Addis Ababa University

Addis Ababa Institute of Technology

School of Electrical and Computer Engineering

Telecommunication Engineering Graduate Program

Quality of Experience Evaluation for Smartphone Video Streaming over 3G Network: In the case of Addis Ababa, Ethiopia

By

Zelalem Fitiwe G/Mariam

Advisor

Dr. Mesfin Kifle

A Thesis Submitted to School of Electrical and Computer Engineering, in Partial Fulfillment of the Requirements for the Degree of Masters of Science in Telecommunications Engineering

December 2019

Addis Ababa, Ethiopia

Addis Ababa University

Addis Ababa Institute of Technology

School of Electrical and Computer Engineering

This is to certify that the thesis prepared by Zelalem Fitiwe G/Mariam, entitled: Quality of Experience Evaluation of Smartphone Video Streaming for Addis Ababa 3G Customers and submitted in partial fulfillment of the requirements for the degree of Master of Science (Telecommunications Engineering Telecommunication Information Systems Track) complies with the regulation of the University and meets the accepted standards with respect to originality and quality.

Signed by the Examining Committee:

____________________	____________	__________
Chair or School Dean	Signature	Date
Mesfin Kifle (PhD)	____________	__________
	Signature	Date
Surafel Lema Abebe (PhD)	____________	__________
Internal Examiner	Signature	Date
Beneyam Berehanu Haile (PhD)	____________	__________
External Examiner	Signature	Date
____________________	____________	__________
Director of Postgraduate Program	Signature	Date

Declaration

I, the undersigned, declare that this thesis is my original work, has not been presented for a degree in this or any other university, and all sources of materials used for the thesis have been fully acknowledged.

Zelalem Fitiwe G/Mariam ________________

Name Signature

Place: Addis Ababa

Date of Submission: ________________

This thesis has been submitted for examination with my approval as a university advisor.

Mesfin Kifle (PhD) ________________

Advisor Signature

Abstract

Smartphones have become key enablers for users to exploit video streaming services such as YouTube, Hulu, Netflix. Quality of experience (QoE) represents the actual visual perception of users and it becomes a prominent concept. Considering the perception of users on the quality of mobile video streaming service evaluation is an important factor to assess the actual performance of the service that is provided to the users. So far, in Ethiopian Mobile Operator a combination of objective and subjective quality evaluation for video streaming service of mobile phone users, particularly for the most widely subscribed mobile 3G network, has not been performed.

In this thesis work, video streaming QoE evaluation over the 3G network in case of Addis Ababa, for smartphone users is made using network-side data from NetProb3010 tool, crowdsourcing tools (YoMoApp and Speedcheck Pro) from the application side and perception of customers from the user side.

The application and user side measurement results were collected from a field experiment campaign of 35 sample users after watching 8 different sample YouTube video sessions. In general, achieved video streaming metric results show that both subjective and objective metrics of QoE are not good and there is the unhappiness of smartphone customers with the current video streaming service quality. For example, based on NetProb3010 results the average initial delay is 19 seconds; the average number of stalling events is 0.38 times per minute and 6.46% for average stalled time rate. The application-level results show that the average initial delay, average number of stalling events, average stalled time duration and average video resolution watched results are 17.3 seconds, 0.47 times per 30 seconds, 3.79 seconds and 230P, respectively. Moreover, application-level average download throughput achieved during the experiment period is 1.78Mbps. The above objective metric results are reflected in the subjective metric (perception) result of the participants where the overall Mean Opinion Score (MOS) value is 2.79.

Keywords—Quality of Experience; Quality of Service; Video Streaming; Initial delay; Stalling events; Stalling duration; Video Resolution; Download Throughput; Mean Opinion Score; KQIs; 3G; Mobile Operators; Customers; Crowdsourcing; Customer perception; Optimization.

Acknowledgement

I would like to thank God for giving me the wisdom and skill to carry out my thesis work. My special thanks go to my advisor Mesfin Kifle (PhD) for his continuous support and advice during the research period. I am also very grateful to thank employees of ethio telecom (Departments of Service Management Center, Internal Audit, Residential Marketing, Quality and process and Customer Service Front Office) who have supported me in collecting data and providing necessary information for the successful delivery of this thesis work. Especially, my gratitude goes to Dawit Kefale and Abera Reasom for their vital support. I would like to thank the thesis progress report examiners Beneyam Berehanu (PhD) and Surafel Lema (PhD) for their valuable comments and suggestions. I am also glad to thank ethio telecom for giving me the opportunity to study my postgraduate program at Addis Ababa University, Addis Ababa Institute of Technology (AAiT). Last but not least, I am pleased to thank my wife and children for their priceless cares throughout my postgraduate program.

Abbreviations and Acronyms

3G	Third Generation
3GPP	Third Generation Partnership Project
AA	Addis Ababa
ACR	Absolute Category Rating
CDMA	Code Division Multiple Access
CN	Core Network
CRRM	Common Radio Resource Manager
CS	Circuit Switched
CSV	Comma Separated Values
DASH	Dynamic Adaptive Streaming over HTTP
DOM	Document Object Model
DPI	Dot Per Inch
ETSI	European Telecommunications Standards Institute
GGSN	Gateway GPRS Support Node
GMSC	Gateway MSC
GPRS	General Packet Radio Service
GPS	Geographical Position System
GSM	Global System for Mobile Communications
GSMA	Global System for Mobile Association
GW	Gate Way
HAS	HTTP Adaptive Video Streaming
HD	High-definition
HQ	Head Quarter
HSDPA	High-Speed Downlink Packet Access
HSPA	High-Speed Packet Access

HSPA+	Evolved HSPA
HSUPA	High-Speed Uplink Packet Access
HTML5	Hypertext Markup Language
HTTP	Hypertext Transfer Protocol
ICT	Information Communication Technology
ITU	International Telecommunication Union
ITU-T	ITU Telecommunication Standardization Sector
Kbps	Kilobits Per Second
KPI	Key Performance Indicators
KQI	Key Quality Indicators
Mbps	Mega Bits per second
MGW	Media GateWay
MNOs	Mobile Network Operators
MOS	Mean Opinion Score
MPD	Media Presentation Description
Ms	Millisecond
MSC	Mobile Switching Center
NMS	Network Management System
NNOC	National Network Operation Center
O & M	Operation and Maintenance
OTT	Over the Top
P	Pixel
PCs	Personal Computers
PQI	Performance Quality Indicator
PS	Packet Switched
PSTN	Public Switched Telephone Network

QoE	Quality of Experience
QoS	Quality of Service
RNAS	RAN Access Server
RAN	Radio Access Network
RAT	Radio Access Technology
RTP	Real-Time Protocol
RTSP	Real Time Streaming Protocol
RNC	Radio Network Controller
SGSN	Serving GPRS Support Node
SLA	Service Level Agreement
SPSS	Statistical Package for the Social Sciences
SMC	Service Management Center
TCP	Transmission Control Protocol
TPO	Transformation Program Office
UDP	User Datagram Protocol
UE	User Equipment
UMTS	Universal Mobile Telecommunications System
URL	Universal Resource Locator
UTRAN	UMTS Terrestrial RAN
VBR	Variable Bit Rate
VoD	Video-on-Demand
Wi-Fi	Wireless Fidelity
YoMoApp	YouTube Monitoring Application

Contents

Declaration i
Abstract ii
Acknowledgement iii
Abbreviations and Acronyms iv
List of Tables x
List of Figures xi
Chapter 1 1
Introduction 1
1.1. Background 1
1.2. Statement of the Problem 3
1.3. The objective of the Thesis 3
1.3.1. General Objective 3
1.3.2. Specific Objectives 3
1.4. Research Questions 4
1.5. Methodology for Experimental Evaluation 4
1.6. Related Works 6
1.7. Scope of the Thesis 8
1.8. Limitation of the Thesis 8
1.9. Contribution 8
1.10. Thesis Organization 9
Chapter 2 10
Background on Quality Evaluation over UMTS 10
2.1. UMTS Network 10
2.1.1. Architecture of UMTS 11
2.1.2. Quality of Service in UMTS Network. 12
2.1.3. UMTS QoS Basic Classes 12
2.2. Quality of Experience 13
2.3. Video Streaming Technologies 14
2.3.1. HTTP Adaptive Video Streaming (HAS) 14
2.3.2. Mobile YouTube Video Streaming 16
2.4. Factors that Affect the QoE of Users in Video Streaming 17
2.5. QoS Metrics in Video Streaming Service 18

2.6. QoE Metrics in Video Streaming Service ... 18
2.7. Video Streaming QoS and QoE Performance Evaluation Methods ... 20
2.7.1. Subjective QoE Evaluation Techniques ... 20
2.7.2. Objective QoE Evaluation Techniques ... 21
Chapter 3 ... 23
Video Streaming Quality Management Practices in Ethio telecom ... 23
3.1. Introduction ... 23
3.2. Video Streaming QoS management in Ethio telecom ... 23
3.3. Video Streaming QoE Management in Ethio Telecom ... 24
3.4. ITU Service Quality Performance Management Process ... 25
3.5. Proposed Solution for Video Quality Management in Ethio Telecom ... 26
Chapter 4 ... 28
Planning and Methodology for System Elements ... 28
4.1. Introduction ... 28
4.2. Plan for Taking YouTube Video Streaming Measurements ... 28
4.3. Methodology for System Elements ... 29
Chapter 5 ... 36
Data Collection Mechanisms ... 36
5.1. Introduction ... 36
5.2. Network Level Evaluation Using NetProb3010 ... 36
5.3. Applications and User Level Experimental Evaluation ... 37
5.3.1. Application Level Evaluation Using YoMoApp ... 37
5.3.2. Application Level Evaluation Using Speedcheck Pro ... 38
5.3.3. User Level Subjective Evaluation Using Questionnaires ... 40
Chapter 6 ... 43
Analysis, Results and Interpretations ... 43
6.1. Introduction ... 43
6.2. Network Level Evaluation of NetProb3010 ... 43
6.3. Application and User Level Experimental Measurement Results ... 46
6.3.1. YouTube Monitoring Application (YoMoApp) Results ... 47
6.3.2. Application Level - Speedcheck Pro ... 50
6.3.3. User Level Subjective Metrics ... 54
6.3.3.1. Quality Category Metrics ... 55

6.3.3.2. Spatiotemporal Metrics 58
6.3.3.3. Other Factors 61
6.4. Correlation Analysis for Experimental Evaluation Results 63
6.5. Discussion 64
Chapter 7 67
Conclusion and Future Work 67
7.1. Conclusion 67
7.2. Future Work 69
References 70
Appendix A - Experiment Instruction 75
Appendix B – YoMoApp & Speedcheck Pro Tools Sample Interface & Results 76
Appendix C - Correlation results of subjective and objective metrics 77
Appendix D - Subjective Metrics Questionnaire 78

List of Tables

Table 2: 1 UMTS Quality of Service Classes 13

Table 4: 1 Summary of Subjects' Profile.. 30

Table 4: 2 Description of Sample Videos for the Experiment... 32

Table 4: 3 Experiment Parts of the Day with Respective No. of Subjects Percentage 33

Table 4: 4 Processes of Evaluation (Rating) of YouTube Video Streaming in the experiments. 34

Table 5: 1 Speedcheck Pro Field Data.. 39

Table 5: 2 Rating Scale for Quality and Annoyance Level 40

Table 5: 3 Contents of Pre-Watching Hard-Copy Questionnaire 41

Table 5: 4 Contents of Post-Watching on YoMoApp Application Questionnaire 42

Table 5: 5 Contents of Post-Watching Hard Copy Questionnaire... 42

Table 6: 1 NetProb3010 - Network Level Measurement Results of Video Streaming Metrics ... 44

Table 6: 2 NetPro3010 Video Streaming KQIs Evaluation Result Summary.............................. 46

Table 6: 3 YoMoApp Evaluation Results of Video Streaming Metrics 47

Table 6: 4 Result Summary for All Subjects' Average Download Throughput during the Specific Period of Video Watching 51

Table 6: 5 Average Content, Video and Streaming Quality and overall MOS Result 55

Table 6: 6 Summary of Average Acceptability Rating for Videos Quality.................................. 56

Table 6: 7 Subjects Perception Rating Result on Annoyance Level of Video Streaming Metrics57

List of Figures

Figure 2: 1 UMTS Architecture [22] 11

Figure 2: 2 Diagram of Http Adaptive Video Streaming [25] 16

Figure 2: 3 Factors that Affect QoE [30] 18

Figure 3: 1 ITUs' QoE Management Process for Telecom Service Quality Assurance [39] 25

Figure 3: 2 Network Architecture Process for video quality management.................................. 27

Figure 4: 1 Representative Frames Screen Shot of the Sample Videos in the zel fit Channel 31

Figure 6: 1 Evaluation Results of Initial Delay Metrics in Seconds... 44

Figure 6: 2 Evaluation Results of Number of Stalling Events per Minute 45

Figure 6: 3 Evaluation Results of Stalled Time Rate Percentage 45

Figure 6: 4 YoMoApp result of Average Video Resolution Watched by Subjects 48

Figure 6: 5 Evaluation Results of average Initial Delay Metric per Experiment Locations........ 48

Figure 6: 6 Evaluation Results of Average Stalling Events and Quality Switches Metrics per Experiment Locations 49

Figure 6: 7 Result of Average Resolution Watched per Location 49

Figure 6: 8 Result of Average Resolution Watched per Environment Context........................... 50

Figure 6: 9 Geographical Locations of the Experiments 51

Figure 6: 10 Statistics of Average Download Throughput per Location.................................... 52

Figure 6: 11 Subjects Smartphone's Average Download Throughput in Mbits/s during the Evaluation Sessions of Video Streaming Experiment 52

Figure 6: 12 Average Download Throughput per the Experiment Environment Context........... 53

Figure 6: 13 Average Download Throughput per Parts of the Experiment Day 54

Figure 6: 14 Average MOS per frequency of subjects 55

Figure 6: 15 Result of Average Acceptability per Experiment Location 57

Figure 6: 16 Impact of Average Video Resolution Watched on QoE 58

Figure 6: 17 Result of Overall MOS per Experiment Location.. 59

Figure 6: 18 MOS per Environment of Subjects 59

Figure 6: 19 Result of Overall MOS per Experiment Parts of the Day 60

Figure 6: 20 Impact of Average Initial Delay on QoE... 60

Figure 6: 21 Average MOS Value per Number of Stalling Events 61

Figure 6: 22 MOS per Subjects' Gender.. 62

Figure 6: 23 MOS Value per Subjects' Age 62

Chapter 1

Introduction

1.1. Background

Customer QoE is dependent on a multitude of factors. It is tightly attached to customer satisfaction and is related to the success or failure of the telecommunication service provider in providing a service. Mobile service is one of the services provided by telecom operators. This mobile service brings new opportunities to develop new services on mobile devices, and traditional services are evolved into smartphone mobile applications. Meanwhile, this trend introduces new technological challenges. The first major challenge involves accommodating growth in data traffic for mobile services in the face of limited available wireless bandwidth throughput. The other challenge involves accommodating the expectations for consistent and high-quality services. One of the services provided by Mobile Network Operators (MNOs) is a video streaming service that customers access online the streaming services on demand through their data connection.

Video streaming services such as Video On Demand (VoD), have become more popular than ever before. There is a constantly growing number of users who are watching video streaming from mobile devices and Cisco predicts that by 2022 Mobile will represent 20 % of total IP traffic, Smartphones will surpass 90 percent of mobile data traffic and Consumer VoD traffic will nearly double [1], [2]. At the same time, video streaming services are constantly enriching their libraries with new and higher bitrate content. The increasing popularity of video streaming among users has led to demands for higher bandwidth availability while maintaining a satisfactory QoE.

To meet these requirements, operators are forced to constantly upgrade their mobile access infrastructure in order to entertain the high traffic volumes that are generated from mobile video streaming services. At the same time, continuous monitoring and evaluating of the different network segments for quality-related performance issues is necessary to ensure that faults and problematic links are identified as soon as possible. In this way, any quality problems that affect the users can be properly addressed before they start to impact subscribers. QoE is a customer-centric measurement and every QoE system measures service quality from the customer's perception. In the other way, traditional QoS is a network-centric, which employs network-level measurement methodologies for service assessment.

Major objective metrics in video streaming services which related to network performance are initial delay, number of stalling, stalling duration, resolution watched and so on. The problem of the objective metric is that network performance cannot directly represent users' perception of services. Many other factors, such as psychological investigation, end-to-end service process, and objective metrics mapping, should be considered as a part of the quality assessment. In recent years, more and more international organizations realize the importance of considering QoE for network evaluation, control and management.

As ethio telecom is evolving its networks to meet the ever-increasing data demands of the market it is important to consider the user perception of how the system is performing efficiently. Ethio telecom is still using the objective measurement of KQIs for evaluation of video streaming service quality in 3G network with little regard for QoE of video streaming from the user's perspective.

Based on our finding from seminar work in title opportunity and challenges of OTT in ethio telecom [3], video streaming usage contributes to a major share of the internet or data traffic. From the data that we analyzed in the seminar work, for the periods of three months (August to October 2018), the total 3G data traffic of the company shows that 24 % of the total traffic share go to OTT video streaming services. From this streaming traffic data, most of the traffic can be shared with one of the biggest OTT providers of video streaming services, YouTube service which accounts 42% of the total streaming services and most of them were smartphone users. Even though the number of customers usage of this OTT service is increasing from time to time, and simultaneously data revenue from the services are increasing, ethio telecom's method of quality evaluation for video streaming service delivered don't consider end-users perception. And our observation shows that 3G customers are not satisfied specially with video streaming services quality they received. The QoE-based service evaluation can reflect the customer's perception, which is a critical factor for telecom service providers such as ethio telecom with the vision of a world-class operator and to be profitable in future's competitive markets.

1.2. Statement of the Problem

In general, ethio telecom evaluates video streaming service quality provided to the 3G network subscribers to identify and address network side quality of service issues. The evaluation method, currently applied in the company, lacks the necessary insight into subscriber actual satisfaction level based on location. Due to that the perception or actual satisfaction level of service quality of mobile video streaming users, particularly smartphone users, based on location and contexts are not known.

Failing of not applying user-level experimental based evaluation for smartphone video streaming service quality leads to an inability of knowing the actual satisfaction level of subscribers. So that that the company is unable to take necessary corrective actions before customers are impacted by the poor video streaming service quality then it will create dissatisfaction of customers. This problem often leads to subscriber frustration on the service. Obviously, frustrated customers will not be motivated to watch poor quality video streaming and then users stop watching. Finally, as a consequence, ethio telecom's revenue from this service will decline in the future competitive market.

1.3. The objective of the Thesis

1.3.1. General Objective

The main aim of this thesis is to evaluate the QoE of YouTube video streaming for smartphone users in 3G network of ethio telecom in case of Addis Ababa.

1.3.2. Specific Objectives

The specific objectives of the thesis are summarized as follows:

- To review works of literature on the evaluation of QoE for mobile video streaming service.
- To review QoS and QoE metrics for YouTube video streaming
- To assess methods of QoE evaluation for YouTube video streaming
- To check the current practice of monitoring video streaming quality in ethio telecom
- To evaluate video streaming quality metrics, result in data which will be extracted from the ethio telecom network during the period of experimental evaluation. - Network Level

- To conduct an experimental evaluation on the user's smartphone to assess YouTube video streaming services quality over the 3G network of Addis Ababa using crowdsourcing applications - Application Level
- To conduct a subjective evaluation based on survey and perception rating assessment of smartphone users on YouTube videos, after watching sample videos, using a questionnaire and based on that to calculate Mean Opinion Score (MOS)- User Level

1.4. Research Questions

1. How is the current performance of objective video streaming quality metrics (network & application-level measurements), and overall quality perceived by the end-users of the smartphone (user-level measurement) over the 3G network?
2. How do different objective metrics, contextual and device factors impact QoE of YouTube Video streaming for smartphone users over 3G Network?
3. How correlate the application-level objective and user-level subjective metrics results obtained from experimental measurements conducted in our study?

1.5. Methodology for Experimental Evaluation

Our methodology has been designed to measure smartphone video streaming quality of experience in 3G mobile broadband and to produce a statistical dataset that treats ethio telecom on a fair basis. In this study, QoE evaluation of smartphone video streaming for Addis Ababa subscribers focuses on evaluating the quality of experience of mobile users when customers are using the 3G network while watching video streaming services. Three methods are identified for collecting the required data after identifying and selecting quality metrics for both objective and subjective measurements.

Data is collected from NetProb3010 (NMS) that is currently implemented in ethio telecom to evaluate network-level QoS of video streaming metrics. Moreover, alternative methods of Application and user-level experimental measurements were done using the system model illustrated in Figure 1: 1 below. The figure describes that two crowdsource applications which are Speedcheck Pro for measuring download throughput, and YoMoApp to measure such as initial delay, stalling etc. for application-level data collection. In addition to that user-level perception assessment on content, video, streaming quality and acceptability has been done.

Based on Hardcopy questionnaire survey, annoyance level of initial delay, annoyance level of a number of stalling events and annoyance level of the time duration of stalling events rating responses are considered as inputs to evaluate subjective video streaming quality. The experiment followed the ITU-T P.913 recommendation [4], and this recommendation explains the procedures to be followed when evaluating subjective video quality.

