	0	0	0	0.77
16	0	0	0	0	1.00
2	14	0	0	0	0.77
15	1	0	0	0	0.88
14	2	0	0	0	0.77
14	2	0	0	0	0.77
15	1	0	0	0	0.88
14	2	0	0	0	0.77
15	1	0	0	0	0.88
15	1	0	0	0	0.88
15	1	0	0	0	0.88
15	1	0	0	0	0.88
15	1	0	0	0	0.88
15	1	0	0	0	0.88
16	0	0	0	0	1.00
16	0	0	0	0	1.00
16	0	0	0	0	1.00
16	0	0	0	0	1.00
16	0	0	0	0	1.00
16	0	0	0	0	1.00
900	466	177	163	22	1728
0.52	0.27	0.10	0.09	0.01	

Table A.4: Factorial Design for 5 parameters with 3 levels

Exp. No	PL	QUA	JIT	PR	BS	Status
1	0.20	12	10	5	10	1
2	0.20	12	10	5	300	
3	0.20	12	10	5	500	
4	0.20	12	10	7	10	
5	0.20	12	10	7	300	
6	0.20	12	10	7	500	
7	0.20	12	10	15	10	A1
8	0.20	12	10	15	300	

Continued on next page

Table A.4 – continued from previous page

Exp. No	PL	QUA	JIT	PR	BS	Status
9	0.20	12	10	15	500	
10	0.20	12	15	5	10	
11	0.20	12	15	5	300	
12	0.20	12	15	5	500	
13	0.20	12	15	7	10	
14	0.20	12	15	7	300	2
15	0.20	12	15	7	500	
16	0.20	12	15	15	10	
17	0.20	12	15	15	300	
18	0.20	12	15	15	500	
19	0.20	12	40	5	10	A2
20	0.20	12	40	5	300	
21	0.20	12	40	5	500	
22	0.20	12	40	7	10	
23	0.20	12	40	7	300	
24	0.20	12	40	7	500	
25	0.20	12	40	15	10	
26	0.20	12	40	15	300	
27	0.20	12	40	15	500	3
28	0.20	26	10	5	10	
29	0.20	26	10	5	300	4
30	0.20	26	10	5	500	
31	0.20	26	10	7	10	
32	0.20	26	10	7	300	
33	0.20	26	10	7	500	A3
34	0.20	26	10	15	10	
35	0.20	26	10	15	300	
36	0.20	26	10	15	500	
37	0.20	26	15	5	10	
38	0.20	26	15	5	300	
39	0.20	26	15	5	500	
40	0.20	26	15	7	10	
41	0.20	26	15	7	300	
42	0.20	26	15	7	500	5

Continued on next page

Table A.4 – continued from previous page

Exp. No	PL	QUA	JIT	PR	BS	Status
43	0.20	26	15	15	10	
44	0.20	26	15	15	300	
45	0.20	26	15	15	500	
46	0.20	26	40	5	10	
47	0.20	26	40	5	300	A4
48	0.20	26	40	5	500	
49	0.20	26	40	7	10	
50	0.20	26	40	7	300	
51	0.20	26	40	7	500	
52	0.20	26	40	15	10	6
53	0.20	26	40	15	300	
54	0.20	26	40	15	500	
55	0.20	51	10	5	10	
56	0.20	51	10	5	300	
57	0.20	51	10	5	500	A5
58	0.20	51	10	7	10	
59	0.20	51	10	7	300	
60	0.20	51	10	7	500	
61	0.20	51	10	15	10	
62	0.20	51	10	15	300	
63	0.20	51	10	15	500	
64	0.20	51	15	5	10	
65	0.20	51	15	5	300	
66	0.20	51	15	5	500	
67	0.20	51	15	7	10	
68	0.20	51	15	7	300	
69	0.20	51	15	7	500	
70	0.20	51	15	15	10	
71	0.20	51	15	15	300	A6
72	0.20	51	15	15	500	
73	0.20	51	40	5	10	
74	0.20	51	40	5	300	
75	0.20	51	40	5	500	
76	0.20	51	40	7	10	

Continued on next page

Table A.4 – continued from previous page

Exp. No	PL	QUA	JIT	PR	BS	Status
77	0.20	51	40	7	300	
78	0.20	51	40	7	500	
79	0.20	51	40	15	10	
80	0.20	51	40	15	300	
81	0.20	51	40	15	500	
82	1.00	12	10	5	10	
83	1.00	12	10	5	300	A7
84	1.00	12	10	5	500	
85	1.00	12	10	7	10	
86	1.00	12	10	7	300	
87	1.00	12	10	7	500	
88	1.00	12	10	15	10	
89	1.00	12	10	15	300	
90	1.00	12	10	15	500	10
91	1.00	12	15	5	10	11
92	1.00	12	15	5	300	
93	1.00	12	15	5	500	
94	1.00	12	15	7	10	A8
95	1.00	12	15	7	300	
96	1.00	12	15	7	500	
97	1.00	12	15	15	10	
98	1.00	12	15	15	300	
99	1.00	12	15	15	500	
100	1.00	12	40	5	10	
101	1.00	12	40	5	300	A9
102	1.00	12	40	5	500	
103	1.00	12	40	7	10	
104	1.00	12	40	7	300	12
105	1.00	12	40	7	500	
106	1.00	12	40	15	10	
107	1.00	12	40	15	300	
108	1.00	12	40	15	500	
109	1.00	26	10	5	10	A10
110	1.00	26	10	5	300	

Continued on next page

Table A.4 – continued from previous page

Exp. No	PL	QUA	JIT	PR	BS	Status
111	1.00	26	10	5	500	
112	1.00	26	10	7	10	
113	1.00	26	10	7	300	
114	1.00	26	10	7	500	
115	1.00	26	10	15	10	
116	1.00	26	10	15	300	
117	1.00	26	10	15	500	A11
118	1.00	26	15	5	10	
119	1.00	26	15	5	300	
120	1.00	26	15	5	500	
121	1.00	26	15	7	10	
122	1.00	26	15	7	300	
123	1.00	26	15	7	500	
124	1.00	26	15	15	10	
125	1.00	26	15	15	300	
126	1.00	26	15	15	500	
127	1.00	26	40	5	10	A12
128	1.00	26	40	5	300	
129	1.00	26	40	5	500	
130	1.00	26	40	7	10	
131	1.00	26	40	7	300	
132	1.00	26	40	7	500	
133	1.00	26	40	15	10	
134	1.00	26	40	15	300	
135	1.00	26	40	15	500	
136	1.00	51	10	5	10	
137	1.00	51	10	5	300	
138	1.00	51	10	5	500	A13
139	1.00	51	10	7	10	7
140	1.00	51	10	7	300	
141	1.00	51	10	7	500	
142	1.00	51	10	15	10	
143	1.00	51	10	15	300	
144	1.00	51	10	15	500	