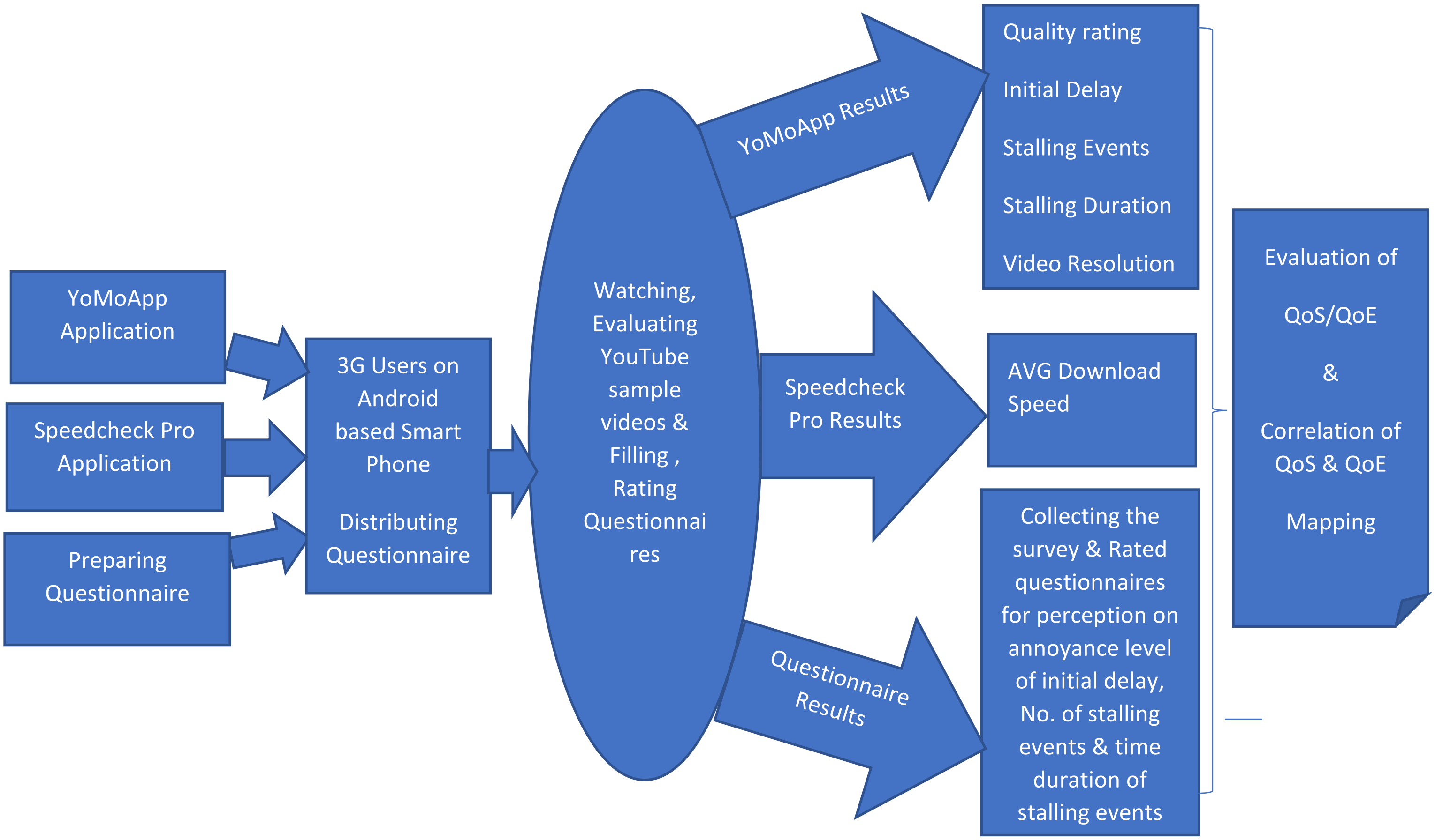


Figure 1: 1 System Model for Application and User Level Experimental Methodology

These three ways of data collection that are network, application and user level measurements have helped us to evaluate and assess the overall satisfaction of the smartphone video streaming users in 3G network. Additional tools like SPSS, Microsoft Excel, Word, Video Downloader, video info viewer, video editor applications (such as Mute video, Video Converter), Google Map and YouTube platform have been used as part of this thesis work. A summary of the methodology used is shown in Figure 1: 2.

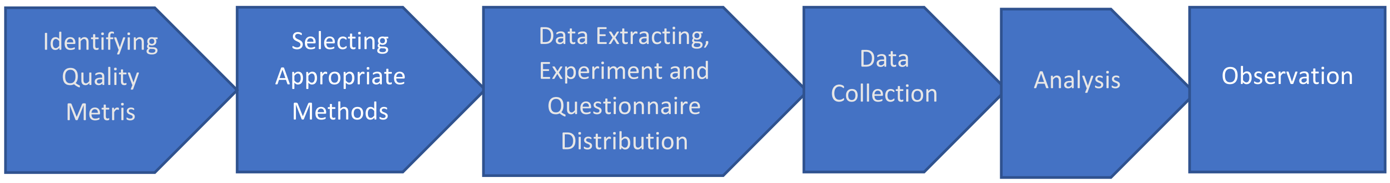


Figure 1: 2 Methodology Used

1.6. Related Works

Prior studies in [5], [6] have shown that stalling/buffering and initial delays on the video playback are the most relevant Key Performance Indicators (KPIs) for QoE in standard HTTP video streaming. The specific case of QoE in YouTube deserves attention, due to the overwhelming popularity and omnipresence of the service. Both studies had considered the "standard" HTTP YouTube video streaming, as well as the more recent Dynamic Adaptive Streaming (DASH) version. In [7] the authors have shown that initial playback delay is generally tolerated by most users, stalling or buffering has a huge influence on user QoE, and a small number of stalling events or frequency severely reduces the QoE [8].

In [9] the authors introduced YoMoApp (YouTube Performance Monitoring Application), an Android application, which passively monitors key performance indicators (KPIs) of YouTube adaptive video streaming on end-user smartphones. The monitored KPIs (i.e., player state/events, buffer, and video quality level) can be used to analyze the QoE of mobile YouTube video sessions. The authors recommend the tool to be used for developing new QoE models of HAS in mobile devices, enabling multiple QoE-based monitoring applications for YouTube traffic, such as dynamic traffic engineering, troubleshooting, load balancing and caching, and many more. Moreover, a subjective study was conducted in which YoMoApp was used to monitor the performance of YouTube video streaming sessions under different network conditions. The measured KPIs for HTTP adaptive video streaming and the quality ratings of the participants are investigated in order to validate the measurement application through subjective QoE tests.

Authors in [10] found that quality adaptation could effectively reduce stalling by 80% when bandwidth decreased in a mobile environment and was responsible for better utilization of the available bandwidth when bandwidth increased. However, quality switches have an impact on perceived quality themselves, as they increase or decrease the video quality according to the switching direction[11]. A more comprehensive survey of the QoE of adaptive streaming can be found in [12].

In [13] the authors created a QoE model by taking into account the content, picture quality, sound quality, matching to interests, fluidness and loading speed as QoE factors. Their work was on interdisciplinary measurement and modelling of Quality of Experience (QoE) related to mobile YouTube video streaming in a Living Lab environment. The relation between QoE factors and technical parameters is also modelled involving network QoS indicators and context information. The concept and implementation of a QoE measurement framework on Android platform are discussed. QoE measurement of mobile YouTube video streams performed in a Living Lab test environment is presented. The result of the measurements is analyzed and the influence of the different QoE factors on the overall QoE related to watching mobile video streams was shown. They found that the content had the largest influence on the QoE of online recommend video content in the mobile context.

Moreover, various research studies have been done regarding the measurement of QoE by considering both measurable and non-measurable metrics in the area of mobile multimedia services. The author in [14] stated measurable metrics are typically those related to the technological aspects of the service. Using these parameters, it is possible to produce quantifiable quality metrics for QoE evaluation of multimedia streaming by performing analysis at the application and network levels. The same author mentioned the so-called ‘non-measurable’ parameters are related to the users’ evaluation of the investigated QoE factors, their behaviour, expectations, emotions, etc. Increasingly, interdisciplinary research is set up in order to take these human aspects into account.

Authors in [15], [16] implemented similar in-application measurements for YouTube-QoE monitoring, relying on users’ application-level tools to directly collect key performance indicators or metrics such as initial delay, stalling events, video resolution, or quality or resolution changes. Application-side measurement method offers exact measurements for QoE evaluation, as these can be directly observed, without the need for extra estimation or mapping methods.

1.7. Scope of the Thesis

The study explores the evaluation of video streaming QoE for android based smartphone users over the 3G network in nine different locations of Addis Ababa where the case study was conducted.

1.8. Limitation of the Thesis

Firstly, the sample groups and the selection of the sample subjects for this study are not big enough to draw a general conclusion, this is due to the subjects are mainly come from the researcher's social network, who are having a similar background. This is due to the limitation of the convenient sampling method for data collection.

Secondly, this study has been done only in nine selected location of Addis Ababa and the number of sample users in the experiment for the selected locations is not equally distributed. Moreover, the number of users in the experiment parts of the day (Morning, Afternoon and Evening) are not proportional these are also due to the reason that the study is conducted based on the users' convenience and willingness. We also tried to study the single parameter of subjective and objective metrics impact on QoE in terms of statistical analysis so, further investigation can be done to combine various parameters and investigate their combined analysis with QoE. Finally, the study has been done only on smartphone users on the YouTube platform, and it does not consider the impacts of Content Delivery Network (CDN). So further studies can be done on by combining feature phones and smartphone users on different video streaming platforms and social media.

1.9. Contribution

In this research, our work mainly focuses on the QoE evaluation of smartphone video streaming from the perspective of network side, crowdsource application and user side over the 3G network. These three methods are used for video streaming quality evaluation in our thesis work. The results of this research help ethio telecom, as a sole mobile service provider, to deal with the quality of the video streaming services provided to its customers. Thus, to consider the perception of its customers on the quality of the video streaming evaluation. Taking the evaluation results as an input for optimizing, enhancing their networks, network planning, network provisioning and traffic management in the 3G network. The author believes the following points are the main contributions to mobile operators, in particular to ethio telecom.

- Addressing video streaming QoE handling process.
- Understanding the QoE of mobile video streaming and additional metrics for evaluation.
- Considering alternative methodologies to use crowdsourcing test tools for evaluating QoE of mobile video streaming metrics.
- Consideration of subjective quality measurement of user perception as an input for future improvements and better customer satisfaction on video streaming services.
- Understanding the current smartphone user's satisfaction level on YouTube video streaming service quality.

In general, the author believes that consideration of network, application and user level measurements for QoE evaluation of smartphone video streaming services in the real environment of the 3G network. Moreover, we tried to customize the subjective quality rating measurement scale of YoMoApp. Here one of the perception ratings, collected from users, are streaming quality that is presented as one single metric and this metric does not show us the actual source of the problem which annoyed users. Therefore, we identified the gap and divided this single metric into three metrics to assess the annoyance level of initial delay, stalling events and stalling duration on hard copy questionnaire using a rating scale for each video session. Finally, we have abled to identify which streaming quality metrics annoyed users more.

1.10. Thesis Organization

The remaining parts of the thesis are ordered as follows. Chapter 2 presents background concepts on 3G networks, and it's quality of services classes, quality of experience and video streaming. Chapter 3 elaborates video streaming quality of experience measurement practice in ethio telecom. Then data planning and methodologies are described in chapter 4 and the data collection instruments that are applied in this thesis are discussed in Chapter 5 which is followed by data analysis, results and interpretations in Chapter 6. Lastly, Chapter 7 presents the conclusion and future work.

Chapter 2

Background on Quality Evaluation over UMTS

2.1. UMTS Network

UMTS in many aspects can be seen upon as an advance on cellular technologies of GSM and GPRS. Third-generation systems (UMTS/3G) have developed with high data rate and image or video quality. The highest changes are related to the access portion of the network. The access network, termed UMTS Terrestrial Radio Network (UTRAN), contains base stations and base stations, controllers.

One of the famous new features of UMTS is supporting higher data rates: on circuit-switched connections 384 kbps, and on packet-switched connections up to 2 Mbps, can be reached. Higher data rates obviously enable some new services [17], such as smooth video streaming and fast downloading of data.

Compared to GSM and other existing mobile networks, UMTS supports a new and important feature namely, it allows negotiation of the properties of a radio bearer. Features that define the characteristics of the transmission may include throughput, transfer delay and data error rate. To be an effective and efficient system, UMTS must assist a wide range of applications that haven't a similar quality of service (QoS) requirements. In the UMTS, there are two access technologies: High-Speed Packet Access (HSPA) and evolved HSPA (HSPA+) which are accessible technologies in Addis Ababa.

GSMA, the representative of mobile operators worldwide, defines HSPA in [18] as a technology which enables operators to upgrade their networks to run at the speed of broadband networks. HSPA includes High-Speed Downlink Packet Access (HSDPA), High-Speed Uplink Packet Access (HSUPA) and HSPA+. HSPA can support a maximum of 14.4 Mbps downlink speed and 5.7 Mbps of peak uplink speed. HSPA+ supports up to 42 Mbps theoretical peak data rate.

Many enhancements in terms of allocated network resources have been made for the handling of "rich" data including multimedia such as video streaming services. However, the available resources still have significant limitations to support smooth video streaming services when a very large number of clients are present in a cell of a network.

2.1.1. Architecture of UMTS

In order to fulfil the expectation of the users, Universal Mobile Telecommunications System (UMTS) is developed in terms of several applications, services and bandwidth. The advancement of the UMTS network is mainly based on the GSM/GPRS network that can support different types of applications such as data, voice and video. UMTS network can efficiently interact with other networks. Figure 2: 1 below describes the UMTS/3G network architecture.

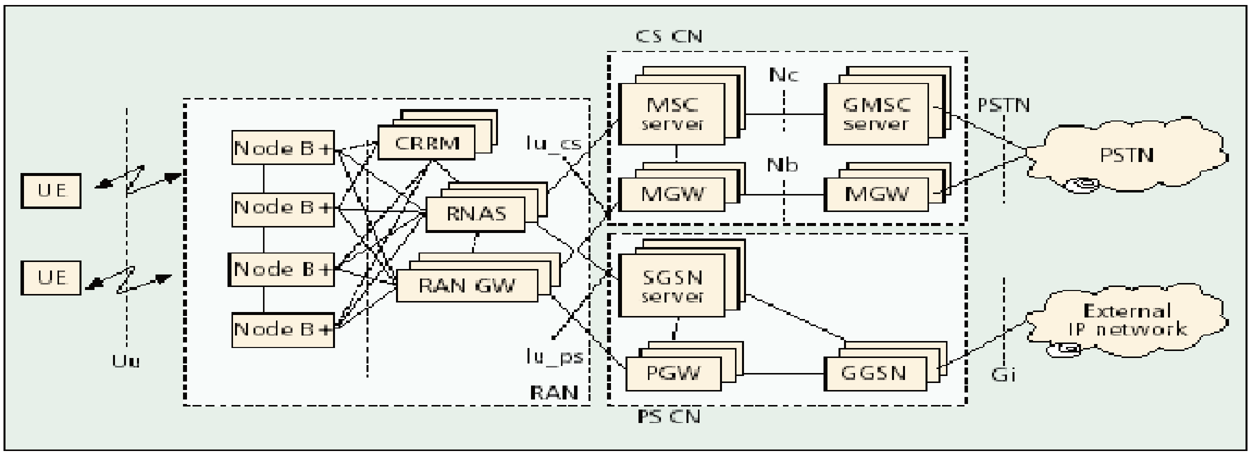


Figure 2: 1 UMTS Architecture [19]

Authors in [19] described the architecture of UMTS and the interaction of network elements to provide service to the users.

The picture in Figure 2: 1 above shows that the Core Network (CN) is the part of the UMTS network that provides services to final users; it can be connected to different types of networks to support different communication protocols. The CN is composed by the Circuit Switched CN (CS CN) and a Packet Switched CN (PS CN). Regarding the CS CN, the Mobile services Switching Centre (MSC) switches the CS connections and manages user mobility. It is interconnected to external networks through the Gateway MSC (GMSC). The MSC is divided into MSC server and Media Gateway (MGW), where all user data are managed. MSC server can manage more MGWs, allowing higher network scalability when newly implemented services increase data flow. The PS CN is composed by two elements: the Serving GPRS Support Node (SGSN), that manages the different sessions (including various QoS requirements) and mobility, and the Gateway GPRS Support Node (GGSN), that performs, among the other things, the QoS mapping among different networks, the packet filtering and the connection with other packet switching external networks (such as Internet). The streaming server providing multimedia services can reside either in the external packet data network or just before the GGSN. The PS CN is connected to the Radio Access Network (RAN) through the RAN Gate Ways (RAN GW) and the RAN Access Server (RNAS). The RAN GW is a routing point for the user traffic between CN and UTRAN. All radio resource management functionality is pooled in the Common Radio Resource Manager (CRRM). It is thus possible to provide the radio resource management network-wide to all radio technologies, which allows an efficient utilization optimization. The RNAS main task is to provide an interface to the CN.

The public land mobile network of UMTS architecture has two main divisions:

- UMTS Terrestrial Radio Access Network (UTRAN):- It manages all radio related functionalities.
- Core Network:- It is responsible for maintaining subscriber data and for switching voice and data connections.

2.1.2. Quality of Service in UMTS Network.

In relation with quality issues, GSM/CDMA network was required only a single QoS to support voice transmission. Unlike GSM/CDMA, different types of applications are supported in the UMTS network in order to accomplish the subscribers' service expectation such as voice and video telephony, web service and video streaming etc.

In a UMTS network, every application needs a different level of QoS according to the service features. QoS is one of the important factors to maintain and provide better service provisioning for subscribers. As a result, different types of classes require different types of QoS in order to provide the service for different types of applications based on their transmission delay sensitivity.

2.1.3. UMTS QoS Basic Classes

There are four different traffic classes that can be obtainable in the UMTS network quality of service provisioning. The authors in [20] the authors described each traffic class in-depth, and we have summarized each traffic classes below:

Conversational class - is highly delay-sensitive because it involves with near to real-time communication. Generally, this class requires a low end-to-end delay for real-time application. This class supports voice and video telephony.

Interactive class - is mostly used for client-server application for remote service. In this type of service, the users request to the server for information and the server responses. Buffering is allowed in this class, there is no guaranteed bit rate because of using the best-effort service model while the delay factor is also kept at the minimum.

Streaming class - is used for streaming multimedia such as high-quality audio and video streaming. Streaming class is little delay tolerance, buffering is allowed in this class. Some of the services categorized in this class are video downloading, high-quality video and audio streaming, still image transferring and watching online TV etc.

Background class: - is used for best-effort service in order to download emails and files. This class is highly delaying tolerance and the traffic of this class has the lowest precedence. Buffering is allowed in this class and high variable delay acceptable. Bit rate does not provide any guarantee because of using the best-effort service model. Table 2: 1 below shows each traffic classes' fundamental characteristics and their application area in the UMTS network of telecommunication services.

Table 2: 1 UMTS Quality of Service Classes

Traffic class	Conversational class	Streaming class	Interactive class	Background class
Fundamental characteristics	-Preserve time relation (variation) between information entities of the streaming with Conversational pattern (stringent and low delay)	-Preserve time variation of services between information entities of the stream with Little delay tolerance	- Request response pattern by Preserve payload content	-The destination is not expecting the data within a certain time and Preserve payload content
E.g of the App.	Voice	video streaming	web browsing	telemetry, emails

2.2. Quality of Experience

There are so many QoE definitions found in different kinds of literature and organizations. Some of the international organizations recommended the following definitions: According to ITU, QoE is defined as the overall acceptability of an application or service, as perceived subjectively by the end-user [21], while the European Telecommunications Standards Institute (ETSI) defines

QoE as a measure of user performance based on both objective and subjective metrics measurement using an ICT service or product [22]. Although there is no common answer for what QoE is, we can draw the assumption that: QoE is a kind of user perception on their satisfaction level of the provided service during the process of the user being in service. Moreover, QoE is a subjective evaluation of user perception and requires conversion into quantitative data for evaluation of multimedia technology such as video streaming.

2.3. Video Streaming Technologies

Video streaming is categorized into two types as progressive and adaptive streaming. In progressive video streaming, the whole video data is downloaded at once with the fixed display quality. An initial delay is introduced at the beginning of the video session in order to accumulate a certain amount of video data that prevents from the buffer starvation (empty buffer). The buffer starvation can, however, happen because the quality of the video is not adapted to the network bandwidth.

In adaptive streaming, while the video is being downloaded to the buffer of the video player, the user starts watching the content. Usually, based on the network bandwidth, the rate at which the video download speed exceeds the rate at which the video bit rates the video playback will be smooth. HAS is one of the video delivery application of adaptive streaming technology for video transmission over the internet.

Streaming technology minimizes the amount of time it takes for a viewer to start viewing media over the internet after starting a data transfer; whatever media you're watching is downloading while you watch. One of the key reasons that producers use streaming servers rather than using regular HTTP web server is because once the video is stored (or cached) on a hard drive, it’s very easy to duplicate. Streaming video can be cache-less, which makes it inherently more secure.

2.3.1. HTTP Adaptive Video Streaming (HAS)

In HAS the content provider streaming server divides the video into chunks of fixed length (e.g. 2 seconds) of different quality, and the player, on behalf of the user, requests the video quality that suits to the device capabilities (e.g. screen size) and available bandwidth [23]. Accordingly, to overcome the problem of smooth video streaming, and to allow for a flexible adaptation of the video quality to the available network resources and device capabilities.

In HAS video content is available in multiple bit rates, i.e., quality levels, and split into small segments each containing a few seconds of playtime. The client measures the current bandwidth and/or buffer status and requests the next part of the video in an appropriate bit rate, such that stalling is avoided, and the available bandwidth is best possibly utilized. Hence, the control intelligence which segment to stream has moved from the servers to the clients.

HAS uses HTTP protocol over TCP and for many content providers, HTTP has become the preferred protocol for video delivery over the last few years. HTTP streaming combines advantages such as firewall pass-through and easy network address translation, but also the benefits of TCP, i.e. congestion control mechanisms and reliable packet delivery. In traditional HTTP video streaming, the video is downloaded as a single continuous manner which represents a single quality setting.

Moreover, video buffering is employed as an additional measure to compensate for jitter and short-term bandwidth variations. Typically, each video session can be divided into two buffering phases, i.e. the start-up phase and the steady-state [24]. During the start-up phase, the player will download the first part of the video as fast as possible to quickly fill the buffer and minimize the initial delay before the playback begins. Once the buffer has been filled up to a specific threshold and the playback has started, the video session goes into the steady-state. This phase is characterized by ON-OFF cycles, also referred to as pacing, where the download is paused as soon as the buffer has been filled and stops when it is reaching depletion.

HAS in contrast to traditional streaming, HAS videos are split on the server into multiple segments, each one corresponding to a few seconds of playback time. Each segment is encoded in a range of different quality profiles which are defined by the content provider. Instead of requesting the entire video, the player performs HTTP requests to fetch consecutive segments. The quality profile of the next segment is determined as a function of the download throughput with which the previous segment was downloaded and the available seconds of playback in the buffer. In this way, the representation of the video can change dynamically to adapt to changes in the network and minimize interruptions or stalls.

The diagram in Figure 2: 1 below illustrates the communication between HAS server and HAS client [25]. HAS server divided the media files into “segments”, which can be encoded into multiple bitrate versions and assigned to a unique URL. Different versions may have different bitrates, resolutions, formats, languages, and other characteristics. A HAS client requests the proper bitrate version to adapt with the dynamic bandwidth variation of the network. Based on the request the HAS server sends the segments or small chunk of the video files to the requester HAS client.

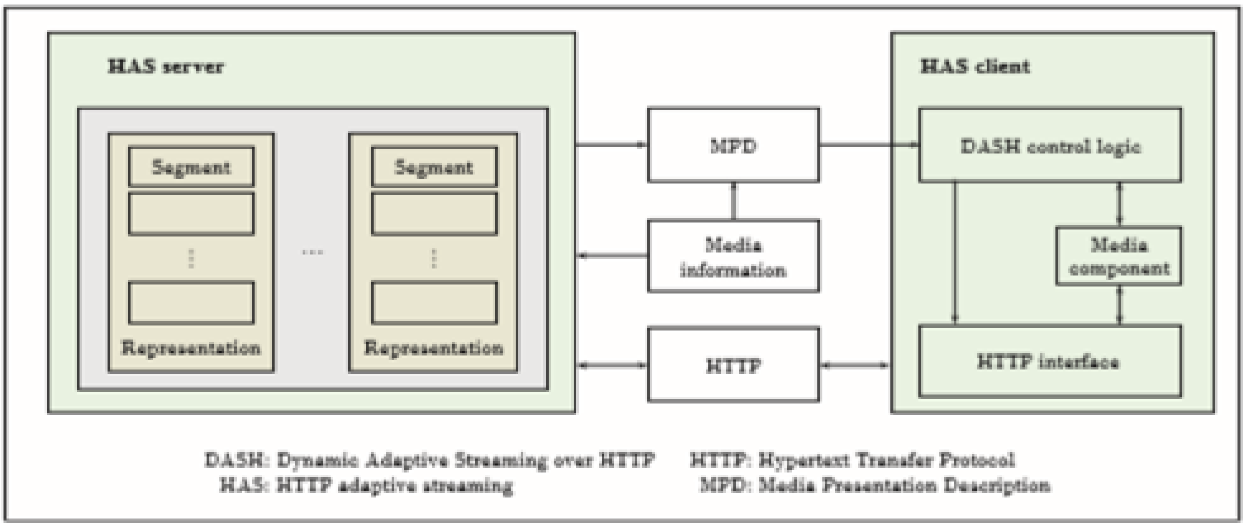


Figure 2: 2 Diagram of Http Adaptive Video Streaming [25]

2.3.2. Mobile YouTube Video Streaming

YouTube is a very popular streaming service watched by worldwide users. YouTube is one of the most famous video content (VoD) provider that is owned by Google and has a high share of video traffic over the internet globally. It uses adaptive streaming for delivering video traffic to users. Over 1.9 Billion logged-in users visit YouTube each month, and every day people watch over a billion hours of video and generate billions of views and More than 70% of YouTube watch time comes from mobile devices [26]. YouTube service uses a transport mechanism based on the HTTP/TCP protocol stack which is called progressive download [27]. This mechanism makes it possible for the video to start playing before its full content has been downloaded. At the receiver side, the video player includes a buffer that stores the video content whilst it is being consumed. The purpose of this buffer is to compensate for the jitter introduced by the network. As soon as there is sufficient amount of data in the buffer, the playback starts.

The strategy of YouTube streaming varies for different applications such as web browser or mobile application containers such as Silverlight, Flash or HTML5 are used and Uses TCP traffic for video streaming [28], [29]. Users watch YouTube either in personal computers (PCs) or in Mobile devices (Android, iOS, Windows etc.) by using the web browser or mobile application. Mobile applications for YouTube streaming used in different mobile devices are different and follow their own streaming strategies.

HTML5 is a container which is used as the default for streaming videos in mobile devices. Recent developments supporting HD videos use Flash containers as default. Streaming strategy for flash containers is independent of the application used but for HTML5 containers, the streaming strategy is dependent on application used [28].

2.4. Factors that Affect the QoE of Users in Video Streaming

Video quality is affected by a multitude of factors in every stage of its existence from the creation of the content to the transmission and finally reproduction. There is a vast amount of research on the quality of the recording equipment, lighting conditions and other factors that contribute to quality at the content creation. Our interest here is to show the general influencing factors that affect the quality of video streaming during the process of encoding, transport and presentation of the multimedia. Figure 3-2 shows factors that affect QoE of network services such as video streaming [30]. As described in the figure, the network (download throughput), users’ device, task application, environment, content context and human factors also influence the Quality of experience of video streaming services provided to the users. Thus, there exist external factors influencing the user-perceived quality. Further, QoE metrics are interdependent.

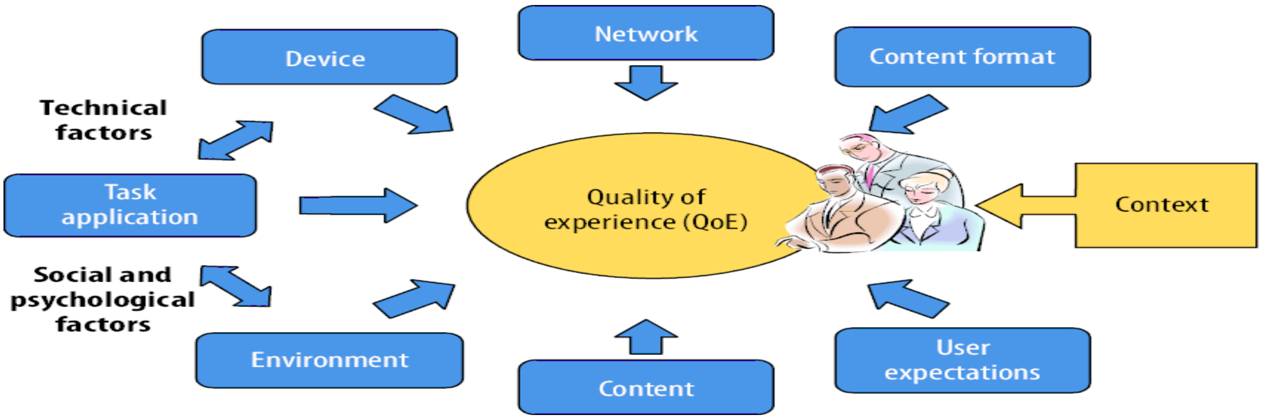


Figure 2: 3 Factors that Affect QoE [30]

2.5. QoS Metrics in Video Streaming Service

Authors in [31] explained the QoE performance metrics of video streaming services as it is important to arrive at service performance metrics through quantifiable means that can be derived from measured parameters. The Quality of Service is expressed by network parameters like bandwidth, delay, jitter or packet loss. However, a good QoS does not guarantee that all customers experience the service to be good. Some of these metrics depend directly on network and plant characteristics that MNO has control of through network maintenance and operations such as; proper configuration, capacity management, QoS strategy, etc. The traditional network-centric QoS metrics such as packet loss, packet reordering, and delay are not sufficient to measure the satisfaction of the users.

2.6. QoE Metrics in Video Streaming Service

The QoE metric is defined and used by the scientific community to quantify the users' perception of the quality of video streaming applications. These metrics are described as QoE Key Performance Indicators (KPI). A QoE KPI is a user-side metric which can capture the user' perception of a service. The VoD streaming sessions last from several minutes to hours. If there is a noticeable delay after a video link is clicked, or if there is an interruption during the playback or any perceivable drop in the visual quality of the video, these can affect the perceived quality of the user. Generally, video streaming QoE metrics have been classified into the following categories based on reference [32].

• **Full reference(FR) metrics:** Both the reference and outcome are available and, thus, enable detailed subjective and objective comparisons of the videos. These metrics are suitable for traditional broadcasting and television systems. Eg. Peak Signal to Noise Ratio (PSNR) [33], Structural Similarity (SSIM) [34], and Video Quality Metric (VQM) [33].

• **No reference(NR) metrics:** Only the outcome is available, and the quality is to be estimated with no reference. These types of metrics are more applicable to online services where only the outcome at the user end is available. In video streaming services, it is hard to determine if the discrepancy in the quality is due to the quality of the reference or due to the intermediate elements.

• **Reduced reference(RR) metrics:** The same set of parameters are derived for both the reference and the outcome. These parameters could be at the application layer: bitrate, framerate or at the network layer: packet loss.

The following some fundamental QoE metrics are used in adaptive video streaming service evaluation:

- Initial delay or Startup delay: it represents the duration (measured in seconds) between the time that a user initiates a video session and the time that the media player starts playing video frames.
- Stalling/Buffering events: When the player buffer is empty, there are no more video frames to display at the screen, the video stops playing (interruption). This event is called buffer starvation.
- Stalling/buffering time: it is the time duration of buffering events occurs during the video session. it is computed as the time duration within each video player interruptions.
- Stalled/Buffered time rate: This metric indicates the percentage of the total stall time out of the total length of the played multimedia file (audio or video).
- Average bitrate/video resolution: The video quality is denoted by the bitrate in kilobits per second (kbps). In Variable Bit Rate (VBR) video or multiple bitrate video, the average bitrate is the mean quality of the streamed video resolutions.
- A number of bitrate/video resolution switching: it denotes the number of times the video quality changes during the session.

2.7. Video Streaming QoS and QoE Performance Evaluation Methods

The term QoE has been frequently used to measure or assess user experience in related to a service, especially in networked services. QoS is an objective measurement of technological performance and it is network-oriented, such as network-level capacity performances (e.g., throughput, initial delay, number of stalling events and it's time duration, resolution, error rate, latency, etc.) and device capabilities and product features (e.g., battery lifetime, video bitrate, frame rate, screen size, operating system and its version etc.), but does not deal with user's overall experience. The goal of QoS of video streaming is to deliver a high QoE which came after the concept of QoS.

QoE of video streaming can be evaluated based on the end-users responses, and it should reflect multi-dimensional effects. To measure QoE of video streaming, a great number of QoE metrics for perceived video quality have been developed and used for quality evaluation in mobile video streaming service. However, these metrics are limited in taking into consideration only some influencing factors of user experience. QoE, therefore, is proposed to involve human dimensions into the measurement of multimedia service performance, together with the objective technical aspects.

QoS of video streaming is what the operator can objectively measure in the networks but QoE is the subjective method which is based on user survey and experiment can reflect the experience of the user on video streaming quality more directly and match well with their perception about a video streaming service.

We have summarized the QoE of video streaming measurement techniques from the overall perspective of HAS in the next section.

2.7.1. Subjective QoE Evaluation Techniques

Subjective QoE evaluations such as surveys are the most effective methodologies to assess Quality of Experience, as they try to consider all the fundamental elements affecting user's experience by gathering scores and opinions provided by human observers. These techniques usually generate more representative results compared to pure technical indicators; unfortunately, they also represent the most expensive and time-consuming way to assess video streaming services. Subjective QoE metrics are used to measure the satisfaction of users in video streaming sessions. The subjective assessment methods are divided into two types: QoE feedback and model-based QoE.

QoE Feedback: The QoE score is decided by the feedback scores or rates collected from the human subjects based on their experience of video streaming quality playback. However, the feedback rate can be influenced across human subjects, since the subjects are diverse in physical and psychological confounding factors and collecting feedback is time-consuming. To get an unbiased and general QoE score, the introduction of statistical analysis techniques is required.

One of the most popular subjective QoE feedback metrics is the Mean Opinion Score (MOS). For getting MOS, limited sets of human subjects are exposed to watch a video and asked to rate the experience of the streaming session. The MOS is a five-point discrete value (e.g. Excellent, Good, Fair, Poor, and Bad). And the QoE score is calculated by averaging the MOS given by the users.

Model-Based Subjective QoE: Collecting feedback from the users is time-consuming and has a limitation on real practice. Thus, it is feasible to establish a QoE model to estimate the subjective QoE scores from objective metrics. This is an automatic, quantitative and repeatable manner. There are two model-based methods: (1) learning-based, which uses learning techniques to map the objective metrics to MOS; (2) heuristic methods, which estimate the subjective QoE scores by some manual functions.

2.7.2. Objective QoE Evaluation Techniques

Using QoE to represent user satisfaction has been widely recognized by the telecom industry, the quality of service quantified using some tools objectively. According to the methodology and location of data collection in the network, they classify QoE measurement methodologies for online video services into the following three categories: client-side, in-network, and server-side measurement.

Client-Side Measurement: There are passive measurement and proactive (active) measurement in the client-side, where some tools are used to measure users' QoE directly. In this case, the measurement is completely depended on the users and the tools have no control over the video content or duration.

In-Network Measurement - Measurements of QoE within the network [35] do not require modifying client or server software. It just overhears IP packets passing through links and estimates the QoE of video streaming in the application layer. It is easy for network operators to deploy these measurement tools. By tracking the packet-level information of each session, the objective QoE can be estimated, including stall duration, startup delay, etc.

Server-Side Measurement - Server-side measurements [36] collect each HTTP data packet on the server-side and rebuild the HTTP session, by which they can measure video streaming metrics.

Streaming video is becoming a common application in 3rd and 4th generation cellular networks. The need for value-added services in cellular networks is growing rapidly and thus the 3rd Generation Partnership Project (3GPP), which defines the standard for the 3rd generation cellular networks, dedicated a special workgroup for streaming video [37].

The problem is the low and fluctuated bandwidth of the network that makes the use of mobile video streaming services such as YouTube while maintaining steady and smooth video streaming with high QoS and QoE is very difficult. So, it is important to evaluate the quality video streaming services provided to customers by considering both subjective and objective measurement methods help MNOs to give better service.

Chapter 3

Video Streaming Quality Management Practices in Ethio telecom

3.1. Introduction

Before collecting data for analysis purposes, document review has been done on the video streaming quality handling process currently in practice in ethio telecom. Within the company, there are divisions, departments, sections and groups for its day to day business activities. From all the divisions, our focus was to study the Customer Service, Customer journey & Quality management, and Network division which are responsible for quality issues and customer complaints related to UMTS video streaming services.

3.2. Video Streaming QoS management in Ethio telecom

In ethio telecom, the Service Management Center (SMC) section which is found under the network division has responsibilities for monitoring and analysis of video Streaming quality from network side using NetProb3010 tool.

NetProb3010 is a tool which vendor dependent owned by Huawei and used to do a day to day work of the SMC section. This tool has the functionality to provide the following jobs regarding video streaming quality [38].

- Monitors KQIs for specified network technologies such as 3G in real-time, and updates the KQI data every 5 minutes, helping ethio telecom to detect potentially service-affecting issues in a timely manner.
- Analyzes KQIs from multiple dimensions and identifies service quality trends, helping ethio telecom to improve streaming service quality and user experience.

Based on the tool, video streaming KQIs are Video Streaming Start Success Rate, Start Delay, Stall Frequency, Stalled Time Rate and Plays Disconnected Rate can be monitored and analyzed every 30 and 60 minutes.

3.3. Video Streaming QoE Management in Ethio Telecom

The preliminary assessment has started by evaluating whether ethio telecom has a process and procedure for handling of complaints, and proactive methods for evaluating video streaming service in UMTS or not. For example, if a smartphone customer complaint about video streaming quality issues when accessing video streaming service in UMTS, how is the complaint handled? And if ethio telecom wants to evaluate the current level of smartphone users on the quality of video streaming services satisfaction levels' in UMTS, how ethio telecom proactively evaluates user-level video streaming quality considering context, and other factors before it impacts customers?

Our document analysis and discussions with Customer Service division proved that there is no clear process or procedure for handling such complaints, and methods of proactive approach for user-level subjective evaluation of video streaming quality. The same assessment and discussions are done in Customer Journey and Quality management, and Network division in the specific section of SMC but we got a similar response which is similar to customer service division.

A summary of the gaps identified during the investigation are listed below:

- There is no clear process for handling and following video streaming service quality-related problems or complaints reactively.
- The objective level measurement method of video streaming quality, currently applied in ethio telecom missed considering some basic important metrics such as average video resolution watched etc,
- There is no clear process for the proactive approach of conducting user level subjective experimental evaluation for video streaming quality based on network technology, location etc.
- Sometimes, video streaming quality-related complaints are handled through public meetings and management meetings.
- video streaming service quality-related complaints are explained by customers as a network problem, and the problems are registered by ethio telecom call centre staffs as complaints of network-related problem. This is due to the lack of knowledge on video streaming quality metrics by both the call centre staffs and customers.

3.4. ITU Service Quality Performance Management Process

By detecting the problems, the assessment was done to see if there can be a process or procedures for handling quality problem issues proactively and reactively. ITU [39] recommends that there must be a quality of service measurement for complaints and performance degradations coming directly from customers and proactively by conducting panels or surveys from the perspective of QoE management. A detailed description of the information flow to support performance management or evaluation is shown below in Figure 3: 1.

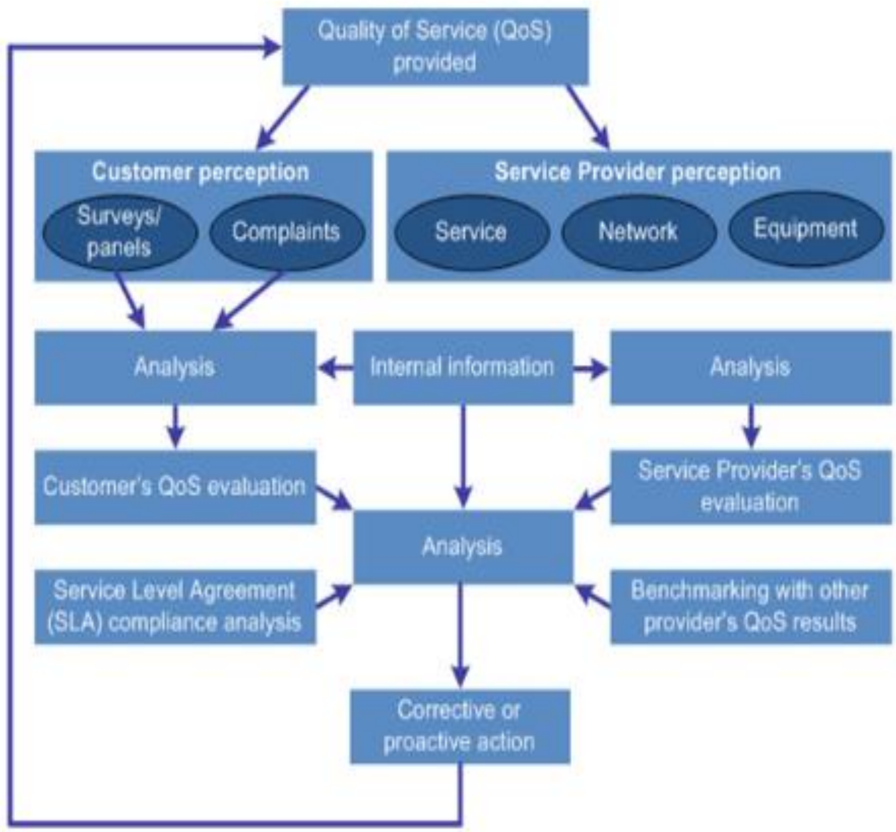


Figure 3: 1 ITUs' QoE Management Process for Telecom Service Quality Assurance [39]

As shown in Figure 3: 1 above, an evaluating or a monitoring mechanism for the proper provisioning of quality of service to customers. In the figure, the network service provided to customers should be monitored or checked in two ways that are based on the perception of customers and service provider.