Continued on next page

Table A.4 – continued from previous page

Exp. No	PL	QUA	JIT	PR	BS	Status
145	1.00	51	15	5	10	
146	1.00	51	15	5	300	
147	1.00	51	15	5	500	
148	1.00	51	15	7	10	
149	1.00	51	15	7	300	
150	1.00	51	15	7	500	
151	1.00	51	15	15	10	
152	1.00	51	15	15	300	8
153	1.00	51	15	15	500	
154	1.00	51	40	5	10	
155	1.00	51	40	5	300	
156	1.00	51	40	5	500	9
157	1.00	51	40	7	10	
158	1.00	51	40	7	300	A14
159	1.00	51	40	7	500	
160	1.00	51	40	15	10	
161	1.00	51	40	15	300	
162	1.00	51	40	15	500	
163	1.80	12	10	5	10	
164	1.80	12	10	5	300	
165	1.80	12	10	5	500	
166	1.80	12	10	7	10	
167	1.80	12	10	7	300	
168	1.80	12	10	7	500	
169	1.80	12	10	15	10	
170	1.80	12	10	15	300	
171	1.80	12	10	15	500	
172	1.80	12	15	5	10	
173	1.80	12	15	5	300	
174	1.80	12	15	5	500	
175	1.80	12	15	7	10	
176	1.80	12	15	7	300	
177	1.80	12	15	7	500	
178	1.80	12	15	15	10	

Continued on next page

Table A.4 – continued from previous page

Exp. No	PL	QUA	JIT	PR	BS	Status
179	1.80	12	15	15	300	A15
180	1.80	12	15	15	500	
181	1.80	12	40	5	10	
182	1.80	12	40	5	300	
183	1.80	12	40	5	500	
184	1.80	12	40	7	10	
185	1.80	12	40	7	300	
186	1.80	12	40	7	500	
187	1.80	12	40	15	10	
188	1.80	12	40	15	300	
189	1.80	12	40	15	500	
190	1.80	26	10	5	10	
191	1.80	26	10	5	300	
192	1.80	26	10	5	500	
193	1.80	26	10	7	10	
194	1.80	26	10	7	300	
195	1.80	26	10	7	500	13
196	1.80	26	10	15	10	
197	1.80	26	10	15	300	
198	1.80	26	10	15	500	
199	1.80	26	15	5	10	
200	1.80	26	15	5	300	A16
201	1.80	26	15	5	500	
202	1.80	26	15	7	10	
203	1.80	26	15	7	300	
204	1.80	26	15	7	500	
205	1.80	26	15	15	10	14
206	1.80	26	15	15	300	
207	1.80	26	15	15	500	
208	1.80	26	40	5	10	
209	1.80	26	40	5	300	15
210	1.80	26	40	5	500	
211	1.80	26	40	7	10	
212	1.80	26	40	7	300	

Continued on next page

Table A.4 – continued from previous page

Exp. No	PL	QUA	JIT	PR	BS	Status
213	1.80	26	40	7	500	
214	1.80	26	40	15	10	
215	1.80	26	40	15	300	
216	1.80	26	40	15	500	
217	1.80	51	10	5	10	A17
218	1.80	51	10	5	300	
219	1.80	51	10	5	500	
220	1.80	51	10	7	10	
221	1.80	51	10	7	300	
222	1.80	51	10	7	500	
223	1.80	51	10	15	10	
224	1.80	51	10	15	300	16
225	1.80	51	10	15	500	
226	1.80	51	15	5	10	
227	1.80	51	15	5	300	
228	1.80	51	15	5	500	17
229	1.80	51	15	7	10	
230	1.80	51	15	7	300	
231	1.80	51	15	7	500	
232	1.80	51	15	15	10	
233	1.80	51	15	15	300	
234	1.80	51	15	15	500	A18
235	1.80	51	40	5	10	
236	1.80	51	40	5	300	
237	1.80	51	40	5	500	
238	1.80	51	40	7	10	18
239	1.80	51	40	7	300	
240	1.80	51	40	7	500	
241	1.80	51	40	15	10	
242	1.80	51	40	15	300	
243	1.80	51	40	15	500	

Table A.5: End to end experiment additional data for validation from 16 raters with mean

V1	**V2**	**V3**	**V4**	**V5**	**V6**	**V7**	**V8**	**V9**	**V10**	**V11**	**V12**	**V13**	**V14**	**V15**	**V16**	**Mean**
4	3	4	4	3	4	4	4	4	3	4	4	4	4	3	4	3.75
3	4	3	3	3	4	3	3	3	3	3	3	4	3	4	3	3.25
2	1	2	2	2	1	1	2	2	2	2	2	2	1	2	2	1.75
2	2	1	2	2	2	2	2	1	2	2	1	2	2	1	2	1.75
2	1	2	2	2	1	2	2	2	2	1	2	2	2	2	3	1.875
3	3	4	3	3	3	2	3	3	4	3	3	3	3	4	3	3.125
4	3	4	4	4	4	4	4	3	5	4	5	4	4	3	4	3.9375
5	5	5	4	5	4	5	5	5	5	5	5	4	5	4	5	4.75
4	3	4	4	4	4	4	4	3	4	4	4	3	4	4	4	3.8125
2	3	3	4	3	3	2	3	3	3	3	3	2	3	2	3	2.8125
3	2	2	2	2	2	2	2	2	3	2	1	2	2	2	3	2.125
3	1	2	2	2	1	2	2	2	1	2	2	2	1	2	2	1.8125
3	2	2	1	2	2	2	2	2	2	2	1	2	2	1	2	1.875
4	4	5	4	4	4	4	4	4	4	5	4	4	4	4	5	4.1875
4	4	3	4	4	4	4	3	4	4	4	4	3	4	4	5	3.875
2	3	3	3	3	2	3	3	3	2	3	3	3	2	3	3	2.75
2	3	3	2	2	3	3	3	3	3	3	3	2	3	2	3	2.6875
3	4	3	3	4	3	3	3	3	3	4	3	3	3	3	4	3.25
3	4	4	3	4	4	4	4	4	3	4	4	4	3	4	4	3.75