The figure describes that telecom service provider assess customers perception proactively by conducting surveys or panels, based on customers complaints and informal information such as meeting, news etc. Then after by analyzing the data collected from the above mentioned three sources together with service level agreement (SLA) compliance analysis the service provider should evaluate the service provided to customers. Based on the analysis and evaluation necessary corrective actions will be taken. Moreover, the telecom service provider assesses and monitors the quality of service (from the perception of service provider) through service, network and equipment point of views. Finally, by taking informal information and benchmarking other provider's QoS results as input to take corrective or proactive actions. In ethio telecom, a focal point for the communication between customers and the service provided, regarding service quality, are mainly customer service and customer journey & Quality management divisions.

3.5. Proposed Solution for Video Quality Management in Ethio Telecom

As stated by the author [40], there has to be a communication with customers when there is mobile data throughput quality received by customers which include video streaming service quality-related problems or complaints.

Customer service is the interface for the customers, and issues which cannot be resolved by customer service will be communicated to SMC if they are related to video streaming quality complaints in UMTS. SMC section can also communicate with other departments of the Network division for resolving the complaints received from customer Service. In addition to this, if there are problems which cannot be resolved by the departments under Network division, there will be communication with vendors for further support and investigation. Finally, as the communication is bi-directional, customers have to be notified through Customer Service for better customer satisfaction by addressing the received complaints.

So, we have done some modifications on the recommendation of authors [40] for complaint handling mechanism to manage the quality of experience. We added a proactive approach of conducting surveys or panels for quality of experience evaluation management of video streaming services based on ITU recommendation stated in the above Figure 3: 1. Then we incorporated the proactive approach of conducting surveys or panels on the authors [40] recommendation. The modified method to evaluate the quality of experience for video streaming service is shown below in Figure 3: 2.

Based on the proposed method depicted in Figure 3: 2, customer journey and quality management division by preparing a user-level subjective survey or panels to assess basic video streaming service quality-related metrics or KQIs. This metric includes such as poor video streaming quality performance variation, initial delay, frequency of stalling, stalling time duration, low quality while watching video streaming online. The division communicates the poor performance evaluation results to network division in order to be taken proper proactive actions before the problems impact customers satisfaction and revenue loss. Therefore, by considering complaints and conducting surveys or panels to assess video streaming quality problems. Finally by introducing and applying the reactive and proactive actions on video streaming quality problems the company can increase customer satisfaction and revenue maximization.

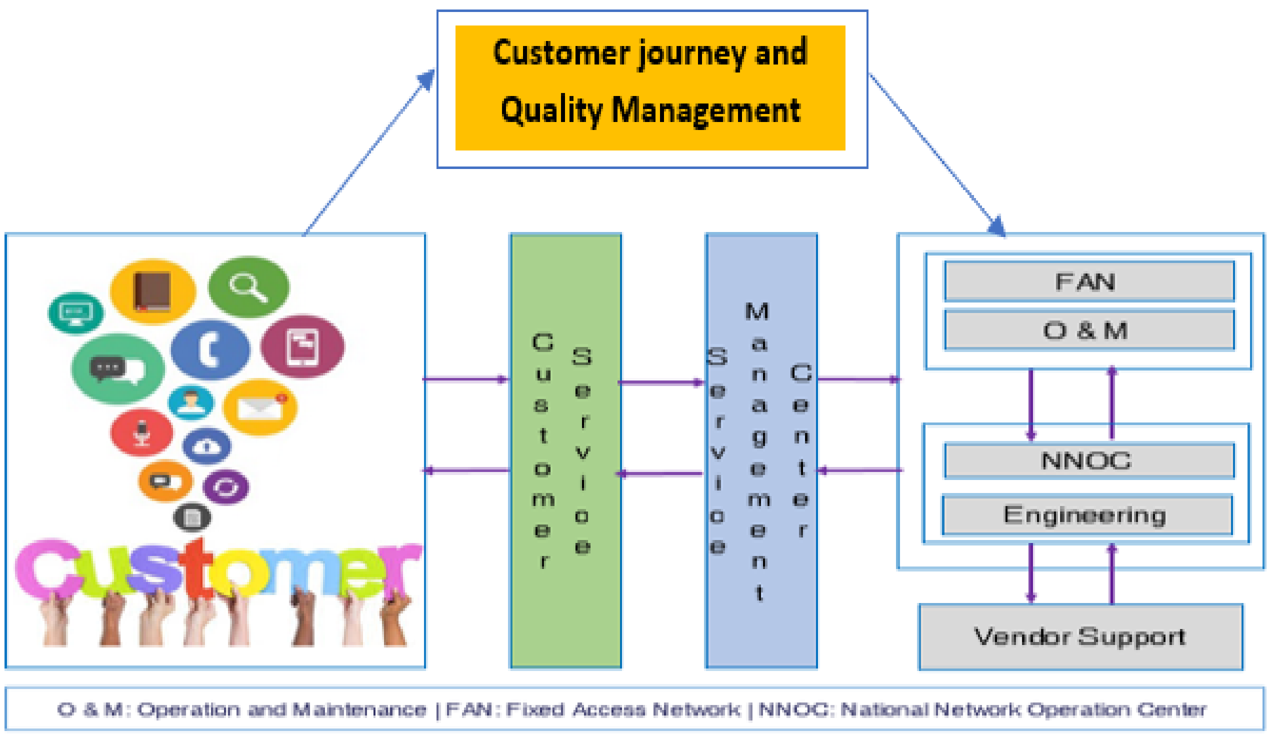


Figure 3: 2 Network Architecture Process for video quality management

Chapter 4

Planning and Methodology for System Elements

4.1. Introduction

The main objective of this chapter is to explain the planning and data collections methodology for evaluation of the quality of experience of video streaming for 3G subscribers in Addis Ababa. We used both subjective and objective measurements methods that are NMS, open-source crowdsourcing applications, survey & perception assessment questionnaire.

4.2. Plan for Taking YouTube Video Streaming Measurements

After assessing the quality handling process, the plan was to conduct the research in Addis Ababa. For network-level QoE related metrics (KQIs), the evaluation has been made using NetProb3010 (NMS). For application-level QoE related metrics two applications will be applied simultaneously, those are network speed test using Speedcheck Pro tool and YouTube Monitoring Application (YoMoApp) tool by considering crowdsource dataset. In addition to objective quality metrics, the data collection considers subjective metrics which shows the perception of the users has been planned to conduct using a questionnaire. Covering all places and all groups of society cannot be accomplished in this thesis work within a short period of time. Then, we have planned to focus on a specific society and location of urban areas that have coverage of 3G access network.

Moreover, for this study we have planned to use one of the methods for non-probability sampling that is convenience sampling, this method has been chosen and considered as the main and suitable approach for data collection due to time limitation for this study. The study involves as many volunteers of 3G mobile users as possible to explore their satisfaction behind YouTube streaming experiences since convenience sampling doesn't seek to choose specific respondents.

Our target is on the 3G mobile customers by identifying those who have a good habit of internet video watching using pre watching survey questionnaire. After selecting the locations for our case study, some indoor such as a residential house and office in the building were chosen. In addition to that outdoor environments or contexts such as open-air restaurant and cafe, were identified. Most of the identified locations had the problem of smooth video streaming service quality.

The next plan is to use an alternative tool for collecting quality indicator metrics is related to video streaming. We have found YouTube Monitoring Application (YoMoApp) and Speed test (Speedcheck Pro) tools which are a crowdsourcing tool and can be downloaded free from Google Play Store.

Selection of the participants in the survey was made through Customers of the 3G mobile network from my social network, and their work or home is near the identified places were selected to conduct the tests. Ethio telecom employees have been selected as a major candidate for conducting the test since they have a free Internet UMTS package which can make the test easier for data collection. The rest participants were compensated with voucher cards, which value 50 Birr, had been given for each individual as motivation and for their participation, which proved to be sufficient for achieving accurate involvement in the study.

4.3. Methodology for System Elements

This section discusses the general overview of the experiment methodology of system elements or parts in this application and user level YouTube video streaming quality evaluation.

Subjects (Users) - From this point onwards we will call the 3G subscribers or users that were involved in the experiment as subjects. The subjects who conducted the video streaming evaluation had not any training or experience on video quality evaluation before. Most of them were employees from governmental and non-governmental organizations. An attempt has been made to include subjects with a different profile including age, gender and level of education. A total of 40 people (22 male and 18 female) from different walks of life participated in the experiments of our video streaming evaluation. Five subjects were filtered as invalid and excluded from the final analysis.

Table 4: 1 below shows a summary of the subjects' profile and most of them are males which accounts for 65% and the rest 35% are females.

Table 4: 1 Summary of Subjects' Profile

Sex	Male	23
	Female	12
Age Groups	Under 18	1
	18 - 24	5
	25 - 34	9
	35 - 44	10
	45 - 54	7
	Above 55	1
Educational Backgrounds	PhD	1
	Masters	5
	Degree	19
	Diploma	5
	Certificate	4
	High school & below	1

In addition to their age, gender and employment status the participants were asked to subjectively assess their video watching habits in order to identify and include the subjects in the evaluation. Then only subjects with a good and excellent experience had been chosen for evaluation and thirty-two of them said that their video-watching habit's or experiences are excellent while three of them said they have a good video watching habit's or experiences. Regarding eye health or normal vision, all the subjects have self-reported as no eye problem in their medical history.

Collection of Videos - Various video databases have been searched to look to choose an appropriate database for the video streaming QoE evaluation that was to be conducted. The major focus was to have original videos that have been presented to the ultimate viewer to watch. There are many sample videos and video databases available for video quality evaluation online with and without impairments and degradations, then we selected video samples from YouTube and free stock video database without impairments and degradation for evaluation. Then we have edited and sorted the video formats for the purpose of our experimental study. In addition to the free stock video database, YouTube is chosen as an additional source database for our evaluation sessions because of its universality and it accounts for 20-35% of all videos on the internet [41].

Finally, eight videos were selected totally from the first 30 long listed sample video collections. All sample videos were undistorted, without impairments and with their average time duration of 30 seconds. Further, in order to accommodate all subject's preference in the evaluation we have considered video categories of the movie, sport, news, advertisement, slow motion, lecture (education), music and cartoon. All videos that we have selected as a sample had MPEG-4 file format, encoded with Advanced Video Codec (AVC) with 30 Frames Per Seconds (FPS), their average bit rate was 1500 Kbps and their video resolution was 720.

For the purpose of this study, we used YouTube service as a platform, and we were created a YouTube channel which has named as zel fit. Afterwards, the eight videos were uploaded on the channel that we had created. Figure 4: 1 below shows the representation frame screenshot of the eight sample video clips used in the experiments.

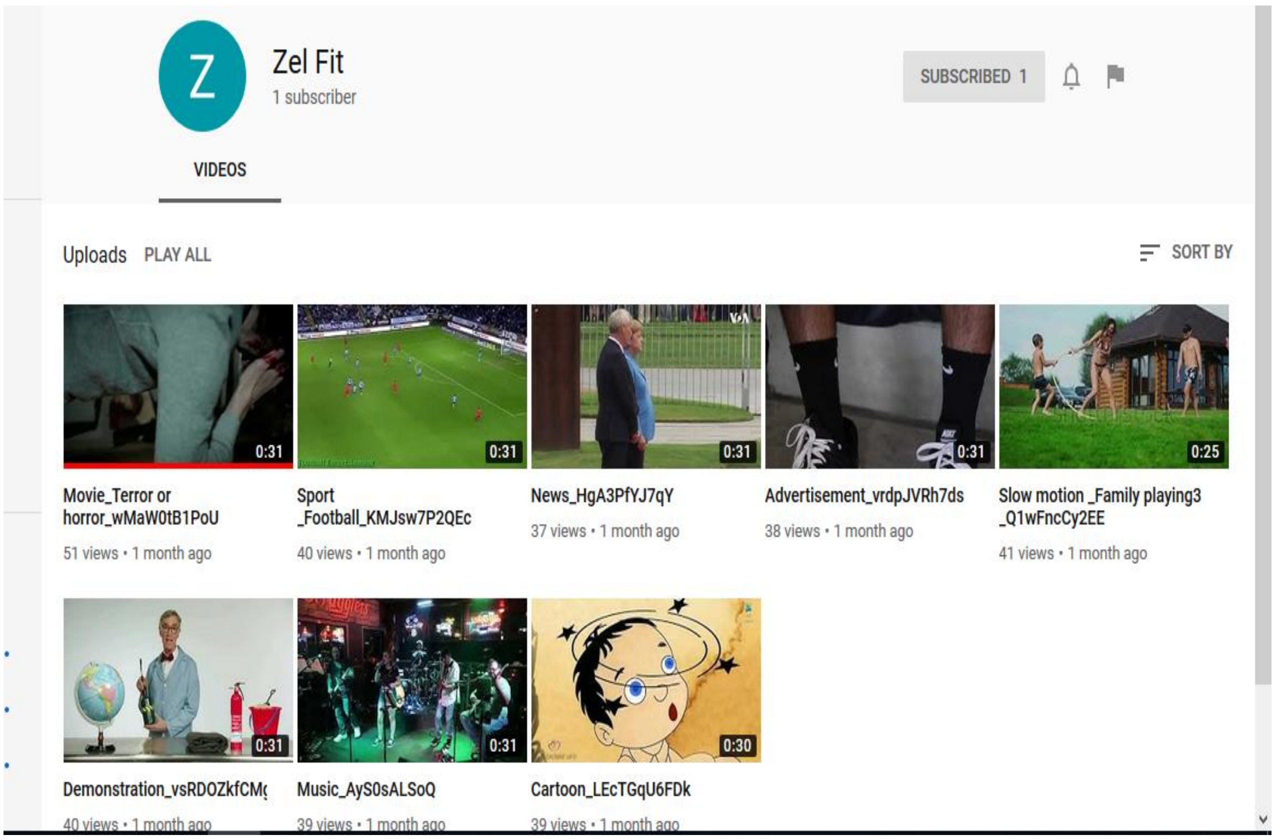


Figure 4: 1 Representative Frames Screen Shot of the Sample Videos in the zel fit Channel

Table 4: 2 shown below describes the category of the video content, the YouTube ID of the videos, time duration of each video and a short description of videos.

Table 4: 2 Description of Sample Videos for the Experiment

S. No	Category	YouTube ID	Duration (Sec.)	Description
1	Movie	wMaW0tB1PoU	30.31	Horror- Action
2	Sport	KMJsw7P2QEc	31.08	Sports scene of football player scoring goal
3	News	HgA3PfYJ7qY	30.38	News on the shocking scene of German Prime minister
4	Advertisement	vrdpJVRh7ds	30.99	Man Running with Nike Sport shoes
5	Slow-motion	Q1wFncCy2EE	24.74	Mother playing with her kids' slow motion
6	Education	vsRDOZKfCMg	31.13	Man demonstrating on Global warming
7	Music	AyS0sALSoQ	31.06	Singer & Full band performance on the stage
8	Cartoon	LEcTGqU6FDK	29.74	Fast-moving cartoon scene on valentine day

Context of Environment - Context of the environment plays a much more critical role in the assessment of QoE in mobile networks than in fixed networks. For example, a user who is indoor would experience a very different service quality from those who are outdoors. Location and time of day are other aspects of the user context that have a big influence on the QoE. The video streaming evaluation for this thesis work was conducted in a real-life environment of users at nine different sample locations of Addis Ababa, the capital city of Ethiopia. Serious attention had been made to eliminate any glare and disturbing noise during the assessment sessions. These conditions were chosen so that everything would be like real-time video watching.

The experiment was done during three different periods of time based on parts of the day at the indoor and outdoor environmental context. The time period from 8:00 AM up to 12:59 AM is considered for the morning session, time period from 1 PM up to 6:59 PM is considered for the afternoon session and time period from 7 PM up to 10 PM is considered for the evening session. Table 4: 3 depicts that the experiment time parts of the day with a respective number of subjects and percentage of subjects involved in the experiment. Most of the video experiment, around 43%, were evaluated in the morning session. The reason for this result was due to the experiment had been made based on subjects' convenience time.

Table 4: 3 Experiment Parts of the Day with Respective No. of Subjects Percentage

Parts of the experiment day	No. of Subjects	Percent
Afternoon	11	31.4
Evening	9	25.7
Morning	15	42.9
Total	35	100.0

User's Device - Mobile devices exhibit significant diversity in their characteristics (radio chipsets, device resolution, operating systems, screen size etc.). So, in this study the subjects are used their own smartphone device, we selected smartphone users with android based mobile with different technical specification of the device. Then totally 6 mobile device brands, 23 different device models with 12 different types of screen sizes and 5 different types of screen density (DPI) applied in the experiment. The screen display orientation and the average distances between the eye and the display of those who conducted are as per the subject's preference or based on their interest and as their usual video watching experience habit.

Crowdsourcing Applications - Crowdsourcing is an emerging measurement technique which relies on user involvement in the measurement procedure. In the crowdsourcing approach, distributed measurement application agents run on end-user devices and perform active, passive or both types of measurements in real-time. It is a scalable, cost-effective approach that captures the real end-user experience.

For our thesis work, we were selected open-source YoMoApp and Speedcheck Pro crowdsourcing applications. The applications were downloaded by the researcher and sent through Bluetooth to some of the subjects who do not have free data package in order to save their data consumption. The rest of the subjects were downloaded the applications directly from google play store. Finally, all subjects were installed the two applications on their smartphone devices before two days of the experiment date. The YoMoApp application [42] is used to evaluate YouTube video streaming metrics or KQIs by enabling the location service of the device on. The Speedcheck Pro application found in [48] is configured to collect data (download throughput) every five minutes before starting to watch the YouTube video uploaded on my channel, through YoMoApp. So that we have measured download throughput metric of QoS

KQIs using Speedcheck Pro. See appendix B for instructions on the usage of the two crowdsourcing applications.

Questionnaire - Users' survey in addition with perception ballots on hard copy questionnaire and electronic questionnaire on YoMoApp were used to conduct the subjective evaluation. The detail of the subjective evaluation procedure is shown at the end of this chapter in Table 4: 4

Evaluation Process - The objective and subjective QoE metrics of video evaluation had provided with eight video samples for each subject to watch and rate the videos using continuous Absolute Category Rating (ACR) methods. Orientation is important to note that the subjects were carefully given a detail orientation about the subjective video assessment: on how to go through the YoMoApp and paper-based ballot questionnaire while rating the videos and so on. Before the actual evaluation session, subjects were oriented to rate the videos one time based on each of the seven Performance Quality Indicators (PQIs), independently. The Orientation comprised of:

- ➢ The definition & description of the seven PQIs & their rating scales.
- ➢ Description of each video session, application of the tools and questionnaire contents.

After the oral orientation, but before the actual rating sessions the subjects were given two video clips followed by rating on YoMoApp and paper-based ballot questionnaire which are similar to the actual ones to check if they had understood the oral orientation given to them. Lastly subjects rate the sample videos on YoMoApp application using 4 rating questions, then on the paper-based ballot questionnaire using 3 rating questions. Table 4: 4 shows that Processes of performance evaluation (Rating) of video streaming in the experiments.

Table 4: 4 Processes of Evaluation (Rating) of YouTube Video Streaming in the experiments

Watching Clip 1	Rating content quality	Watching Clip 2	Rating content quality	Watching Clip 3 up to clip 7 &Rating	Watching Clip 8	Rating content quality
	Rating video quality		Rating video quality			Rating video quality
	Rating streaming quality		Rating streaming quality			Rating streaming quality
	Overall quality Acceptability		Overall quality Acceptability			Overall quality Acceptability
	Annoyance level- initial delay		Annoyance level- initial delay			Annoyance level for initial delay
	Annoyance level- staling event		Annoyance level-staling event			Annoyance level of staling event
	Annoyance level-stalling event duration		Annoyance level of stalling event duration			Annoyance level of stalling event duration

Generally, our experiment has been designed to produce 8 sample points for each YoMoApp video streaming objective and subjective metrics evaluation, and 6 sample points for each Speedcheck Pro related metrics evaluation. We did this by setting a target of 9 test sites in AA where 3G is fully available. For each user, average data will be calculated from all watched sample videos to produce an overall average for video streaming subjective, and objective metrics of the experimental evaluation. Then Each user evaluation experiment became one sample point by calculating the average value. Finally, from each subject on average 56 ratings from each video have been collected. Table 4: 4 shows the process of video streaming evaluation.

Filtering of Data and Unreliable User Ratings - In order to obtain reliable results despite the lack of control over the anonymous subjects, methods were designed and deployed to counteract cheating and unreliable test subjects. Participants failing to respond real, if there is a big deviation between subjective perception and objective metrics, results were discarded from the results. As a result of the overall application and user level studies, we found that 12.5 % of the participants failed to respond real and filtered as invalid.

Chapter 5

Data Collection Mechanisms

5.1. Introduction

The main objective of this chapter is to explain the planning and data collections mechanisms for evaluation of the quality of experience of video streaming for 3G subscribers in Addis Ababa by considering both subjective and objective measurements using NMS, open-source crowdsourcing applications, survey & perception assessment questionnaire of watched videos.

The upcoming clauses illustrate each mechanism of data collection- Network level, Application level and User level quality evaluation of YouTube video. A task has been created to collect network, application and user-level video streaming data. The data was collected from July 17 up to August 7, 2019, for a period of two weeks. The 3G technologies considered during data collection are HSPA and HSPA+ by taking video streaming metrics.

5.2. Network Level Evaluation Using NetProb3010

Objective mobile video streaming quality metrics of the 3G network have been collected from NMS with the help of NetProbe3010, a vendor dependent tool, which is owned by Huawei. The product name of the tool is SEQ Analyst version: V200R003C00. The NetProbe3010 tool is integrated with sites located in Addis Ababa. The tool has functionalities of monitoring and analysis video streaming quality.

In this case, aggregate summarized video streaming metrics of the 3G network are collected every 60-minutes usage of subscribers in Addis Ababa when users accessed UMTS/3G network and watched video streaming services. The video streaming quality metrics considered for this thesis work of the network side evaluation are video Streaming Start Delay in milliseconds (ms), video Streaming Stall Frequency per minute, and video Streaming Stalled Time Rate percentage. We collected one-month network-level measurement data for the above three video streaming metrics on an hourly average basis.

We only filtered two weeks video streaming metrics for network-level measurement data out of the one month. The purpose of filtering two weeks data was to compare the results with the other alternative measurements, which have been done only for two weeks only, in this thesis work. Then after completion of filtering task, the data has been extracted in excel (Xls) format file. The data have been manipulated using SPSS & microsoft excel. Finally we have calculated hourly average values for network-level objective metrics for each date & time, and the whole two weeks period of July 17 up to August 7, 2019.

5.3. Applications and User Level Experimental Evaluation

The other alternative measurements that have been done in this thesis work are crowdsourcing applications and subjective questionnaire survey. The applications that we used are open-source crowdsourcing applications, YoMoApp and Speedcheck Pro. The subjective survey has been conducted to assess subjects’ perception on the YouTube videos which had been uploaded for the purpose of this reaserarch study. Based on that we have evaluated YouTube video streaming performance on android-based smartphones at nine locations of Addis Ababa in the real-life environment of 3G users in the field. The experimental measurements of applications and questionnaire survey had been took place for the period of July 17 up to August 7, 2019, during the morning, afternoon and evening of the week.

Finally, we were able to collect all the necessary information needed for the evaluation of objective and subjective video quality metrics such as initial delay, No. stalling events, stalling time duration, resolution, download throughput and MOS.

5.3.1. Application Level Evaluation Using YoMoApp

The YoMoApp tool is free and can be downloaded from Google Play. Subjects of video streaming QoE were able to download the tool and make a test based on procedures depicted in Appendix A. YoMoApp (YouTube Monitoring App) [43] is based on an Android WebView browser element, in which the mobile YouTube website is loaded. The video playback on the mobile YouTube website is integrated via an HTML5 video element and uses DASH technology. With the Android app, JavaScript functions are injected to the website. These functions detect the video element in the DOM tree and perform the monitoring of the playback. Therefore, event listeners are added to the HTML5 video element to monitor all changes of the player state, and the height and width of the video element, which corresponds to the video resolution.

In [44] YoMoApp documentation log files are periodically captured every second, the current playback time and buffered playtime are polled. Additionally, the YouTube ID, video title, video duration, and statistics about video streaming metrics are collected. The data is sent to the Android app for postprocessing and logging. Due to inconsistencies and errors, such as missing or incorrect values, which may arise from the usage of JavaScript, the postprocessing is required to ensure high consistency of the resulting streaming logs. YoMoApp also monitors network and context parameters. The network usage, i.e., the total amount of uploaded and downloaded data, is logged periodically for both mobile and Wi-Fi networks. The app also logs changes of the operator, cell ID, signal strength, RAT, and GPS position.

Moreover, it retrieves several device characteristics and monitors their changes. These include screen size, screen orientation, player size, player mode (normal/full screen), and volume. The monitored data are stored in separate log files for each video session. If the video has ended or was aborted after a minimum session length of the 20s. The user is asked 4 different question to rate the QoE of the streaming session on a MOS scale ranging from 1 (bad) to 5 (excellent) however, users have an option that can also close the rating dialogue. For this thesis work, the user is required to submit a rating and the time duration of each videos are more 20s. All log files are locally cached on the device and they are transmitted to an external database when the app is closed, at fixed time intervals, or if triggered manually by the user. Sample of the YoMoApp user interface and rating scale are included in Appendix B.

5.3.2. Application Level Evaluation Using Speedcheck Pro

The other tool applied for data collection is a crowdsourcing tool called Speedcheck Pro which measures the download throughput of the users' device while the users are watching videos. The tool is free and can be downloaded from Google Play. Subjects, who are involved in the video streaming QoE evaluation, had downloaded the application tool and made a test based on procedures depicted in Appendix A by locking their mobile devices to 3G only.

The owner of the tool is Etrality GmbH, which is legally registered and operating in Germany. The description of the tool is available in Speedcheck.org official website [45] and the metric for this specific tool evaluation is downlink throughput. We selected Speedcheck Pro as one part of methodologies, from other speed test application available freely, is because it has capability feature to Schedule automatic checks to monitor the connection over time. For example, if you are planning to test or evaluate video streaming metrics around a certain time of day and time you can schedule a speed check to run multiple tests during this particular time window. The technical setup for the Speedcheck Pro is illustrated in Appendix A and a sample of the results collected from the participants are included in Appendix B.

This tool would help end-users better understand the performance they are getting from their network access and how it relates to their expected QoE. On the other side, it would allow network operators and content providers to troubleshoot access network performance issues, and better design their systems so that Quality of Experience of their customers is optimized.

For our experimental evaluation around 100 3G users were requested in person in the months of June and July 2019. Then only 40 subjects were willing and involved in the experiment. Finally, based on procedures in Appendix A, the subjects had sent details of the test with the raw data from Speedcheck Pro directly to the researcher gmail address in excel (Xls) format. The log files are filtered and preprocessed. Invalid logs were removed and logs without speed test results were sorted out. After the filtering of all video sessions of users, 240 sessions of speed test have been taken for analysis purpose. Table 5: 1 below shows a summary of collected fields from Speedcheck Pro data within their descriptions.

Table 5: 1 Speedcheck Pro Field Data

S.No	Field Name	Description
1	Test Type	This field help us to identify whether the test is repeating (scheduled) or manual
2	Date	This field helps us to identify the date in which the test is conducted.
3	Time	This field helps us to identify the time in which the test is conducted. We can extract day and hour of measurements conducted from this field by converting the time into the ethiopian time zone. The time zone had three hours difference during the test period.
4	Ping [ms]	This field helps us to extract the international latency in ms
5	Download [Mbps]	This filed shows the result of the downlink throughput in Megabits Per Second (Mbps)
6	Connection Sub	We choose this field to know the technology of 3G network selected by the user during the test period. For example, HSPA and HSPA+ technologies are identified using this field.
7	Latitude	This field shows the location of each test in a coordinate system
8	Longitude	This field shows the location of each test in a coordinate system.

5.3.3. User Level Subjective Evaluation Using Questionnaires

The aim of identifying subjective metrics is to correlate the objective metrics results with the perception of the subjects in the experiment about the services being provided. To evaluate the perception of the subjects, a questionnaire which consists of subjective questions or metrics has been prepared. The questionnaire had two parts that are pre-watching and post-watching. The evaluation experiment of video streaming quality and annoyance-related metrics assessment is based on ACR (Absolute Category Rating) of 5-point Likert Scale which is described in Table 5: 2 below.

Table 5: 2 Rating Scale for Quality and Annoyance Level

Quality	Perception of Annoyance Level on streaming Impairment	Rating scale
Excellent	Imperceptible	5
Good	Noticeable, not annoying	4
Fair	Noticeable, Slightly annoying	3
Poor	Noticeable, annoying	2
Bad	Perceptible or Very Annoying	1

Pre-watching Questionnaire: - The pre-watching questionnaire consists of 5 questions about demography of the subjects, 3 questions about their device which were used in the experiment, 1 question about their video streaming habit, 1 question for the location of the subjects where the experiment is done, 1 question for the date and time of the experiment and 1 question related to the mobile connectivity during the experiment. Totally 12 questions response was collected from survey assessment of 35 subjects. Table 4: 7 below shows that questions in pre-watching hard copy questionnaire.

Table 5: 3 Contents of Pre-Watching Hard-Copy Questionnaire

QUESTIONS No	Hard copy Questionnaire Contents
Q01	The environment context
Q02	Age group
Q03	Educational background
Q04	Job status
Q05	Gender
Q06	Access network connectivity
Q07	Video habit
Q08	Smart phone Brand
Q09	Smart phone model
Q10	Android version
Q11	The name of Location (Place) of experiment
Q12	Date of Experiment

Post Watching Questionnaire: - Totally 7 questions were asked to each subject in relation with video streaming quality and impairment metrics, the 4 questions were in YoMoApp and 3 questions in hard copy questionnaire. A summary of the subjective metrics questions is displayed in Table 5: 4 & 5: 5. Each subjective metric is given identification starting from Q01 to Q15. Appendix C displays details of the subjective metrics of hard copy questionnaire. Totally 3 quality rating and 1 acceptability questions on YoMoApp application questionnaire, and 3 viewing impairment rating questions on hard copy questionnaire so totally 7 questions were prepared for each sample videos for 35 subjects.

We used single stimulus ACR rating scale for the six perception assessment questions on YoMoApp and hard copy questionnaires. Additionally, we used scalar type questiones (Yes or No) for acceptability rating on YoMoApp then the result was converted to an integer value, where a 1 corresponds to acceptable, a -1 corresponds to inacceptable. Finally, the three YoMoApp quality rating, excluding the acceptability question, based on the eight sample videos on YoMoApp were collected and analyzed to calculate MOS value. Figure 5: 4 below shows the contents of the post watching questions on YoMoApp.

Table 5: 4 Contents of Post-Watching on YoMoApp Application Questionnaire

QUESTIONS No	YoMoApp Rating questions
Q01	Perception of Content quality
Q02	Perception of Video quality
Q03	Perception of Streaming quality
Q04	Acceptability of video quality

In addition to that, to identify and differentiate each streaming quality metric which are not listed on YoMoApp, 3 viewing impairment rating questions had prepared on hard copy questionnaire. The rating results were collected and analyzed to evaluate the annoyance level of the users on the initial delay, stalling events and stalling time duration of each event while, users were watching the sample videos. Figure 5: 5 below displays the contents of the post watching questions on the hard copy questionnaire.

Table 5: 5 Contents of Post-Watching Hard Copy Questionnaire

QUESTIONS No	Hard copy Questionnaire Contents
Q01	Perception of annoyance level of initial delay
Q02	Perception of annoyance level of frequency of interruptions(stalling)
Q03	Perception of annoyance level of duration time of interruptions (stalling)

Chapter 6

Analysis, Results and Interpretations

6.1. Introduction

As stated in Chapter 4, subsection 4.3 data has been collected using four mechanisms NetProb3010 (NMS), YouTube Monitoring Application (YoMoApp), Speedcheck Pro application and users survey & perception assessment questionnaire. The data analysis will be performed by using statistical analysis to evaluate both the objective and subjective data that have been collected from the above three sources. Therefore, network, application and user-perceived parameters data have been evaluated and analyzed statistically. Moreover, correlation analysis has been done between subjective and objective metrics results. The analysis of the results obtained from the above scenario evaluations and their interpretations are illustrated in Subsections of this chapter.

6.2. Network Level Evaluation of NetProb3010

Network level video streaming QoS data from NetProb3010 was collected from July 17 through August 7, 2019, for two weeks. We selected this period in order to make the dates similar with the period of measurement using application and user level experiment. So that we can make comparative analysis easily between the three measurements methods.

The metrics that we had considered for the evaluation were an initial delay in millisecond, stalled frequency per minute and stalled time rate percentage, which are the hourly average for all subscribers in Addis Ababa when they access video streaming services in 3G network. The extracted metrics data consists of the hourly average values for each date. The raw data was extracted in excel file format and analysis was done using Microsoft Excel and SPSS. The hourly metrics data were analyzed based on statistical methods. Moreover, the hourly average values of the metrics had been taken as an input to evaluate the service based on the parts of the day.

Table 6: 1 below shows that th minimum, maximum and mean values of network level measurement results of video streaming metrics for initial delay, number of stalling events and stalled time percentage of subjects in the experiment. Based on the experiment results for

average value for initial delay, number of stalling events and stalled time percentage had been 19.06 Sec, 0.38 times per minute and 6.46% respectively.

Table 6: 1 NetProb3010 - Network Level Measurement Results of Video Streaming Metrics

Metrics	No. of Users	Minimum	Maximum	Mean	Std. Deviation
Initial delay	35	12.50	29.50	19.06	4.02
No. of Stalling events	35	0.01	0.80	0.38	0.21
Stalled time Rate Percentage	35	0.15	13.58	6.46	3.26

Figure 6: 1 below shows that the initial delay experienced by the 3G users for the evaluation period. Based on our data analysis, 48% of the users experienced an initial delay between 15.74 and 19.87 seconds. In addition to that, Twenty-six percent of users were experienced an initial delay between 12.5 and 14.84 and the rest 26% of the users experienced between 20.29 and 29.50.

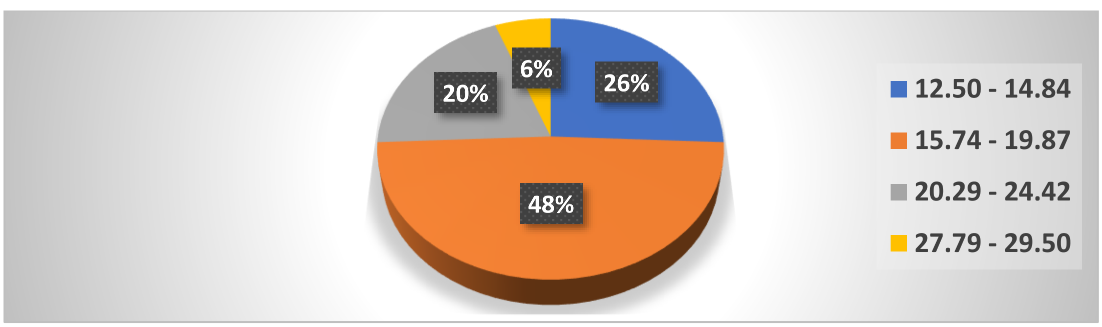


Figure 6: 1 Evaluation Results of Initial Delay Metrics in Seconds

Figure 6: 2 below displays that the number of stalling events that users were experianced during the experiment date. Based on our data analysis 86% of the dataset or users experienced stalling events between 0.01 and 0.45 and for the rest 14% of the dataset the number of stalling events between 0.56 and 0.80. The result shows that all the 3G users has experienced stalling event while watching video streaming services during the evaluation period.

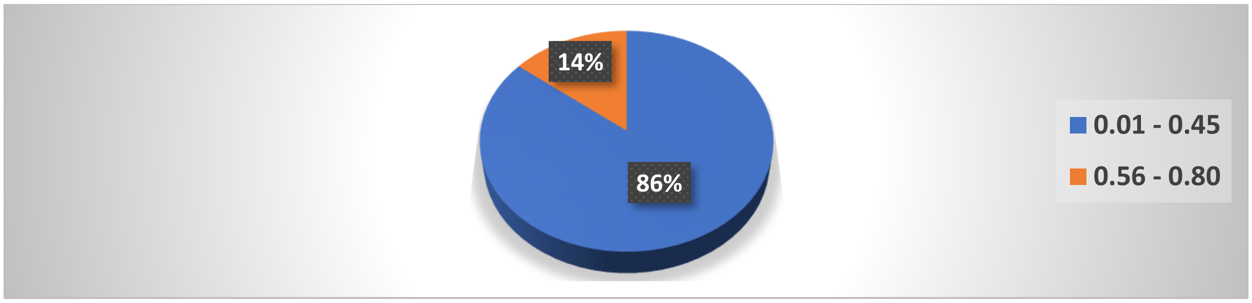


Figure 6: 2 Evaluation Results of Number of Stalling Events per Minute

Figure 6: 3 below shows that the percentage of the hourly stalled time rate observed in NetProb3010. The stalled time rate percentage had been experienced between 0.15 and 4.66 by 83% of the 3G users. The rest 17% of the 3G users had experienced stalled time rate between 6.99 and 13.58. The result shows that users who watched video streaming services during the period of evaluation had been added an average additional time of minimum 0.15% and maximum 13.58%-time duration of the original video duration due to stalling or buffering.

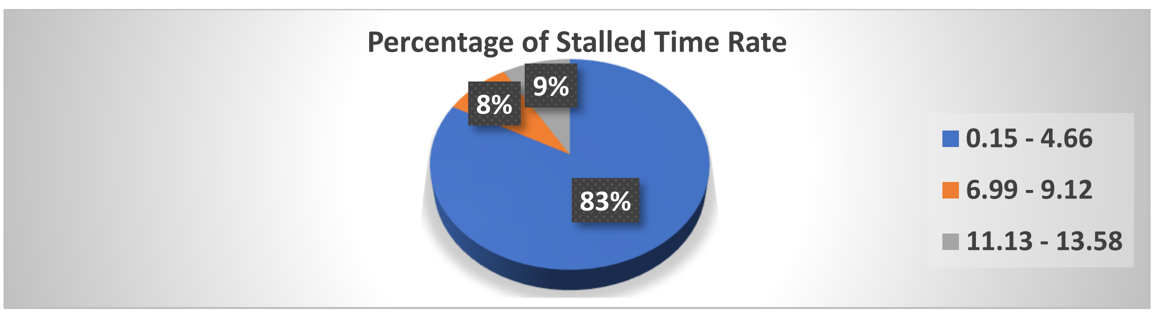


Figure 6: 3 Evaluation Results of Stalled Time Rate Percentage

In order to see further analysis, we have also evaluated the average values of the data based on part of the day and overall evaluation period. Table 6: 2 below shows the summarized average results, and initial delay, stalled frequency and stalled time rate were 19069 ms (19 Sec), 0.38 times per minute and 6.46% stalled time rate respectively. Based on our analysis for network-level evaluation, we observed that initial delay was worst during the nighttime. Furthermore, stalling frequency & stalled time rate percentage were worst in the morning & afternoon parts of the date.

Table 6: 2 NetPro3010 Video Streaming KQIs Evaluation Result Summary

Evaluation Period	Parts of The Day	Video Streaming Start Delay(ms)	Video Streaming Stall Frequency(times/min)	Video Streaming Stalled Time Rate (%)
July 17,2019 - August 7,2019	Morning Average (MA)	19166	0.53	9.83
	Afternoon Average (AA)	18527	0.41	6.17
	Evening Average (EA)	19516	0.21	3.38
	Experiment Date Average (MA+AA+EA)/3	19069	0.38	6.46

6.3. Application and User Level Experimental Measurement Results

There were 40 participants who are found in all age categories have participated in this QoE of video streaming evaluation experiment in the field study. Most participants finished watching the 8 video clips. The field study of the YoMoApp application resulted in 320 streaming sessions test results and speedcheck Pro resulted in 240 network throughput tests were generated by 40 different users during the evaluation period.

The participants were distributed in nine locations of Addis Ababa, Addis Ababa Commercial college: 5.7%, Eyore building (Tewodros street): 5.7%, Ethio telecom HQ: 22.9%, Karakore around Ayertena: 5.7%, Lancha: 8.6%, Meskel flower: 22.9 %, Riche: 5.7 %, TPO (near Licey Gebremariam school): 8.6% and Yode Restaurant(Olympia behind Mega building): 11.4%.

All participants were used their own smartphone and the 3G cellular access network to stream videos using YoMoApp. 62.9% of participants had been used Android versions 5.0 & 6.0, followed by Android 7.0 & 8.0 with 38.1%. In terms of the display size, the participants used 12 different types. 57 % of users had a resolution of 480 x 800, 480 x 854, 540x960 and 720x1280 pixels at 240 (dot per inch) dpi, followed by 22.5% with a significantly higher resolution of 1080 x 1920, 1080 x 2076, 1080 x 2094 and 1080 x 2131, at 420 and 480 dpi. The rest 20.5 % were used 720 x 1184, 720 x 1192, 720 x 1280, 720 x 1384, 720 x 1396 pixels at 280 and 320 dpi.

After the evaluation, the log files were filtered and preprocessed. Invalid logs were removed and logs without video statistics were sorted out. Finally, after the filtering, 35 valid subjects with 273 sessions were used for evaluation purposes in the results section. The analysis of the field study participants includes device characteristics and the technical characteristics of the objective and subjective metrics of video streaming sessions (KQIs) were considered.

6.3.1. YouTube Monitoring Application (YoMoApp) Results

YoMoApp data was collected from July 17 through August 7, 2019, for two weeks in the field. The evaluation metrics considered were initial delay, a number of stalling events, stalling time duration, video quality watched, a number of video quality switches etc. We calculated average value for each user during the subjects are watching the YouTube video streaming using the 3G network.