Continued on next page

Table A.5 – continued from previous page

V1	V2	V3	V4	V5	V6	V7	V8	V9	V10	V11	V12	V13	V14	V15	V16	Mean
4	3	4	4	4	3	4	4	4	4	4	3	4	3	4	3	3.6875
3	4	3	4	4	4	4	4	4	4	3	4	3	4	4	4	3.75
3	4	3	4	4	4	4	4	4	4	3	4	3	3	4	4	3.6875
3	4	4	4	4	4	4	4	4	4	4	5	4	4	4	4	4
3	4	3	3	3	4	3	3	3	3	3	4	3	3	4	3	3.25
2	1	1	1	1	1	1	1	1	2	1	1	1	1	1	2	1.1875
2	1	2	1	1	1	1	1	1	1	1	2	2	1	1	1	1.25
2	1	2	1	1	1	1	1	1	1	1	2	1	1	1	1	1.1875
1	2	2	1	1	1	1	1	1	1	1	2	1	1	1	2	1.25
1	1	1	1	1	1	1	1	1	1	1	2	1	1	1	1	1.0625
1	1	1	1	1	1	1	1	1	1	1	2	1	1	1	1	1.0625
2	1	2	2	2	1	2	2	2	2	2	2	1	2	2	2	1.8125
1	2	1	1	1	1	1	2	1	1	1	1	1	2	1	1	1.1875
2	3	2	2	2	2	2	2	2	3	2	3	2	2	2	2	2.1875
2	2	2	2	2	2	2	2	2	2	2	3	2	3	2	3	2.1875
2	1	2	2	2	2	2	1	2	2	2	2	2	2	1	2	1.8125
1	1	2	1	1	1	1	1	1	2	1	2	1	1	2	1	1.25
2	2	1	2	2	2	2	2	2	1	2	2	2	2	2	1	1.8125
2	2	1	2	2	2	1	2	2	2	2	2	2	1	2	2	1.8125
1	2	2	2	2	2	2	2	2	2	2	1	2	1	2	1	1.75
2	2	1	2	2	1	2	2	2	2	2	1	2	2	2	1	1.75

Continued on next page

Table A.5 – continued from previous page

V1	V2	V3	V4	V5	V6	V7	V8	V9	V10	V11	V12	V13	V14	V15	V16	Mean
1	1	1	1	1	2	1	1	2	1	1	1	1	1	1	2	1.1875
1	2	2	2	2	2	2	2	2	1	2	2	2	2	1	2	1.8125
2	3	2	2	2	2	2	3	2	2	2	3	2	2	2	3	2.25
2	2	2	2	3	2	2	2	2	3	2	2	2	2	2	3	2.1875
2	2	2	3	2	2	2	2	3	3	2	2	2	2	3	2	2.25
3	3	2	2	2	3	2	2	2	2	2	3	2	2	3	2	2.3125
2	1	2	1	1	1	1	1	1	2	1	1	1	1	2	1	1.25
1	1	1	1	1	2	1	2	1	1	1	2	1	1	2	1	1.25
1	1	2	1	2	1	1	1	1	1	1	2	1	2	1	1	1.25
2	1	2	2	2	1	2	2	2	2	2	1	2	2	1	2	1.75
2	1	2	1	2	2	2	2	1	2	2	2	1	2	2	2	1.75
2	1	2	2	2	1	2	2	2	1	2	2	1	2	1	2	1.6875
2	1	2	2	2	1	2	2	2	1	2	2	1	2	2	2	1.75
1	1	2	1	2	2	1	1	1	2	1	1	1	1	1	1	1.25
2	1	1	1	2	1	1	2	1	1	1	2	1	1	2	1	1.3125
1	1	1	1	1	1	1	2	1	1	1	1	1	1	2	1	1.125
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1

Continued on next page

Table A.5 – continued from previous page

V1	V2	V3	V4	V5	V6	V7	V8	V9	V10	V11	V12	V13	V14	V15	V16	Mean
1	2	2	2	1	2	2	2	2	1	2	2	2	2	1	2	1.75
1	1	1	1	1	1	2	1	1	1	1	1	1	1	1	1	1.0625
2	2	1	2	1	2	2	2	2	2	2	2	1	2	1	2	1.75
1	1	1	1	2	1	1	1	1	1	1	1	2	1	2	1	1.1875
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
1	2	2	2	2	1	2	2	2	2	1	2	2	2	1	2	1.75
1	2	2	2	1	2	2	2	2	2	1	2	2	2	1	2	1.75
1	2	1	1	1	2	1	1	1	1	2	1	1	1	1	2	1.25
2	2	2	2	1	1	2	2	2	2	2	2	1	2	2	1	1.75
1	1	1	1	1	1	2	1	1	1	1	2	1	1	1	2	1.1875
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1	1	1	1	1	1	2	1	1	1.0625
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
1	1	1	1	1	1	2	1	1	1	1	1	1	1	1	1	1.0625
1	1	1	1	1	1	1	1	1	1	1	1	1	2	1	1	1.0625
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1

Continued on next page

Table A.5 – continued from previous page

V1	V2	V3	V4	V5	V6	V7	V8	V9	V10	V11	V12	V13	V14	V15	V16	Mean
1	1	1	1	1	1	1	1	2	1	1	1	1	1	1	1	1.0625
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
2	2	2	2	2	3	2	2	2	2	2	2	3	2	2	2	2.125
1	1	1	1	1	2	1	1	1	1	1	1	1	1	1	1	1.0625
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1.0625
1	1	1	1	1	1	1	1	1	1	1	1	2	1	1	1	1.0625
1	1	1	1	1	1	1	1	1	2	1	2	1	1	1	1	1.125
2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1.0625
1	1	1	1	1	1	1	2	1	1	2	1	1	1	1	1	1.125
1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	1	1.0625
2	2	2	2	1	2	2	2	2	2	2	2	1	2	2	2	1.875
1	1	1	1	1	1	2	1	1	1	1	1	1	1	1	1	1.0625
2	1	1	1	1	1	1	1	1	2	1	1	1	1	1	1	1.125
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	1.0625
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1

Continued on next page

Table A.5 – continued from previous page

V1	V2	V3	V4	V5	V6	V7	V8	V9	V10	V11	V12	V13	V14	V15	V16	Mean
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1

Table A.6: MOS per video per condition

PL	QUA	JIT	PR	BS	Motion	Mean	Group1	Group2
1	1	1	1	1	1	2.19	3	1
1	1	1	1	1	2	3.94	4	2
1	1	1	1	1	3	2.00	2	1
1	1	1	1	1	4	3.00	3	2
1	1	1	1	1	5	1.19	2	1
1	1	1	1	1	6	3.81	4	2
1	1	2	2	2	1	2.00	2	1
1	1	2	2	2	2	3.00	3	2
1	1	2	2	2	3	2.13	3	1
1	1	2	2	2	4	3.88	4	2
1	1	2	2	2	5	2.13	3	1
1	1	2	2	2	6	1.13	2	1
1	1	3	3	3	1	1.13	2	1
1	1	3	3	3	2	2.94	3	2
1	1	3	3	3	3	3.88	4	2
1	1	3	3	3	4	2.06	3	1
1	1	3	3	3	5	3.88	4	2
1	1	3	3	3	6	3.13	4	1
1	2	1	1	2	1	4.00	4	2
1	2	1	1	2	2	4.06	5	2
1	2	1	1	2	3	4.00	4	2
1	2	1	1	2	4	4.00	4	2
1	2	1	1	2	5	4.88	5	2
1	2	1	1	2	6	3.31	4	2
1	2	2	2	3	1	2.13	3	1
1	2	2	2	3	2	3.00	3	2
1	2	2	2	3	3	1.94	2	1
1	2	2	2	3	4	1.00	1	1
1	2	2	2	3	5	2.94	3	2
1	2	2	2	3	6	2.00	2	1
1	2	3	3	1	1	4.13	5	2
1	2	3	3	1	2	2.94	3	2
1	2	3	3	1	3	3.94	4	2