The results are analyzed averagely per each subject for all sample video sessions based on the geographical location of the subscribers. The extracted data consists of average values for each unique location. The raw data for each subject were extracted in CSV file format from YoMoApp log portal and exported and converted to XLS format using Excel and then the analysis was done using Microsoft Excel and SPSS tools.

All subjects were experienced an initial delay when they were watching all sample video sessions. Around 25.7% of the sessions had a smooth playback without stalling event, and the rest 74.3% experienced stalling events. More than 63% of the streaming sessions showed an initial delay greater than 10s,

Table 6: 3 depicts YoMoApp evaluation results of KQIs (metrics) for 273 mobile YouTube streaming sessions. Based on the experiment, the average length of the initial delay was 17.32s, with a maximum of 97.69s. the average total stalling time was 3.79 seconds and the maximum total stalling time duration was 88.15s. The maximum number of monitored stalling events were 2.25, while the average number of stalling events was 0.47.

Table 6: 3 YoMoApp Evaluation Results of Video Streaming Metrics

YoMoApp KQIs	No. of users	Minimum	Maximum	Mean	Std. Deviation
Average Initial delay	35	1.96	97.69	17.32	21.25
Average Video Resolution Watched	35	144.00	360.00	229.8	78.07
Average Stalling events	35	.00	2.25	.47	.60
Average stalling time duration	35	.00	88.15	6.04	10.07

Figure 6: 4 below shows the average overall sample videos resolution watched by the subjects in the experimental evaluation. Based on our findings that 55% of the sample video streaming sessions were watched with resolution 144p, followed by 36% watched with resolution 360p and the rest 9% watched with resolution 240p. Further, we observed the lowest start up resolution quality which supports the conservative behaviour of YouTube video resolution as mainly 144p is streamed from the start in case of poor network condition.

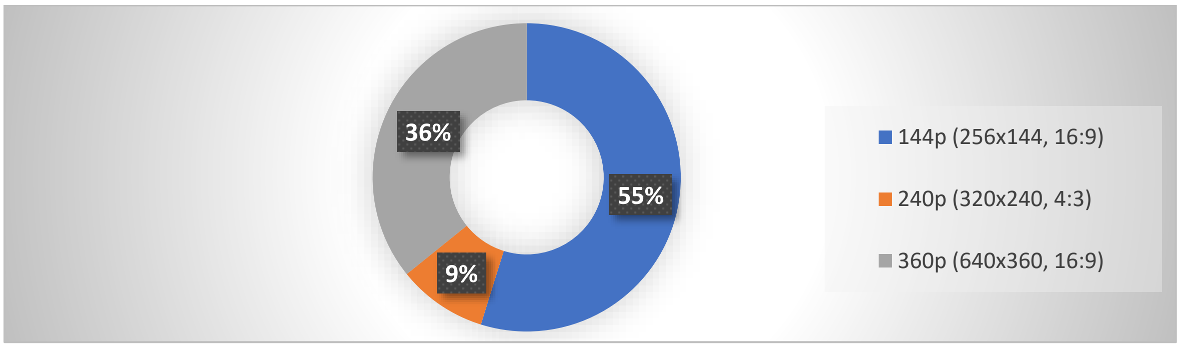


Figure 6: 4 YoMoApp result of Average Video Resolution Watched by Subjects

The results in Figure 6: 5 shows that the maximum average initial delay of 36.26 seconds were observed in ethio telecom HQ and minimum average initial delay of 3.3 seconds were detected in Yode Restaurant around Olympia. The rest seven locations where subjects were experienced average initial delay time of almost more than 10 seconds during the period of evaluation.

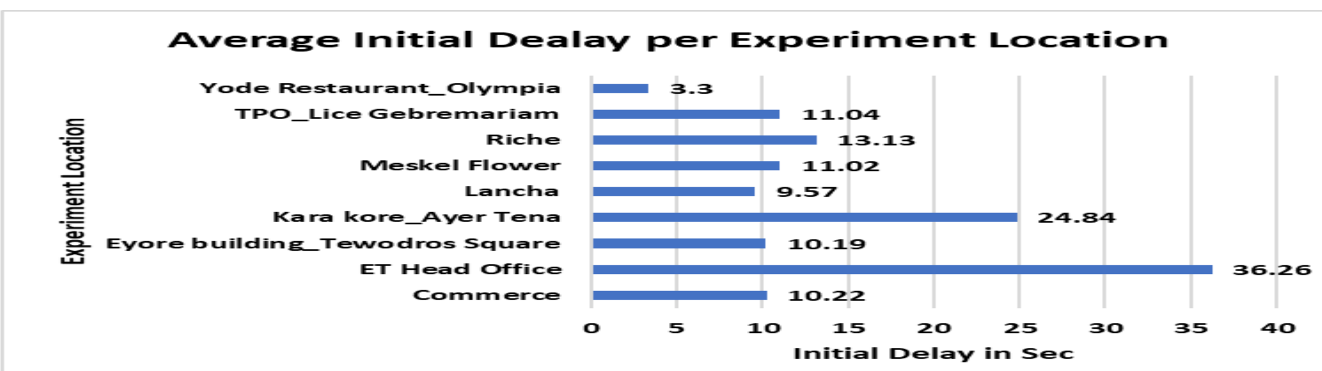


Figure 6: 5 Evaluation Results of average Initial Delay Metric per Experiment Locations

From Figure 6: 6 plots below, a few observations can be made. The average number of stalling events was worst with 1.27 times per 30 seconds at Riche. The best average number of stalling events results were found at Karakore around Ayertena and Yode restaurant around Olympia.

The worst average stalling time durations were observed at ethio telecom HQ, Commercial College of Addis Ababa university. Further more the best stalling time duration was observed at Yode restaurant around Olympia. Therefore, the result found from this application ensures the average number of stalling events and stalling time duration were not good.

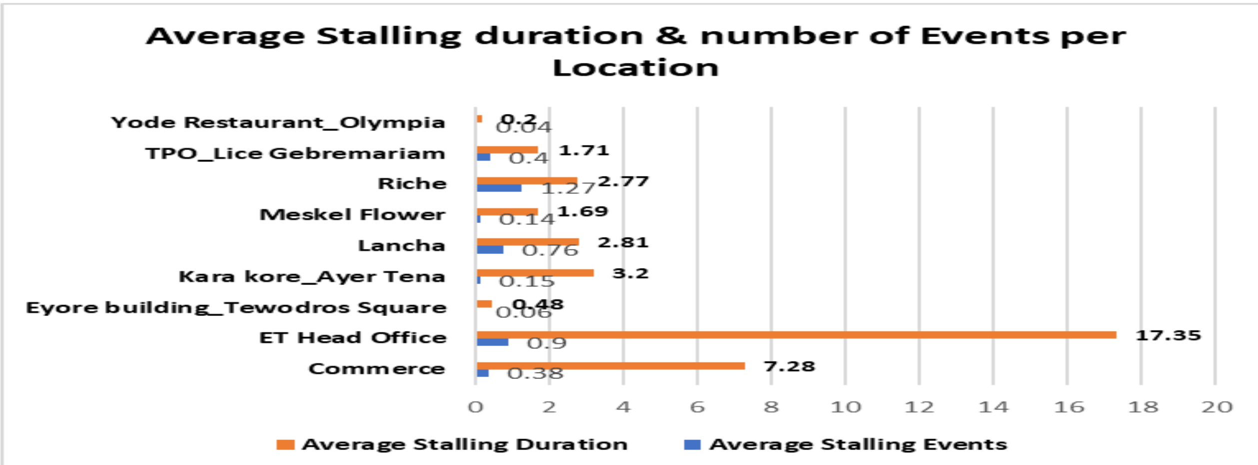


Figure 6: 6 Evaluation Results of Average Stalling Events and Quality Switches Metrics per Experiment Locations

Figure 6: 7 shows that the maximum or relatively best average video resolution watched at Yode Restaurant around Olympia with resolution 360P and the worst average video resolution watched at commerce with resolution 158P. The result shows that no subjects didn't watch videos with SD and HD video quality standards during the experiment. Even though the original videos were uploaded with 720p HD quality and subjects's device average download throughput was more than 1.78 Mbps, the subjects were watched poor video resolution qualities during the experiment.

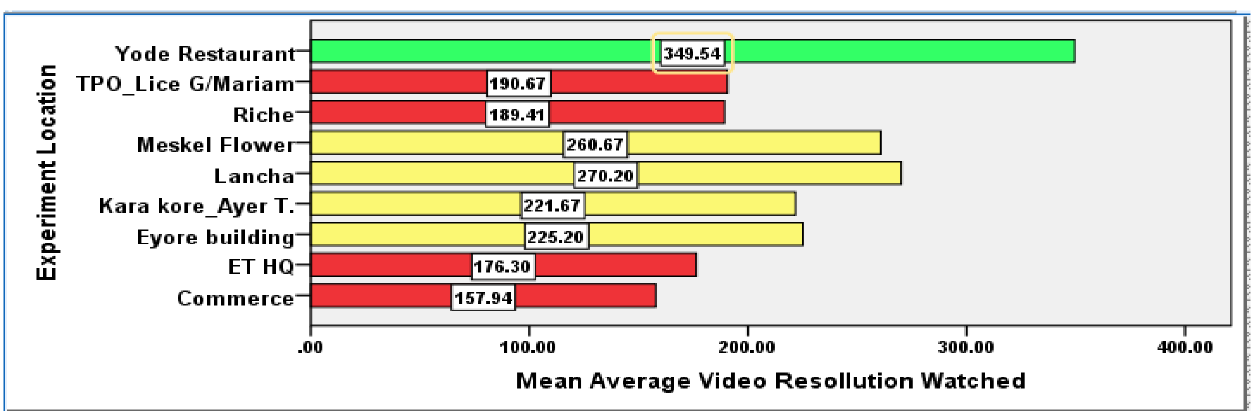


Figure 6: 7 Result of Average Resolution Watched per Location

Figure 6: 8 shows that the highest average video resolution of 235.5p was watched in the outdoor environment, whereas the lowest average video resolution 224.9p was watched at the indoor environment.

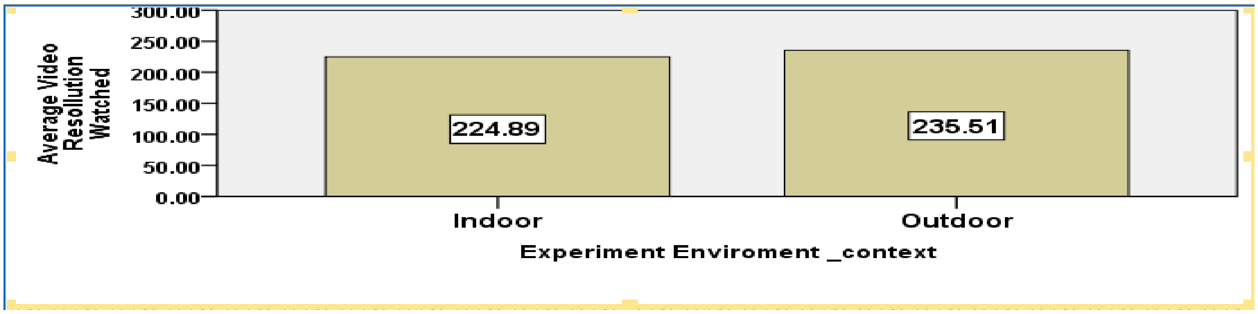


Figure 6: 8 Result of Average Resolution Watched per Environment Context

6.3.2. Application Level - Speedcheck Pro

The speed test was conducted using Speedcheck Pro tool simultaneously while subjects were watching YouTube sample video streaming sessions on YoMoApp in their smartphone devices. Here our focus was evaluating the real figure of the customers' download throughput while they were watching YouTube video streaming through YoMoApp application during different parts of the day and at different experimental locations.

The evaluation was performed on the stationary place of the indoor and outdoor environment in the morning, afternoon and evening or night. We preferred the stationary places for evaluation of the experiment as most customers are expected to be in a stationary place when they watch video streaming services. It was also mentioned in [46] that most of the mobile traffic generated is from indoor. 70% - 80% of the traffic generated is inside buildings and in some cities, 50% of the traffic generated is from 10% of the buildings.

As we stated before, the identified locations of the experiment were central parts of AA with public areas which accommodate a high number of users in their day to day activities. Figure 6: 9 below shows the geographical locations of the experiment which was extracted based on latitude and longitude information found from the speedcheck Pro application

Figure 6: 9 Geographical Locations of the Experiments

As stated in the above speed test has been done in selected nine sample locations in a different time & parts of the day. Table 6: 4 below demonstrates a summary for the outcome of Speedcheck Pro that carried out for the evaluation of download throughput while users were watching sample video sessions. The evaluation considered in this work was downlink throughput. The result of the download throughput measurement for minimum, maximum and mean values were 0.20, 9.61 and 1.78 Mbps respectively.

Table 6: 4 Result Summary for All Subjects' Average Download Throughput during the Specific Period of Video Watching

Metric	No. of Users	Minimum	Maximum	Mean	Std. Deviation
Average Download in Mbps	35	.20	9.61	1.78	2.41

Figure 6: 10 below shows a result of the average download throughput for each location. A maximum average of 7.03 Mbps download speed was achieved at Yode Restaurant around Olympia, 3.22Mbps was achieved at TPO around Lice G/Mariam School. The minimum average download throughput of 0.46 Kbps was achieved at Eyore building around Tewodros square. Moreover, in the other 4 locations the average download throughput measured were between 1 and 1.31Mbps, and for the rest two locations average download throughput were less than

1Mbps. To summarize 77.8% of the experimental evaluation sites where users were watching sample videos achieved download throughput below 1.31Mbps. The figure below reveals average download throughputs were not uniform for all the locations because of the nature of the users' devices and the spatiotemporal distribution of the tests.

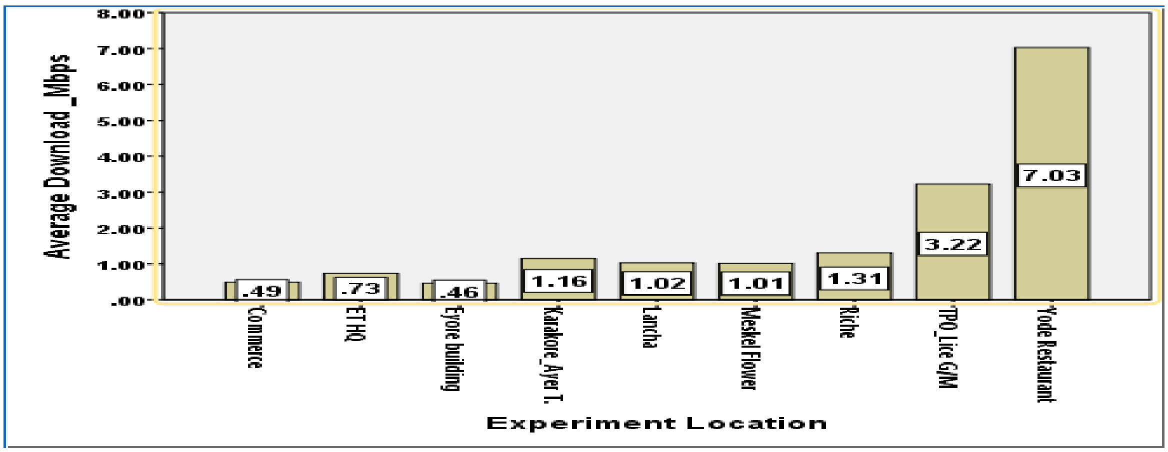


Figure 6: 10 Statistics of Average Download Throughput per Location

Figure 6: 11 below shows that 57% of subject's smartphone download throughput was less than 1Mbps that were between 0.2 and 0.99 Mbps and 11.4% of the subjects were experienced download throughput between 1.02 and 2.05 during watching the sample video sessions. The remaining 31.5% of the subjects were experienced download throughput between 2.06 and 3.73Mbps, between 6.93 and 9.61Mbps. To summarize the download throughput achieved or measured less than 1Mbits/s for more than fifty percent of the video streaming evaluation session. Moreover, 22.9% of evaluation sessions the download throughput measured had been between 1.02 and 2.05Mbits/s. Thus, this poor performance of download throught acheived might have its own impact on the quality of videos perceived by the user.

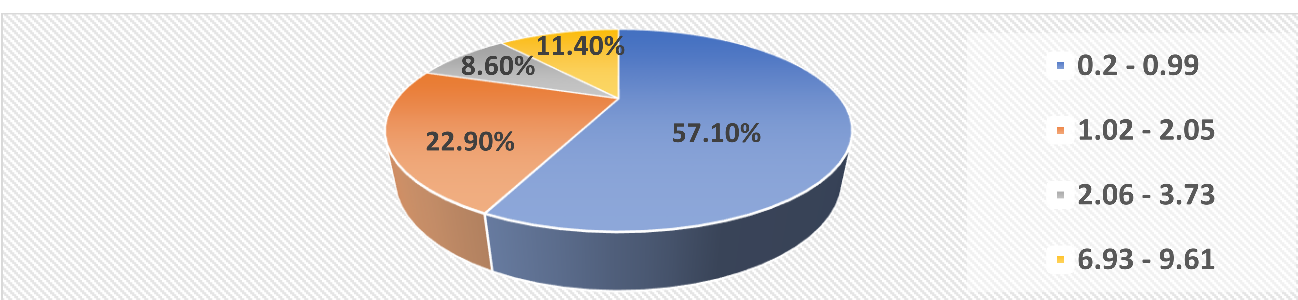


Figure 6: 11 Subjects Smartphone's Average Download Throughput in Mbits/s during the Evaluation Sessions of Video Streaming Experiment

Figure 6: 12 below illustrates that the average download throughput of subjects while watching the sample videos per experiment environment context. So, the best average download throughput during the evaluation session was measured in an outdoor environment which achieved 2.67Mbps. For the indoor environment the average download throughput 1.03Mbps. The maximum average download throughput was 9.61Mbps for outdoor environment context and the minimum download throughput was 0.20Mbps for the indoor environment. In [46] the authors are stated that poor performance of download throughput in the indoor environment could be because of the building wall attenuation. Therefore the level of electromagnetic fields will be affected strongly.

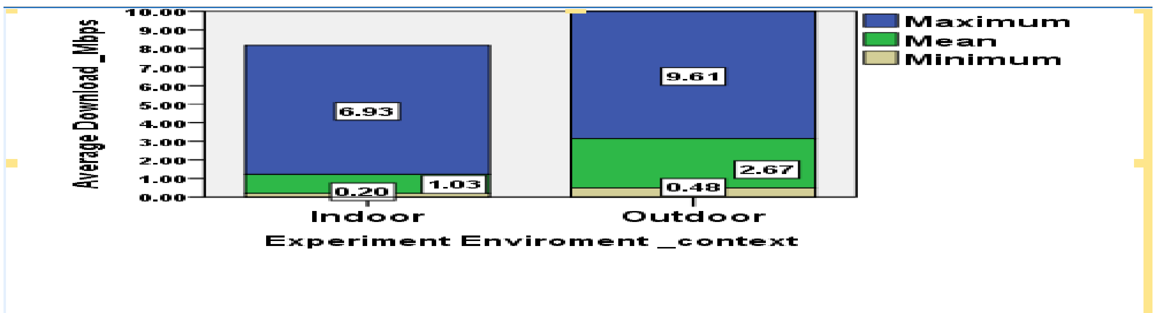


Figure 6: 12 Average Download Throughput per the Experiment Environment Context

As illustrated or shown in Figure 6: 8 before and Figure 6: 12 above, the subjects achieved the best average download throughput and watched the best video resolution quality at the location of the outdoor environment. In contrast, subjects were achieved poor average download throughput and they watched poor video resolution quality at an indoor environment. We can conclude that poor Download throughput is one of the factors that can influence on the QoE of users. So high attention must be given to indoor coverage, as most traffic generated at the indoor environment, and users can watch smooth video streaming services.

Figure 6: 13 below shows that average download throughput achieved per parts of the day. So subjects who watched the videos achieved relatively the best download throughput measurement results in the afternoon and evening parts of the day with 2.76 and 2.30Mbps respectively. Subjects who watched videos in the morning part of the day achieved relatively poor measurement results of 0.66Mbps. This measurement result might be because of most users were focusing with their work and didn’t access mobile data in the afternoon. So that users who access

mobile data can get best average download throughput during this off-peak time. However, the average download throughput is reduced in the evening and morning parts of the experiment period. The result shows that most users of 3G network access data in the evening and morning which are a busy hour for the 3G network in ethio telecom context.