Continued on next page

Table A.6 – continued from previous page

PL	QUA	JIT	PR	BS	Motion	Mean	Group1	Group2
1	2	3	3	1	4	2.19	3	1
1	2	3	3	1	5	2.13	3	1
1	2	3	3	1	6	2.81	3	2
2	3	1	2	1	1	1.00	1	1
2	3	1	2	1	2	1.06	2	1
2	3	1	2	1	3	1.06	2	1
2	3	1	2	1	4	1.13	2	1
2	3	1	2	1	5	1.00	1	1
2	3	1	2	1	6	1.00	1	1
2	3	2	3	2	1	1.00	1	1
2	3	2	3	2	2	1.13	2	1
2	3	2	3	2	3	1.00	1	1
2	3	2	3	2	4	1.00	1	1
2	3	2	3	2	5	1.00	1	1
2	3	2	3	2	6	1.13	2	1
2	3	3	1	3	1	1.00	1	1
2	3	3	1	3	2	1.19	2	1
2	3	3	1	3	3	1.06	2	1
2	3	3	1	3	4	1.00	1	1
2	3	3	1	3	5	1.25	2	1
2	3	3	1	3	6	1.00	1	1
2	1	1	3	3	1	1.81	2	1
2	1	1	3	3	2	2.06	3	1
2	1	1	3	3	3	1.00	1	1
2	1	1	3	3	4	2.13	3	1
2	1	1	3	3	5	1.13	2	1
2	1	1	3	3	6	1.19	2	1
2	1	2	1	1	1	2.00	2	1
2	1	2	1	1	2	1.13	2	1
2	1	2	1	1	3	2.13	3	1
2	1	2	1	1	4	2.25	3	1
2	1	2	1	1	5	2.00	2	1
2	1	2	1	1	6	1.31	2	1
2	1	3	2	2	1	2.19	3	1

Continued on next page

Table A.6 – continued from previous page

PL	QUA	JIT	PR	BS	Motion	Mean	Group1	Group2
2	1	3	2	2	2	2.00	2	1
2	1	3	2	2	3	2.00	2	1
2	1	3	2	2	4	2.00	2	1
2	1	3	2	2	5	2.19	3	1
2	1	3	2	2	6	1.19	2	1
3	2	1	2	3	1	1.00	1	1
3	2	1	2	3	2	1.13	2	1
3	2	1	2	3	3	1.81	2	1
3	2	1	2	3	4	1.06	2	1
3	2	1	2	3	5	2.00	2	1
3	2	1	2	3	6	1.13	2	1
3	2	2	3	1	1	1.06	2	1
3	2	2	3	1	2	1.13	2	1
3	2	2	3	1	3	2.00	2	1
3	2	2	3	1	4	1.00	1	1
3	2	2	3	1	5	1.19	2	1
3	2	2	3	1	6	1.06	2	1
3	2	3	1	2	1	1.13	2	1
3	2	3	1	2	2	1.13	2	1
3	2	3	1	2	3	1.13	2	1
3	2	3	1	2	4	1.06	2	1
3	2	3	1	2	5	1.13	2	1
3	2	3	1	2	6	1.00	1	1
3	3	1	3	2	1	1.88	2	1
3	3	1	3	2	2	1.06	2	1
3	3	1	3	2	3	1.13	2	1
3	3	1	3	2	4	1.13	2	1
3	3	1	3	2	5	1.06	2	1
3	3	1	3	2	6	1.13	2	1
3	3	2	1	3	1	1.06	2	1
3	3	2	1	3	2	1.06	2	1
3	3	2	1	3	3	1.06	2	1
3	3	2	1	3	4	1.06	2	1
3	3	2	1	3	5	1.06	2	1

Continued on next page

Table A.6 – continued from previous page

PL	QUA	JIT	PR	BS	Motion	Mean	Group1	Group2
3	3	2	1	3	6	1.06	2	1
3	3	3	2	1	1	1.00	1	1
3	3	3	2	1	2	1.00	1	1
3	3	3	2	1	3	1.00	1	1
3	3	3	2	1	4	1.00	1	1
3	3	3	2	1	5	1.00	1	1
3	3	3	2	1	6	1.00	1	1

Bibliography

[1] W. Stallings, *High-speed networks and internets: Performance and quality of service*. Pearson Education India, 2002.

[2] S. Wenger, "H. 264/avc over ip," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 13, no. 7, pp. 645–656, 2003.

[3] G. O'Driscoll, *Next generation IPTV services and technologies*. John Wiley & Sons, 2008.

[4] S. Winkler, *Digital video quality: vision models and metrics*. John Wiley & Sons, 2005.

[5] J. K. Choi, G. M. Lee, and H. J. Park, "Web-based personalized iptv services over ngn," in *Computer Communications and Networks, 2008. ICCCN'08. Proceedings of 17th International Conference on*, pp. 1–6, IEEE, 2008.

[6] W. Liang, J. Bi, R. Wu, Z. Li, and C. Li, "On characterizing ppstream: measurement and analysis of p2p iptv under large-scale broadcasting," in *Global Telecommunications Conference, 2009. GLOBECOM 2009. IEEE*, pp. 1–6, IEEE, 2009.

[7] S. Spoto, R. Gaeta, M. Grangetto, and M. Sereno, "Analysis of pplive through active and passive measurements," in *Parallel & Distributed Processing, 2009. IPDPS 2009. IEEE International Symposium on*, pp. 1–7, IEEE, 2009.

[8] X. Zhang, J. Liu, B. Li, and T.-S. P. Yum, "Coolstreaming/donet: a data-driven overlay network for peer-to-peer live media streaming," in *INFOCOM 2005. 24th Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings IEEE*, vol. 3, pp. 2102–2111, IEEE, 2005.

[9] H. J. Kim and S. G. Choi, "A study on a qos/qoe correlation model for qoe evaluation on iptv service," in *Advanced Communication Technology (ICACT), 2010 The 12th International Conference on*, vol. 2, pp. 1377–1382, IEEE, 2010.