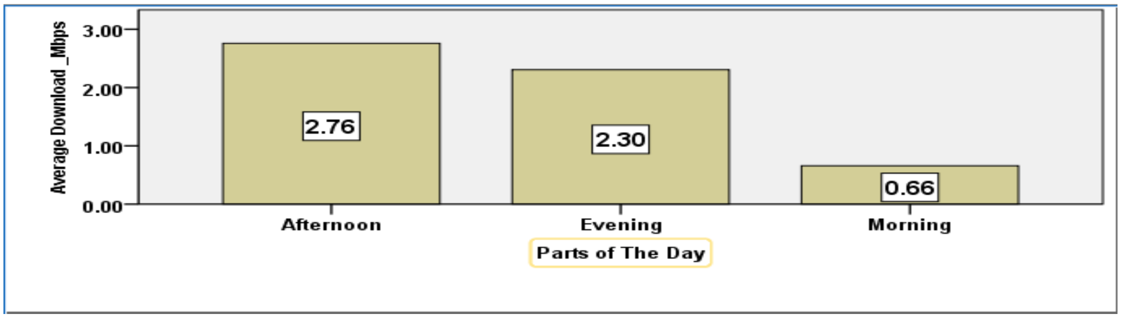


Figure 6: 13 Average Download Throughput per Parts of the Experiment Day

As a part of solutions to provide better customer experience on video streaming services ethio telecom should optimize and improve network coverage through complementary public wi-fi hotspot coverage by providing high-speed access at selected indoor locations such as buildings based on comprehensive survey assessment.

6.3.3. User Level Subjective Metrics

Subjective metrics are a result of responses from subjects who watched the sample videos based on video streaming metrics explained in chapter four of this thesis work. As described in Chapter 4, a video streaming evaluation consisting of 11 survey questions, 3 perceptual annoyance level impairment rating on the paper-based questionnaire and 4 perceptual quality rating on YoMoApp application conducted at nine different locations in Addis Ababa.

The rating from YoMoApp application and paper-based ballot questionnaire ratings are combined to calculate the overall MOS value of the evaluation of video streaming experiment. For analysis purpose, the QoE evaluation of video streaming experiment has been analyzed with respect to their MOS value, based on three categories as follows:

- Quality category metrics
- Spatiotemporal metrics
- Other factors

6.3.3.1. Quality Category Metrics

Perception questions of post watching of video streaming metrics on YoMoApp and paper-based questionnaire are categorized as quality-related metrics. The results of quality-related metrics collected based on 5-point likert scale were shown in chapter 5.

Figure 6: 14 below shows that 82.85% of subjects were achieved (rated) MOS value below 3.5 and the rest 18.15 % of subjects were achieved MOS value between 3.5 and 4.33. The standard deviation was 0.549. This shows that most users were not satisfied with the current video streaming quality.

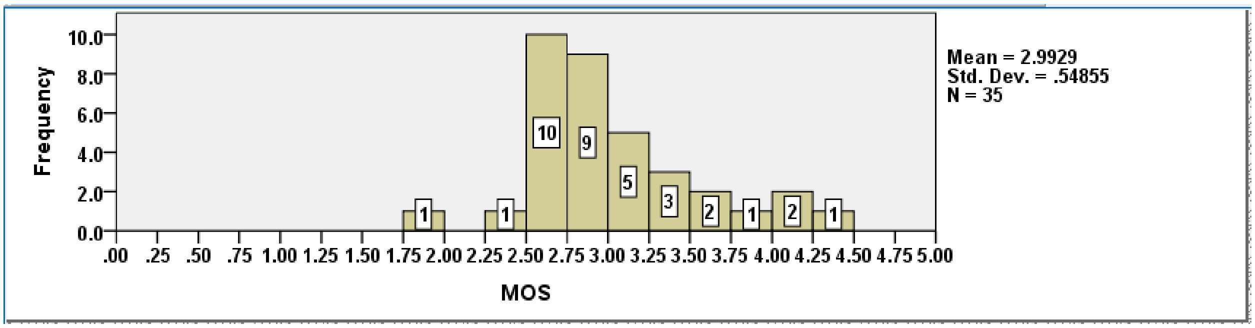


Figure 6: 14 Average MOS per frequency of subjects

Table 6: 5 below shows the summary of MOS results on the data analysis. The overall MOS result consists of three different types of quality that are content quality, video or resolution quality and streaming quality. Based on the subject's perception result of content has got the highest score 3.15, the video or resolution of the videos were rated 2.57, and streaming quality was rated with 2.66. The overall MOS result for the experiment period was 2.79. The finding of this study shows that video or resolution quality metric had impacted users' QoE, more than content and streaming quality.

Table 6: 5 Average Content, Video and Streaming Quality and overall MOS Result

Quality	Rating
Average Content	3.15
Average Video or resolution	2.57
Average Streaming	2.66
Overall MOS	2.79

As stated in [47], the average MOS value must be greater than 3.5 for better customer satisfaction. However, the results that was shown in Table 6: 5 above the overall MOS value for evaluated quality-related metrics was 2.79 which shows that subjects were not satisfied with the current service of video streaming quality provided by ethio telecom.

Table 6: 6 below shows that 37.1% of the subjects, who were involved in the experiment, accepted the video streaming quality. 48.6% of the subjects were rejected the video streaming quality as unacceptable and 14.3% of the subjects were rated equally, as acceptable and unacceptable, for the number of sample videos they were watched. The formula that we used to calculate the average acceptability rating for each subject is written below.

$$\text{Average Acceptability Rating of Subjects} = \frac{\text{Summation Of Acceptability Rating of each Video}}{\text{Total Number of Watched Videos}}$$

Table 6: 6 Summary of Average Acceptability Rating for Videos Quality

Average Acceptability Rating of Subjects	**No. of Subjects**	**Percentage**
Accepted (Yes or 1)	13	37.1%
Not Accepted (No or -1)	17	48.6%
Equal result for accepted and not accepted (offset or 0)	5	14.3%

Figure 6: 15 below shows that the average acceptability rating of subjects based on experiment locations. The result depicts that subjects who were located at Yode restaurant around Olympia, TPO around Licey Gebre Mariam school, Eyore building around Tewodros square and Kara kore around Ayertena accepted the overall average video streaming session quality they had watched. Whereas subjects who were located at Riche, Meskel flower, Lancha, ethio telecom HQ and AA commercial college didn't accept the overall average video streaming session quality they had watched. The formula that we used to calculate the average acceptability rating per location is written below.

$$\text{Average Acceptability rate per location} = \frac{\text{Summation of Average Acceptability rating of subjects}}{\text{Total Number of Experiments in each Location}}$$

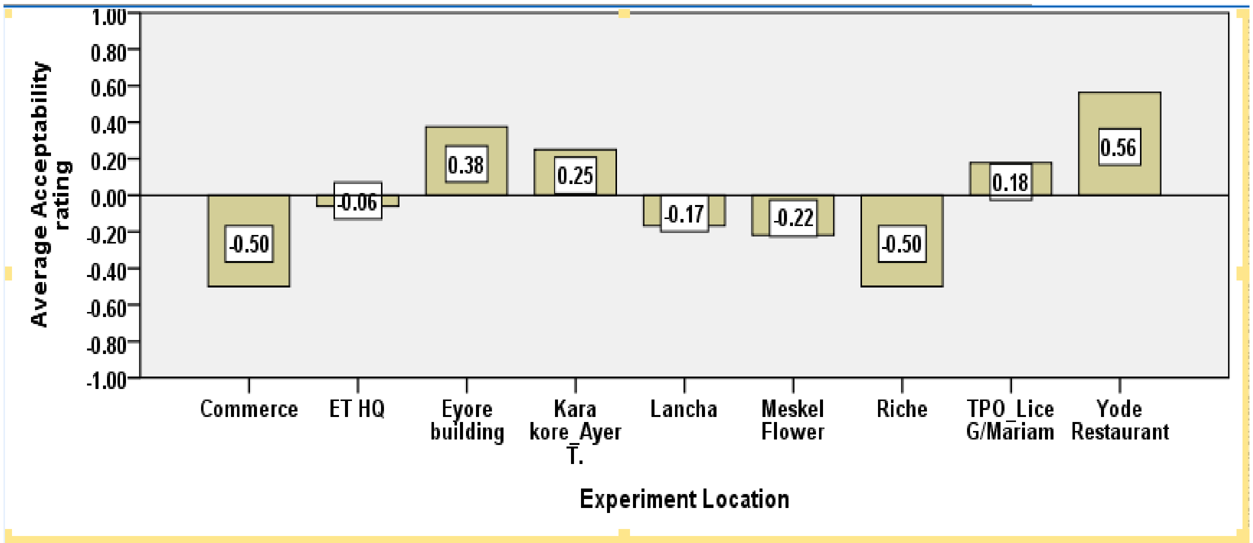


Figure 6: 15 Result of Average Acceptability per Experiment Location

From the analysis results found from this experimental evaluation study, we can conclude that most users had not accepted the video streaming quality provided by ethio telecom at the specified most locations.

Table 6: 7 below depicts that subjects' perception of annoyance level on video streaming quality that had been watched by the subjects. The result shows the overall average annoyance level of initial delay, a number of stalling events and stalling time duration have been achieved with 2.78, 4.54 and 4.26 respectively. Based on the results we can conclude that the subjects were annoyed more on initial delay metric of the sample videos they watched. The result shows that initial delay had an impact more on the users' perception of the video streaming quality with an overall average annoyance level value of 2.78.

Table 6: 7 Subjects Perception Rating Result on Annoyance Level of Video Streaming Metrics

Average Annoyance level Perception	No. of users	Minimum	Maximum	Mean	Std. Deviation
Annoyance level on Initial delay	35	1	4.25	2.78	0.88
Annoyance level of number of stalling events	35	2.75	5	4.54	0.58
Annoyance level of stalling time duration	35	1	5	4.26	0.84

Figure 6: 16 below illustrates 16 users who had watched average video resolution between 144P and 196P, and they achieved the lowest average MOS of 2.42. 11 users who had watched

average video resolution between 211P and 283P, and they achieved MOS of 2.72, and the rest 8 users had watched average video resolution between 302p and 360p, and they achieved MOS value of 3.63. The result shows that subjects who watched the lowest resolution had the smallest MOS value, and subjects who watched higher resolution achieved a higher MOS.

From this result, we can conclude that when average video resolution watched by the user increases, their quality of experience has also raised significantly. This confirms that when users watched videos with the lower resolution has a negative impact on users' perception.

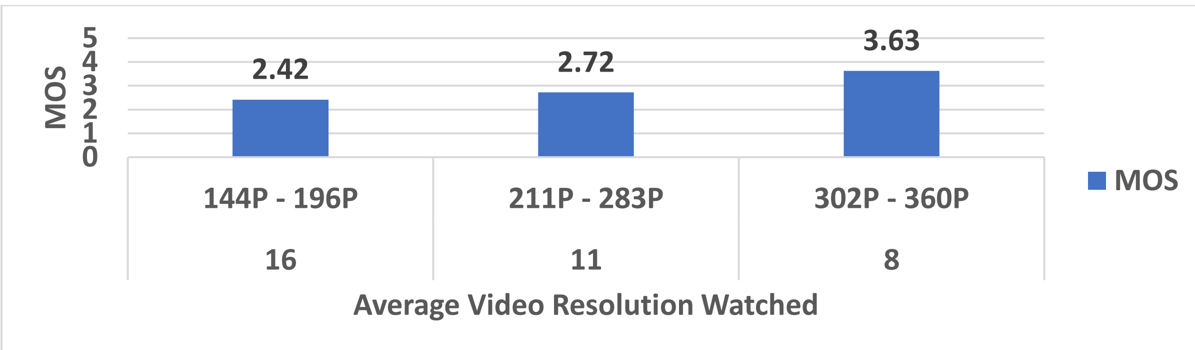


Figure 6: 16 Impact of Average Video Resolution Watched on QoE

6.3.3.2. Spatiotemporal Metrics

Spatiotemporal metrics are included in this study to see where and when 3G customers perceive the worst and best quality of video streaming service in the mobile 3G network. From the analysis results found from this experimental evaluation study, 88.8% of the sample locations (8 locations out of 9) where subjects were not satisfied on the video streaming quality provided by ethio telecom. Only at one location users were relatively satisfied with MOS value of 3.82 which is greater than 3.5.

Figure 6: 17 below shows the maximum, average and minimum MOS value that was achieved in the specified location of the experiment. The maximum and minimum MOS were achieved at Yode restaurant and ethio telecom headquarter with 4.29 and 1.46 respectively. The highest average MOS was achieved with 3.82 at Yode restaurant around Olympia. The lowest average MOS value was achieved with 2.10 at Riche. The MOS result that were achieved greater than 3.5 was at one location out of the nine specified locations of the experiment.

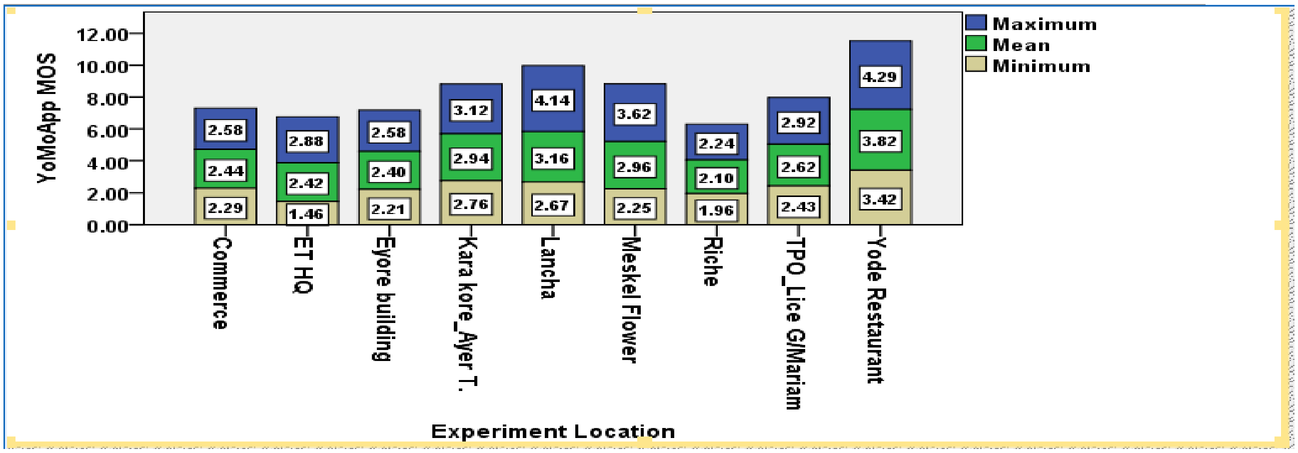


Figure 6: 17 Result of Overall MOS per Experiment Locations

The environment factors in Figure 6: 18 below shows that, users who watched video clips at outdoor environment have relatively the lowest MOS value of 2.78 and, whereas users who watched video at indoor environment had the highest MOS value of 2.81. However, we had seen in Figure 6: 12 of this thesis, the outdoor environment which achieved download throughput measurement results of 2.67 Mbps and the indoor environment achieved download throughput of 1.03Mbps. This finding confirms that download throughput is not the only factor that influences the QoE of video streaming but other factors such as environment context can impact on QoE of users.

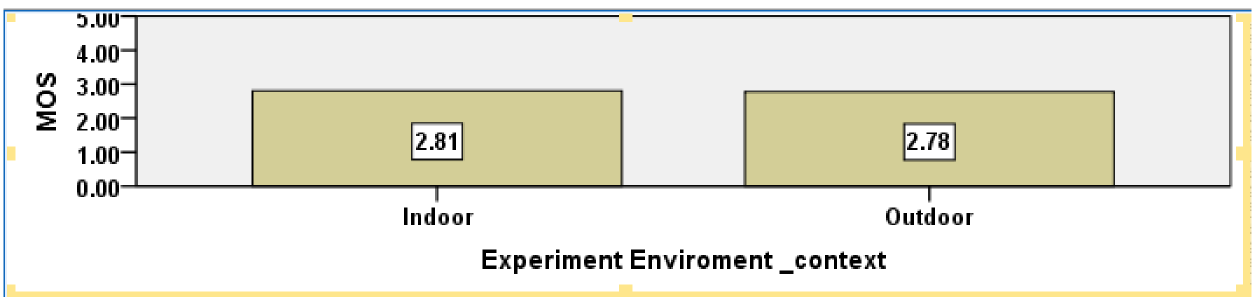


Figure 6: 18 MOS per Environment of Subjects

Perceptions of the subjects for the best and worst result on the parts of the day for their video streaming quality with respect to the MOS value are illustrated in Figure 6: 19 below. The result for the parts of the day in which subjects have rated the videos with the smallest MOS value during the morning this part of the day is considered as a busy hour for 3G service in ethio telecom context, which supports the rating of the customers with 2.67 MOS. Furthermore, the result from subjective assessment shows that the largest MOS value of 2.92 was achieved during the evening, and the MOS value 2.87 was acheived during the afternoon. The highest download throughput was achieved during the afternoon, but the highest MOS value was achieved in the evening parts of the day. The result confirms that better performance of download throughput only does not give good perception of QoE for video streaming.

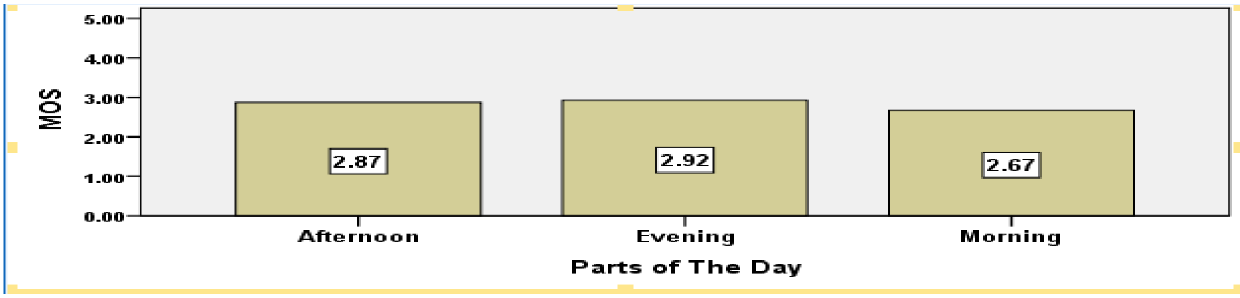


Figure 6: 19 Result of Overall MOS per Experiment Parts of the Day

Figure 6: 20 below illustrates that, 13 users who have experienced average initial delay between 1.96 and 9.68 had the highest average MOS of 3.23, 16 users who have experienced average initial delay between 10.22 and 19.70 had achieved MOS of 2.59 and the rest 6 users experienced the initial delay more than 24.55 had achieved MOS value of 2.40. From this thesis result, we can conclude that when users' initial delay duration increases, their quality of experience has dropped significantly. The result shows that subjects who had experienced longer initial delay duration had the smallest MOS value and subjects who had experienced shorter initial delay had the highest MOS value. This result confirms that a longer initial delay duration has a negative impact on users' QoE perception.

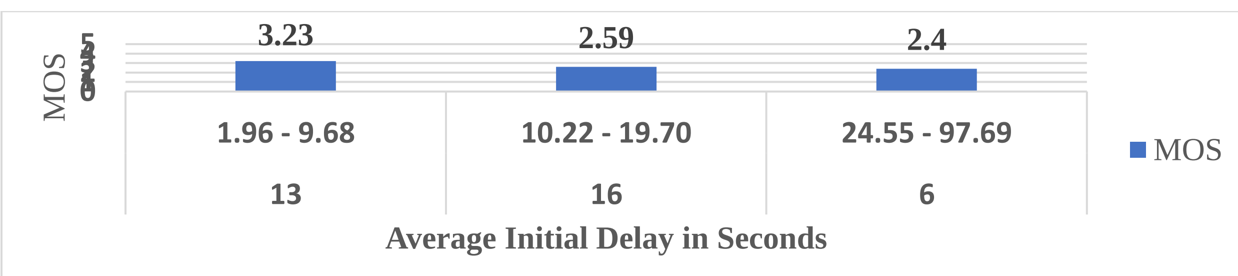


Figure 6: 20 Impact of Average Initial Delay on QoE

Figure 6: 21 below illustrates that, users who have not experienced stalls had the highest average MOS of 3.24, by experiencing average stalling events between 0.1 and 0.5 it has affected user

quality of experience and dropped down a MOS value to 2.71. Those who were experienced average stalling events between 0.57 and 0.88 affected quality experience of users with a MOS value of 2.52. Furthermore, users who experienced average stalling events between 1.63 and 2.25, their QoE have dropped significantly with a MOS value of 2.49. Based on this result, the number of stalling events increases, we observed that a negative impact on the user's QoE perception.

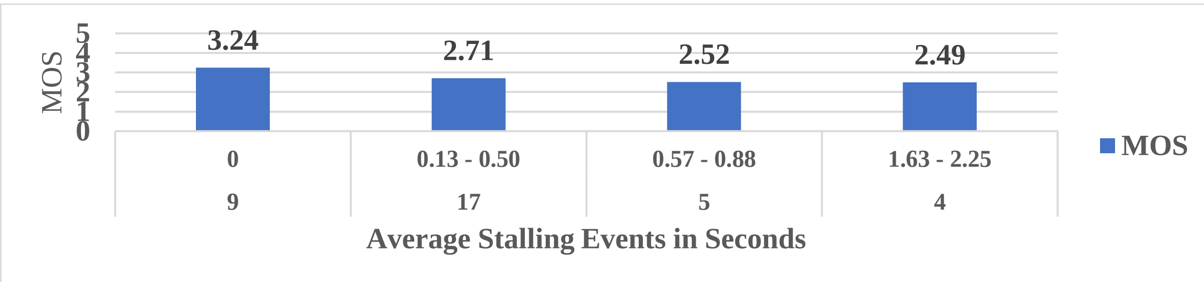


Figure 6: 21 Average MOS Value per Number of Stalling Events

Figure 6: 22 below illustrates that users who have experienced longer average total stalling duration had achieved lower MOS value of 3.24, by experiencing average total stalling duration between 0.09 and 5.82 seconds the user quality of experience dropped down to a MOS value of 2.66. Those who were experienced an average total stalling duration between 6.29 and 88.15 seconds also affected a little bit the quality experience of users with a MOS value of 2.60. Based on this result, when the average total stalling duration increases, we observed that a negative impact on the user's QoE.