[10] H.-J. Kim, D. H. Lee, J. M. Lee, K.-H. Lee, W. Lyu, and S.-G. Choi, "The qoe evaluation method through the qos-qoe correlation model," in *Networked*

Computing and Advanced Information Management, 2008. NCM'08. Fourth International Conference on, vol. 2, pp. 719–725, IEEE, 2008.

[11] K. Wolter, A. van Moorsel, and Q. B. QoE, "The relationship between quality of service and business metrics: Monitoring, notification and optimization," *Hewlett-Packard Labs Technical Report HPL-2001-96*, 2001.

[12] D. Collange and J. L. Costeux, "Passive estimation of quality of experience," *Journal of Universal Computer Science*, vol. 14, no. 5, p. 625641, 2008.

[13] A. Perkis, "Does quality impact the business model? case: Digital cinema," pp. 151–156, 2009.

[14] M. Venkataraman, S. Sengupta, M. Chatterjee, and R. Neogi, "Towards a video QoE definition in converged networks," p. 1616, 2007.

[15] J. Hassan, S. Das, M. Hassan, C. Bisdikian, and D. Soldani, "Improving quality of experience for network services [guest editorial," *IEEE Network*, vol. 24, pp. 4–6, Mar. 2010.

[16] P. Brooks and B. Hestnes, "User measures of quality of experience: why being objective and quantitative is important," *IEEE Network*, vol. 24, pp. 8–13, Mar. 2010.

[17] M. Fiedler, T. Hossfeld, and P. Tran-Gia, "A generic quantitative relationship between quality of experience and quality of service," *IEEE Network*, vol. 24, pp. 36–41, Mar. 2010.

[18] M. N. Zapater and G. Bressan, "A proposed approach for quality of experience assurance of IPTV," vol. 0, (Los Alamitos, CA, USA), p. 25, IEEE Computer Society, 2007.

[19] S. Khirman and P. Henriksen, "Relationship between quality-of-service and quality-of-experience for public internet service,"

[20] J. Goldberg and T. Kernen, "Network structures the internet, iptv and qoe," *EBU Technical Review*, pp. 1–11, Oct. 2007.

[21] M. Häsel, T. Quandt, G. Vossen, T. Hassner, M. Rehbein, P. A. Stokes, and L. Wolf, "Dagstuhl manifestos, vol. 2, issue 1 issn 2193-2433," 2013.

[22] K. Brunnström, S. A. Beker, K. De Moor, A. Dooms, S. Egger, M.-N. Garcia, T. Hossfeld, S. Jumisko-Pyykkö, C. Keimel, C. Larabi, *et al.*, "Qualinet white paper on definitions of quality of experience," 2013.

[23] B. Davie, D. Oran, S. Casner, and J. Wroclawski, "Integrated services in the presence of compressible flows," *Work in Progress (draft-davie-intservcompress-00. txt)*, 1999.

[24] K. McCloghrie, "Differentiated services quality of service policy information base," 2003.

[25] F. De Simone, M. Naccari, M. Tagliasacchi, F. Dufaux, S. Tubaro, and T. Ebrahimi, "Subjective assessment of h. 264/avc video sequences transmitted over a noisy channel," in *Quality of Multimedia Experience, 2009. QoMEx 2009. International Workshop on*, pp. 204–209, IEEE, 2009.

[26] T. K. Srull, "Memory, mood, and consumer judgment.," *Advances in consumer research*, vol. 14, no. 1, 1987.

[27] T. K. Srull, "The effects of subjective affective states on memory and judgment.," *Advances in consumer research*, vol. 11, no. 1, 1984.

[28] T. Hobfeld, R. Schatz, and S. Egger, "SOS: The MOS is not enough!," in *Quality of Multimedia Experience (QoMEX), 2011 Third International Workshop on*, pp. 131–136, IEEE, 2011.

[29] O. Werner, "Broadcasters requirements for iptv," *EBU Technical Review*, pp. 1–9, Apr. 2007.

[30] C. Mantel, P. Ladret, and T. Kunlin, "A temporal mosquito noise corrector," in *Quality of Multimedia Experience, 2009. QoMEx 2009. International Workshop on*, pp. 244–249, IEEE, 2009.

[31] A. Punchihewa and D. G. Bailey, "Artefacts in image and video systems; classification and mitigation," in *Proceedings of image and vision computing New Zealand*, p. 197202, 2002.

[32] L. Tionardi and F. Hartanto, "The use of cumulative inter-frame jitter for adapting video transmission rate," vol. 1, pp. 364–368 Vol.1, 2003.

[33] G. Liang and B. Liang, "Balancing interruption frequency and buffering penalties in VBR video streaming," pp. 1406–1414, 2007.

[34] L. Janowski and Z. Papir, "Modeling subjective tests of quality of experience with a generalized linear model," pp. 35–40, 2009.

[35] S. Wolf, M. H. Pinson, A. A. Webster, G. W. Cermak, and E. P. Tweedy, "Objective and subjective measures of mpeg video quality," *Society of Motion Picture and Television Engineers*, pp. 160–178, 1997.

[36] E. Cerqueira, L. Janowski, M. Leszczuk, Z. Papir, and P. Romaniak, "Video artifacts assessment for live mobile streaming applications," in *Future Multimedia Networking*, pp. 242–247, Springer, 2009.

[37] A. Maalouf, M.-C. Larabi, and C. Fernandez-Maloigne, "A grouplet-based reduced reference image quality assessment," in *Quality of Multimedia Experience, 2009. QoMEx 2009. International Workshop on*, pp. 59–63, IEEE, 2009.

[38] N. D. Narvekar and L. J. Karam, "A no-reference perceptual image sharpness metric based on a cumulative probability of blur detection," in *Quality of Multimedia Experience, 2009. QoMEx 2009. International Workshop on*, pp. 87–91, IEEE, 2009.

[39] X. Zhu and P. Milanfar, "A no-reference sharpness metric sensitive to blur and noise," in *Quality of Multimedia Experience, 2009. QoMEx 2009. International Workshop on*, pp. 64–69, IEEE, 2009.

[40] A. Ninassi, O. Le Meur, P. Le Callet, and D. Barba, "Considering temporal variations of spatial visual distortions in video quality assessment," *Selected Topics in Signal Processing, IEEE Journal of*, vol. 3, no. 2, pp. 253–265, 2009.

[41] F. N. Rahayu, U. Reiter, M. T. Nielsen, T. Ebrahimi, P. Svensson, and A. Perkis, "Analysis of ssim performance for digital cinema applications," in *Quality of Multimedia Experience, 2009. QoMEx 2009. International Workshop on*, pp. 23–28, IEEE, 2009.

[42] N. Staelens, S. Moens, W. Van den Broeck, I. Mariën, B. Vermeulen, P. Lambert, R. Van de Walle, and P. Demeester, "Assessing the perceptual influence of h. 264/svc signal-to-noise ratio and temporal scalability on full length movies," in *Quality of Multimedia Experience, 2009. QoMEx 2009. International Workshop on*, pp. 29–34, IEEE, 2009.

[43] U. Reiter and J. Korhonen, "Comparing apples and oranges: subjective quality assessment of streamed video with different types of distortion," in *Quality of Multimedia Experience, 2009. QoMEx 2009. International Workshop on*, pp. 127–132, IEEE, 2009.

[44] T. Liu, H. Yang, A. Stein, and Y. Wang, "Perceptual quality measurement of video frames affected by both packet losses and coding artifacts," in *Quality of Multimedia Experience, 2009. QoMEx 2009. International Workshop on*, pp. 210–215, IEEE, 2009.

[45] S. Winkler and P. Mohandas, "The evolution of video quality measurement: from psnr to hybrid metrics," *Broadcasting, IEEE Transactions on*, vol. 54, no. 3, pp. 660–668, 2008.

[46] O. Hohlfeld, "Stochastic packet loss model to evaluate qoe impairments," *PIK-Praxis der Informationsverarbeitung und Kommunikation*, vol. 32, no. 1, pp. 53–56, 2009.

[47] F. Babich, M. D'Orlando, and F. Vatta, "Video quality estimation in wireless IP networks: Algorithms and applications," *ACM TRANSACTIONS ON MULTIMEDIA COMPUTING COMMUNICATIONS AND APPLICATIONS*, vol. 4, no. 1, 2008.

[48] K. Gaitanis, S. Al Chikhani, and C. de Vleeschouwer, "Temporal optimization of quality during video compression," pp. 233–237, 2009.

[49] M. N. Garcia and A. Raake, "Frame-layer packet-based parametric video quality model for encrypted video in IPTV services," in *2011 Third International Workshop on Quality of Multimedia Experience (QoMEX)*, pp. 102–106, IEEE, Sept. 2011.

[50] L. Aspirot, P. Belzarena, G. Perera, B. Bazzano, and U. MONTEVIDEO, "End to end quality of service prediction based on functional regression," *HET-NETs*, vol. 5, 2005.

[51] K. Kamimura, H. Hoshino, and Y. Shishikui, "Constant delay queuing for jitter-sensitive IPTV distribution on home network," p. 16, 2008.

[52] B. T. M. Committee, "Maximizing the quality of sdtv in the flat-panel environment," *EBU technical review, European Broadcasting Union*, Apr. 2004.

[53] S. Tourancheau, P. Le Callet, and D. Barba, "Impact of the resolution on the difference of perceptual video quality between crt and lcd," in *Image Processing, 2007. ICIP 2007. IEEE International Conference on*, vol. 3, pp. III–441, IEEE, 2007.

[54] M. H. Pinson and S. Wolf, "The impact of monitor resolution and type on subjective video quality testing," tech. rep., NTIA Technical Memorandum TM-04-412, National Telecommunications and Information Administration, Mar. 2004.

[55] S. Péchard, S. Tourancheau, P. Le Callet, M. Carnec, D. Barba, *et al.*, "Towards video quality metrics for hdtv," in *Proceedings of the Second International Workshop on Video Processing and Quality Metrics for Consumer Electronics*, 2006.

[56] S. Péchard, M. Carnec, P. Le Callet, and D. Barba, "From SD to HD television: effects of H. 264 distortions versus display size on quality of experience," in *Image Processing, 2006 IEEE International Conference on*, pp. 409–412, IEEE, 2006.

[57] S. Deshpande, "A method for synchronization mismatch perception evaluation for large ultra high resolution tiled displays," pp. 238–243, 2009.

[58] M. Waltl, C. Timmerer, and H. Hellwagner, "A test-bed for quality of multimedia experience evaluation of sensory effects," pp. 145–150, 2009.

[59] T. H. Szymanski, "Bounds on end-to-end delay and jitter in input-buffered and internally-buffered IP networks,"

[60] I. Wechsung and A. Naumann, "Evaluating a multimodal remote control: The interplay between user experience and usability," pp. 19–22, 2009.

[61] S. Moller, K.-P. Engelbrecht, C. Kuhnel, I. Wechsung, and B. Weiss, "A taxonomy of quality of service and quality of experience of multimodal human-machine interaction," pp. 7–12, 2009.

[62] M. Deschamps, P. R. Band, and A. J. Coldman, "Assessment of adult cancer pain: shortcomings of current methods," *Pain*, vol. 32, no. 2, pp. 133–139, 1988.

[63] T. E. Gear, A. G. Lockett, and A. P. Muhlemann, "A unified approach to the acquisition of subjective data in r&d," *Engineering Management, IEEE Transactions on*, no. 1, pp. 11–19, 1982.

[64] L. Weinstein, X. Xie, and C. C. Cleanthous, "Purpose in life, boredom, and volunteerism in a group of retirees," *Psychological Reports*, vol. 76, no. 2, pp. 482–482, 1995.

[65] M. Ruano, J. Ribes, J. Ferrer, and G. Sin, "Application of the morris method for screening the influential parameters of fuzzy controllers applied to wastewater treatment plants.," *Water Science & Technology*, vol. 63, no. 10, 2011.

[66] D. Hamby, "A review of techniques for parameter sensitivity analysis of environmental models," *Environmental Monitoring and Assessment*, vol. 32, no. 2, pp. 135–154, 1994.

[67] P. SandipS, M. Patil Sudhir, P. Bhattu Ajay, and A. Sahasrabudhe, "Fem analysis and optimization of two chamber reactive muffler by using taguchi method,"

[68] E. B. Wilson, *An introduction to scientific research*. Courier Dover Publications, 1990.

[69] J. D. Barrow, *New theories of everything.* Oxford University Press, 2007.

[70] T. S. Kuhn, *The structure of scientific revolutions .* University of Chicago press, 1996.

[71] S. J. Hussain, G. A. Punchihewa, and R. Harris, "A survey on quality of experience for IPTV and taxonomy of influencing factors," in *International Conference on Intelligence and Information Technology*, (Lahore, Pakistan), pp. 308–315, Institute of Electrical and Electronic Engineers, Oct. 2010.

[72] R. Hamberg and H. de Ridder, "Time-varying image quality: Modeling the relation between instantaneous and overall quality," *SMPTE journal*, vol. 108, no. 11, p. 802811, 1999.

[73] 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*, pp. 321 –325, Jan. 2012.

[74] V. Seferidis, M. Ghanbari, and D. Pearson, "Forgiveness effect in subjective assessment of packet video," *Electronics Letters*, vol. 28, pp. 2013 –2014, Oct. 1992.

[75] R. Aldridge, J. Davidoff, M. Ghanbari, D. Hands, and D. Pearson, "Measurement of scene-dependent quality variations in digitally coded television pictures," *Vision, Image and Signal Processing, IEE Proceedings* -, vol. 142, pp. 149 –154, June 1995.

[76] I. Wakeman, F. Wilson, and W. Smith, "Quality of service parameters for commercial application of video telephony," in *Human Factors in Telecommunication Symposium*, 1993.

[77] R. A. Fisher, "The design of experiments.," 1935.

[78] D. K. Lin, "Making full use of taguchi's orthogonal arrays," *Quality and reliability engineering international*, vol. 10, no. 2, pp. 117–121, 1994.

[79] Y. Wu and A. Wu, *Taguchi methods for robust design.* ASME press New York, 2000.

[80] C. Shen, L. Wang, W. Cao, and L. Qian, "Investigation of the effect of molding variables on sink marks of plastic injection molded parts using taguchi doe technique," *Polymer-Plastics Technology and Engineering*, vol. 46, no. 3, pp. 219–225, 2007.

[81] J. Darwin, D. M. Lal, and G. Nagarajan, "Optimization of cryogenic treatment to maximize the wear resistance of 18% cr martensitic stainless steel by taguchi method," *Journal of materials processing technology*, vol. 195, no. 1, pp. 241–247, 2008.

[82] W.-C. Chen, Y.-Y. Hsu, L.-F. Hsieh, and P.-H. Tai, "A systematic optimization approach for assembly sequence planning using taguchi method, doe, and bpnn," *Expert Systems with Applications*, vol. 37, no. 1, pp. 716–726, 2010.

[83] S. Ickin, L. Janowski, K. Wac, and M. Fiedler, "Studying the challenges in assessing the perceived quality of mobile-phone based video," in *Quality of Multimedia Experience (QoMEX), 2012 Fourth International Workshop on*, pp. 164–169, IEEE, 2012.

[84] J. Sim and C. C. Wright, "The kappa statistic in reliability studies: use, interpretation, and sample size requirements," *Physical therapy*, vol. 85, no. 3, pp. 257–268, 2005.

[85] K. El Emam, "Benchmarking kappa: Interrater agreement in software process assessments," *Empirical Software Engineering*, vol. 4, no. 2, pp. 113–133, 1999.

[86] A. J. Viera, J. M. Garrett, *et al.*, "Understanding interobserver agreement: the kappa statistic," *Fam Med*, vol. 37, no. 5, pp. 360–363, 2005.

[87] ITU, "ITU-T P.910 Subjective video quality assessment methods for multimedia applications," tech. rep., International Telecommunication Union, 2006.

[88] EBU, "Samviq subjective assessment methodology for video quality," Tech. Rep. BPN 056, European Broadcasting Union, 2003.

[89] J. L. Fleisś, "Measuring nominal scale agreement among many raters.," *Psychological bulletin*, vol. 76, no. 5, p. 378, 1971.

[90] J. R. Landis and G. G. Koch, "The measurement of observer agreement for categorical data," *Biometrics*, vol. 33, pp. 159–174, Mar. 1977. ArticleType: research-article / Full publication date: Mar., 1977 / Copyright 1977 International Biometric Society.

[91] G. Holt, M. Buser, D. Harmel, K. Potter, M. G. Pelletier, and S. E. Duke, "Engineering and ginning," *Natural gas*, vol. 10, pp. 10–86, 2003.

[92] R. S. Pindyck and D. L. Rubinfeld, *Econometric models and economic forecasts*, vol. 2. McGraw-Hill New York, 1981.

[93] D. W. Hosmer Jr and S. Lemeshow, *Applied logistic regression*. John Wiley & Sons, 2004.

[94] T. M. Khoshgoftaar and E. B. Allen, "Logistic regression modeling of software quality," *International Journal of Reliability, Quality and Safety Engineering*, vol. 6, no. 04, pp. 303–317, 1999.

[95] S. Menard, *Applied logistic regression analysis*, vol. 106. Sage, 2002.

[96] M. Ries, C. Crespi, O. Nemethova, and M. Rupp, "Content based video quality estimation for h. 264/AVC video streaming," in *Wireless Communications and Networking Conference, 2007. WCNC 2007. IEEE*, p. 26682673, 2007.

[97] K. Yamagishi and T. Hayashi, "Parametric packet-layer model for monitoring video quality of IPTV services," vol. 19, 2008.

[98] J. Joskowicz, J. C. Lpez-Ardao, M. Gonzlez Ortega, and C. Garca, "A mathematical model for evaluating the perceptual quality of video," *Future Multimedia Networking*, p. 164175, 2009.

[99] J. Joskowicz and J. López Ardao, "A general parametric model for perceptual video quality estimation," in *Communications Quality and Reliability (CQR), 2010 IEEE International Workshop Technical Committee on*, pp. 1–6, IEEE, 2010.

[100] D. Loguinov and H. Radha, "Measurement study of low-bitrate internet video streaming," in *Proceedings of the 1st ACM SIGCOMM Workshop on Internet Measurement*, pp. 281–293, ACM, 2001.

[101] J. Mullin, L. Smallwood, A. Watson, and G. Wilson, "New techniques for assessing audio and video quality in real-time interactive communications," *IHM-HCI Tutorial*, 2001.

[102] A. Watson and M. A. Sasse, "Measuring perceived quality of speech and video in multimedia conferencing applications," in *Proceedings of the sixth ACM international conference on Multimedia*, pp. 55–60, ACM, 1998.

[103] J. Joskowicz and J. C. L. Ardao, "A parametric model for perceptual video quality estimation," *Telecommunication Systems*, vol. 49, no. 1, pp. 49–62, 2012.

[104] H. Koumaras, A. Kourtis, D. Martakos, and J. Lauterjung, "Quantified PQoS assessment based on fast estimation," in *of the Spatial and Temporal Activity Level?, Multimedia Tools and Applications*, p. 355374, Springer Editions, 2007.

[105] M. Garcia, A. Raake, and P. List, "Towards content-related features for parametric video quality prediction of IPTV services," in *IEEE International Conference on Acoustics, Speech and Signal Processing, 2008. ICASSP 2008*, pp. 757 –760, Apr. 2008.

[106] M. Garcia and A. Raake, "Impairment-factor-based audio-visual quality model for iptv," in *Quality of Multimedia Experience, 2009. QoMEx 2009. International Workshop on*, pp. 1–6, IEEE, 2009.

[107] S. J. Hussain, R. J. Harris, and G. A. Punchihewa, "Dominant factors in the content domain that influence the qoe of an iptv service," in *TENCON Spring Conference, 2013 IEEE*, pp. 572–577, IEEE, 2013.

[108] A. Khan, L. Sun, and E. Ifeachor, "Content clustering based video quality prediction model for mpeg4 video streaming over wireless networks," in *Communications, 2009. ICC'09. IEEE International Conference on*, pp. 1–5, IEEE, 2009.

[109] A. Raake, M.-N. Garcia, S. Moller, J. Berger, F. Kling, P. List, J. Johann, and C. Heidemann, "Tv-model: Parameter-based prediction of iptv quality," in *Acoustics, Speech and Signal Processing, 2008. ICASSP 2008. IEEE International Conference on*, pp. 1149–1152, IEEE, 2008.

[110] S. Argyropoulos, A. Raake, M.-N. Garcia, and P. List, "No-reference video quality assessment for sd and hd h. 264/avc sequences based on continuous estimates of packet loss visibility," in *Quality of Multimedia Experience (QoMEX), 2011 Third International Workshop on*, pp. 31–36, IEEE, 2011.

[111] Q. Dai and R. Lehnert, "Impact of packet loss on the perceived video quality," in *Evolving Internet (INTERNET), 2010 Second International Conference on*, pp. 206–209, IEEE, 2010.

[112] F. Boulos, B. Parrein, P. Le Callet, D. S. Hands, *et al.*, "Perceptual effects of packet loss on h. 264/avc encoded videos," *Perceptual Effects of Packet Loss on H. 264/AVC Encoded Videos*, 2009.

[113] N. Feamster and H. Balakrishnan, "Packet loss recovery for streaming video," in *12th International Packet Video Workshop*, pp. 9–16, PA: Pittsburgh, 2002.

[114] S. Kanumuri, P. C. Cosman, A. R. Reibman, and V. A. Vaishampayan, "Modeling packet-loss visibility in mpeg-2 video," *Multimedia, IEEE Transactions on*, vol. 8, no. 2, pp. 341–355, 2006.

[115] P. Calyam and C.-G. Lee, "Characterizing voice and video traffic behavior over the internet," in *International Symposium on Computer and Information Sciences (ISCIS)*, 2005.

[116] G. Sarwar, E. Lochin, and R. Boreli, "Mitigating the impact of packet reordering to maximize performance of multimedia applications," in *Communications (ICC), 2011 IEEE International Conference on*, pp. 1–5, IEEE, 2011.

[117] Y.-C. Chang, T. Carney, S. A. Klein, D. G. Messerschmitt, and A. Zakhor, "Effects of temporal jitter on video quality: assessment using psychophysical and computational modeling methods," in *Photonics West'98 Electronic Imaging*, pp. 173–179, International Society for Optics and Photonics, 1998.

[118] M. Claypool and J. Tanner, "The effects of jitter on the peceptual quality of video," in *Proceedings of the seventh ACM international conference on Multimedia (Part 2)*, pp. 115–118, ACM, 1999.

[119] S. Tao, J. Apostolopoulos, and R. Guérin, "Real-time monitoring of video quality in ip networks," *IEEE/ACM Transactions on Networking (TON)*, vol. 16, no. 5, pp. 1052–1065, 2008.

[120] D. Loguinov and H. Radha, "End-to-end internet video traffic dynamics: Statistical study and analysis," in *INFOCOM 2002. Twenty-First Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings. IEEE*, vol. 2, pp. 723–732, IEEE, 2002.

[121] M. Li, M. Claypool, and R. Kinicki, "Mediaplayer versus realplayer: a comparison of network turbulence," in *Proceedings of the 2nd ACM SIGCOMM Workshop on Internet measurment*, pp. 131–136, ACM, 2002.

[122] O. Issa, F. Speranza, T. H. Falk, *et al.*, "Quality-of-experience perception for video streaming services: Preliminary subjective and objective results," in *Signal & Information Processing Association Annual Summit and Conference (APSIPA ASC), 2012 Asia-Pacific*, pp. 1–9, IEEE, 2012.

[123] A. Jurgelionis, J.-P. Laulajainen, M. Hirvonen, and A. I. Wang, "An empirical study of netem network emulation functionalities," in *Computer Communications and Networks (ICCCN), 2011 Proceedings of 20th International Conference on*, pp. 1–6, IEEE, 2011.

[124] S. Hemminger *et al.*, "Network emulation with netem," in *Linux Conf Au*, pp. 18–23, Citeseer, 2005.

[125] S. R. Gulliver and G. Ghinea, "The perceptual and attentive impact of delay and jitter in multimedia delivery," *Broadcasting, IEEE Transactions on*, vol. 53, no. 2, pp. 449–458, 2007.

[126] Q. Huynh-Thu and M. Ghanbari, "Impact of jitter and jerkiness on perceived video quality," in *Second International Workshop on Video Processing and Quality Metrics for Consumer Electronics (VPQM-06), Scottsdale, Arizona*, 2006.

[127] J. Pouwelse, K. Langendoen, R. Lagendijk, and H. Sips, "Power-aware video decoding," in *22nd Picture Coding Symposium, Seoul, Korea*, pp. 303–306, 2001.

[128] W. B. Wilson, "Method for reducing processing power requirements of a video decoder," May 14 2002. US Patent 6,389,071.

[129] J. Pouwelse, K. Langendoen, and H. Sips, "Dynamic voltage scaling on a low-power microprocessor," in *Proceedings of the 7th annual international conference on Mobile computing and networking*, pp. 251–259, ACM, 2001.

[130] S. Mohapatra, R. Cornea, N. Dutt, A. Nicolau, and N. Venkatasubramanian, "Integrated power management for video streaming to mobile handheld devices," in *Proceedings of the eleventh ACM international conference on Multimedia*, pp. 582–591, ACM, 2003.

[131] F. Kozamernik, V. Steinmann, P. Sunna, and E. Wyckens, "Samviqa new ebu methodology for video quality evaluations in multimedia," *SMPTE Motion Imaging Journal*, vol. 114, no. 4, pp. 152–160, 2005.

[132] T. Kim and M. H. Ammar, "Receiver buffer requirement for video streaming over tcp," in *Electronic Imaging 2006*, pp. 607718–607718, International Society for Optics and Photonics, 2006.

[133] R. Rejaie, M. Handley, and D. Estrin, "Quality adaptation for congestion controlled video playback over the internet," in *ACM SIGCOMM Computer Communication Review*, vol. 29, pp. 189–200, ACM, 1999.

[134] Z. Iqbal, N. P. Grigg, R. Govinderaju, and N. M. Campbell-Allen, "Statistical comparison of final weight scores in quality function deployment (qfd) studies," *International Journal of Quality & Reliability Management*, vol. 31, no. 2, pp. 5–5, 2013.

[135] R. Brant, "Assessing proportionality in the proportional odds model for ordinal logistic regression," *Biometrics*, pp. 1171–1178, 1990.

[136] J. Guerin, "The psychology behind movie trailers@MISC," June 2014.