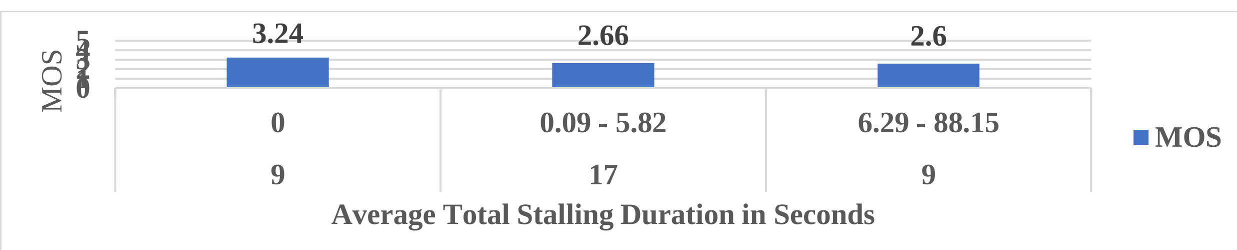


Figure 6: 22 Average MOS Value per Average Total Stalling Duration

6.3.3.3. Other Factors

One of the other factors that can impact the QoE of subjects while watching video streaming session is context. Contextual factors, in general, can cover many different aspects, which are all important for QoE measurement. In this study, environment, gender and age factors have been considered and analyzed for their impact on MOS value.

Figure 6: 23 below shows that females were acheived an average MOS of 3.2, whereas males acheived an average MOS value of 2.98. However, this result could also due to not considering an equal number of genders in the sample users, or the sample video content hadn't been in favour of the male. Therefore, it is recommended to consider gender equality when selecting sample users in future studies.

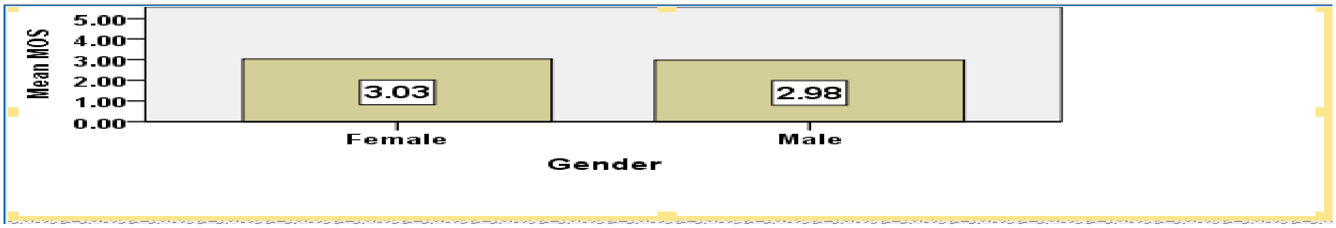


Figure 6: 23 MOS per Subjects' Gender

Based on our study shown in Figure 6: 24 below that, subjects below 35 years old are relatively better MOS value of 2.94. The participants with the age greater than or equal to 35 years old had with less MOS value of 2.75. However, this result could also be due to not considering an equal number of age groups when selecting sample users. Therefore, it is recommended to consider age group equality when selecting sample users in future studies.

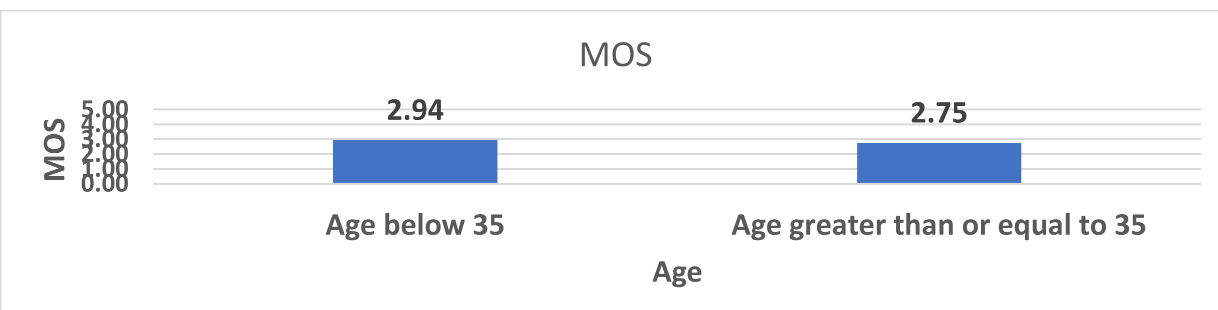


Figure 6: 24 MOS Value per Subjects' Age

Moreover, based on our study results the device factors such as application factors for example, device model, Android version, CPU speed, the number of cores in the CPU, device RAM and

Smartphone screen size of users had not great impact in affecting user QoE or MOS of YouTube video streaming. Author in [23] stated that, in HAS the content provider streaming server divides the video into chunks of fixed length (e.g. 2 seconds) of different quality, and the player, on behalf of the user, requests the video quality that suits to the device capabilities (e.g. screen size) and available bandwidth. In [48] also evaluated the influence of different types of mobile devices, ranging from a smartphone to a laptop, on the perceived quality of multimedia. The authors in [48] found that the impact of the device and screen size have no substantial influence on the quality perception of users in the case of watching quality videos.

6.4. Correlation Analysis for Experimental Evaluation Results

Metrics that have been considered in correlation analysis cover application-level parameters, namely, average video resolution watched, average initial delay, average number of stalling events, average total stalling length and average buffer level. Moreover, the network level parameter of average download throughput is considered. Finally, the actual video quality subjective rating given by the user as the perceived quality was measured in terms of video quality acceptability and MOS. Moreover, we found some relation between the subjective (perception) and the achieved objective metrics results. The table in Appendix D shows the linear correlation analysis results between the measured parameters of subjective and objective metrics results of the experimental measurements of the thesis work.

The cells of the table show the spearman rank-order correlation coefficient between the two parameters in the respective row and column. Based on that we have observed some parameters had a significant correlations in our experimental study, e.g., average initial delay had a strong negative linear correlation with MOS, average acceptability rating, average download throughput, average video resolution watched and average device buffer level with correlation coefficient value of -0.489, -0.707, -0.607 & -0.747 respectively at 0.01 significance level (p-value).

The average stalling event has a negative strong linear correlation with MOS, average acceptability rating and device average buffer level with correlation coefficient value of -0.457, -0.624 & -0.523 respectively at 0.01 significance level (p-value).

The average total stalling time duration had a strong negative linear correlation with MOS, average acceptability rating, Average download throughput and Average video resolution watched with correlation coefficient value of -0.414, -0.618, -0.422 & -0.567 respectively at 0.01 and 0.05 significance level (p-value).

We also observed that Average video resolution watched metric had a strong positive linear correlation with MOS, average acceptability rating, average download throughput and average buffer level with correlation coefficient value of 0.776, 0.891, 0.532 & 0.891 respectively at 0.01 significance level (p-value).

6.5. Discussion

The average initial delay metrics on the network level shows that the users were experienced longer initial delay around 19 seconds, and the application-level based on experimental evaluation result shows that on average 17.32 seconds. Further, the subjective assessment on the annoyance level of users shows that on average 2.78 which means that the initial delay was on average perceived by the users as noticeable and annoying. All in all, the result from the three different measurements results shows that users were experienced a longer initial delay and which was annoying the user. In [49] the authors confirmed that the initial delay has more impact on the video abandonment than the stalling event in the middle of playback. In [50], a large-scale user study shows that the startup delay has a significant impact on a user’s online time, and if the startup delay exceeds two seconds, the user may stop watching the video.

The average number of stalling events metrics on the network level during the period of experimental evaluation shows that all users have experienced stalling events on average 0.38 times per minute, and the application level experimental evaluation result shows that 0.47 times per 30 seconds. Further, the subjective assessment on the annoyance level of users rating shows that 4.54 rating which means that the stalling events were perceived by the users as noticeable but not annoying. All in all, the result from the three different measurements results shows that all users were experienced few stalling events, and which was not annoying. This confirms the findings of [51] that adaptive video streaming is mostly able to avoid stalling in mobile video streaming sessions. In [4] most users can tolerate moderate initial delays, stalling has a significant impact, as already little stalling event severely reduces the perceived quality by the users.

The average total stalling time duration metrics on the network level during the period of experimental evaluation shows that on average users were experienced an additional 6.46% stalled time rate on the video duration time, and the application level experimental evaluation result showed that on average 6.04 seconds, additional time due to stalling time, per 30 seconds. Further, the subjective assessment on the annoyance level, for a stalling time duration, of users rating showed that 4.26 which means that the stalling time duration was perceived by the users as noticeable but not annoying. All in all, the result from the three different measurements results showed that users were experienced a shorter time of stalling duration, and which was not annoying. Authors in [52] their results show that Video MOS (vMOS) is essentially affected by many QoS parameters such as initial buffering latency, stalling ratio, and stalling times.

The average video resolution, that the users watched, metric on the network level are not measured in ethio telecom, and the application level experimental evaluation result shows that users watched video on average 230p resolution. Further, the subjective assessment on the video resolution rating showed that MOS value of 2.57 which means that the quality of video resolution was very low, so that users perceived poor quality of the video. All in all, the result from the two different measurements results shows that users were experienced lower video resolution or poor video quality, and which leads user’s dissatisfaction.

It has been found from our evaluation results and data analysis, that was not only the network performance affect QoE of video streaming, but also contextual factors do, such as location, and user profiles, i.e. age, gender.

From the analysis of the Experiment result it is clear that the male users do have much higher average MOS value compare with female, and the younger users have much higher average MOS compare with the elders, which shed some light on female users are harder to be satisfied with the service provided compare with male, so does the elders. Based on further comprehensive study ethio telecom as a sole service provider should pay more attention to the female and the elder users in order to satisfy them in a better approach, i.e. Advertisement.

Moreover, our study results show that the smartphone users were most positive while watching the YouTube clips in the evening but much less pleased at the morning and afternoon. This can be easily explained that users are more relax and casual to watch videos in private and freely without daily routine compare with watching videos in the public together with their day to day tasks. Even though the average throughput was better in the afternoon, the users who had watched sample videos at that time could still be stressed by the atmosphere which leads to lower satisfaction and leading to rate lower values from those who had watched the samples in the evening. Additionally, Quality of experience also strongly relies on many subjective factors, such as user experience, user interest and expectation. The users might receive very high QoS from the network provider, however, they could still possible be less satisfied due to their expectation has not been met due to bad video quality, or they do not like the video content.

We have also observed a positive and negative linear relationship between subjective and objective metrics of the video streaming evaluation study of this thesis work. See Appendix C for correlation table results of subjective and objective metrics. The average initial delay had a strong negative linear correlation coefficient of -0.489 at 0.01 significance level (p-value) with the MOS.

The Average stalling event had a negative strong linear correlation coefficient of -0.457 at 0.01 significance level (p-value) with MOS, and stalling duration also had a strong negative linear correlation coefficient of -0.414 at 0.05 significance level (p-value) with MOS. Regarding stalling author in [30] presented that not only the number of stalling events but also the stalling duration of the playback of video which occurs due to buffer depletion, have a high correlation with poor quality perceived by the end-users. Specifically, the authors reached to conclude that a video with 2 stalling events with 3 seconds stalling duration each, will lead to significantly lower Mean Opinion Score (MOS). In [53], the authors found that the users with the stall duration were more than three seconds, the dissatisfaction increased. The author in [54] found that the number of stalling events has the highest impact on QoE and that a medium number of stalling events can result in decreasing a MOS value by 2 points. Average video resolution watched metric had a strong positive linear correlation coefficient of 0.776 at 0.01 significance level (p-value)with MOS value.

Chapter 7

Conclusion and Future Work

7.1. Conclusion

In this paper, we have presented the results of a field study for performance evaluation of YouTube streaming quality which was conducted for two weeks in ten locations of Addis Ababa, using crowdsourcing applications with a connection of 3G mobile networks. The video QoE was computed using standardized objective and subjective methods. In this paper, we have presented results from network-level data analysis and experimental measurement of user study for QoE in the context of adaptive video streaming of YouTube videos. The objectives of the study presented in this paper were twofold: first we aimed to evaluate QoE in the real environment of smartphone users in 3G network of ethio telecom by considering different contextual and device factors then we have done correlation analysis on the results found from objective and subjective evaluations.

After dealing with the video streaming quality management process in ethio telecom, four basic video quality metrics have been identified – initial delay, stalling events, stalling duration and resolution watched from application side and download throughput evaluation for network side assessment. In addition to the quality metrics, subjective metrics have been collected and identified which can show us the perception of the users. Data has been collected for quality evaluation from NetProb3010, YouTube Monitoring Application (YoMoApp) and Speedcheck pro together using a crowdsourced approach in Addis Ababa. The subjective quality indicator metrics have also been collected by preparing a questionnaire before and after watching the sample eight videos for 35 sample 3G users of ethio telecom in Addis Ababa.

From the collected data, network-level NetProb3010 evaluation results for the average daily initial delay, stalling event and stalling duration was 19 seconds, 0.38 times per minute and 6.46% for average stalled time rate respectively. Based on application-level crowdsourcing YoMoApp evaluation results the average initial delay, average number of stalling events, average stalled time duration and average video resolution watched results were 17.3 seconds, 0.47 times per 30 seconds, 3.79 seconds and 230P respectively and we also measured application-level download throughput of each device while the users were watching the sample videos and the average download throughput was 1.78Mbps. For better customer satisfaction, the recommended video streaming service has to be High Definition (HD) smooth streaming without interruption and the recommended download throughput has to be above 4 Mbps. However, the results obtained here are lower than the recommended values. Thus, this result can bring to customer dissatisfaction and which has been reflected in the responses of the 3G customers with an overall MOS value of 2.79 for the content, streaming and video quality evaluation of YouTube videos.

QoE of users was also impacted by several subjective factors in addition to objective factors, such as the context of watching, age, gender, video content etc. Not always the highest throughput will lead to the best QoE, and not always the best technology smartphone devices deserve the best QoE or MOS achieved by the users. Nevertheless the user received very high downlink throughput, their MOS value may still be lower because of stress by the context of the environment and users might not like the video content. In addition, objective metrics results had some relation with subjective metrics (MOS) results, see appendix D. However, multiple parameters shall be considered together in the future to draw a more accurate correlation with QoE.

From both the objective and subjective metrics evaluation result, continuous assessment for the satisfaction level of users on video streaming service must be a crucial task for ethio telecom. Thus, ethio telecom as an operator and service providers should incorporate user-level subjective evaluation method in its video streaming quality or performance evaluation to meet customer satisfaction and expectation. Moreover, ethio telecom should not only focus on capturing the attention of new customers but also try to guarantee the quality level of video streaming, and the download throughput by optimizing its 3G network infrastructure based on location and coverage assessment in order to provide a better quality of video streaming service.

Finally, ethio telecom does not only need to spend too much money on focusing on developing the newest technology and trying to provide the customers with the highest throughput. In addition to that, it shall also give attention to user’s perception results that mention above in this study, and try to cooperate well with service providers and content providers in order to reach the maximum customer satisfaction. By doing so, ethio telecom can generate more profit margin and save the cost in order to reduce the revenue gap.

7.2. Future Work

This study was only a first step and further research is needed to better understand video streaming quality and user perception in the context of adaptive video streaming service. As a future work, we suppose to further details on the possible impacts of the social setting (through different setups, e.g., in laboratory and naturalistic settings with higher ecological validity). Additionally, we aim to take a closer look at the interplay between content likeability, mood and QoE; and to investigate how traditional and alternative measures of QoE relate to each other. Other researchers can also consider other adaptive streaming solutions and terminal devices, additional bitrate profiles (e.g., from low to high bitrate), alternative test setups as well as different types of content.

The sample devices considered in the experiment were only smartphone users so further studies can be done on by combining feature phones and smartphone users on different video streaming platforms and social medias. Further, the sample videos considered in the experiment were 8 different types of video content; however, these videos' resolution is uploaded to my youtube channel all in 720p HD video resolution. Thus, different types of videos with different resolutions (such as 1080p, 480p, 360p) shall be used instead of the same resolution. Since different video resolutions require different codecs which is necessary for QoE measurement. In addition, all video should be selected from the 'QoE certified database', where all video clips have been tested. By doing so, the result of this study can be used to compare with our thesis work so that the findings can be easily verified by the prior art. Moreover, a more comprehensive Android application could be developed to measure objective and subjective metrics QoE at the same time, this can avoid users skipping between different Application and hard copy questionnaire during the evaluation.

Finally, if the advanced tool will be available in the future, different videos with different resolution will be chosen from well known certified databases and wide-ranging unbiased sample group can be target. Hence, a more comprehensive further study can be possibly done based on this research to find similarities and extrapolate the results with the Network Speed Test result and YouTube video monitoring tool such as YoMoApp to find a complete QoE-QoS mapping scale. Moreover, it is recommended to consider gender equality and video neutrality when selecting sample users and videos in future studies.

References

[1] Cisco, Cisco Visual Networking Index: Forecast and Trends, 2017–2022, Cisco White Paper, 2019.

[2] Cisco, Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2017–2022, Cisco White Paper, 2018.

[3] B. Tamene, E. Kebede and Z. Fitiwe, “opportunity and challenges of OTT in ethio telecom,” MSc. seminar, Dept. Elect. and Comp. Eng, Addis Ababa Univ., Addis Ababa, 2019.

[4] International Telecommunication Union, Methods for the subjective assessment of video quality, audio quality and audiovisual quality of Internet video and distribution quality television in any environment, ITU-T P.913, 2016.

[5] T. Hoßfeld, M. Seufert, M. Hirth, T. Zinner, P. Tran-Gia, and R. Schatz, “Quantification of youtube qoe via crowdsourcing,” in Multimedia (ISM), 2011 IEEE International Symposium on, Dec 2011, pp. 494–499.

[6] R. K. Mok, E. W. Chan, X. Luo, and R. K. Chang, “Inferring the qoe of http video streaming from user-viewing activities,” in Proceedings of the First ACM SIGCOMM Workshop on Measurements Up the Stack, ser. W-MUST ’11. New York, NY, USA: ACM, 2011, pp. 31–36. [Online]. Available: http://doi.acm.org/10.1145/2018602.2018611

[7] T. Hoßfeld, S. Egger, R. Schatz, M. Fiedler, K. Masuch, and C. Lorentzen, “Initial Delay vs. Interruptions: Between the Devil and the Deep Blue Sea,” in QoMEX, 2012.

[8] T. Hoßfeld, M. Seufert, M. Hirth, T. Zinner andP. Tran-Gia “Quantification of YouTube QoE via Crowdsourcing,” IEEE International Symposium on Multimedia, 2011.

[9] F. Wamser, M. Seufert, P. Casas, R. Irmer, P. Tran-Gia and R. Schatz, "YoMoApp: A tool for analyzing QoE of YouTube HTTP adaptive streaming in mobile networks," *2015 European Conference on Networks and Communications (EuCNC)*, Paris, 2015, pp. 239-243

[10] J. Yao, S. S. Kanhere, I. Hossain, and M. Hassan, “Empirical Evaluation of HTTP Adaptive Streaming Under Vehicular Mobility,” in Proceedings of the 10th International IFIP TC 6 Networking Conference: Networking 2011, Valencia, Spain, 2011.

[11] B. Lewcio, B. Belmudez, A. Mehmood, M. W¨altermann, and S. M¨oller, “Video Quality in Next Generation Mobile Networks – Perception of Time-varying Transmission,” in Proceedings of the 2011 IEEE International Workshop Technical Committee on Communications Quality andReliability (CQR), Naples, FL, USA, 2011.

[12] M. Seufert, S. Egger, M. Slanina, T. Zinner, T. Hoßfeld, and P. Tran Gia, “A Survey on Quality of Experience of HTTP Adaptive Streaming,” IEEE Communications Surveys & Tutorials, vol. 17, no. 1, pp. 469–492, 2015.

[13] I. Ketykó, K. Moor, T. Pessemier, A. Verdejo, K. Vanhecke, W. Joseph, L. Martens and L. Marez, “QoE Measurement of Mobile YouTube Video Streaming,” Department of Information Technology, Ghent University IBBT, Proceedings of the 3rd workshop on Mobile video delivery - MoViD' 10, 2010. [Online], Available: https://www.academia.edu/18508895/QoE_measurement_of_mobile_YouTube_video_streaming/, [Accessed Jun. 11, 2019].

[14] K. De Moor, I. Ketyko, W. Joseph, T. Deryckere, L. De Marez, L. Martens, and G. Verleye. “Proposed framework for evaluating quality of experience in a mobile, testbed-oriented living lab setting,” Mobile Networks and Applications, Springer Netherlands. vol. 15, no. 3, pp. 378-391, 2010.

[15] N. Hyunwoo & K. Kyung-Hwa & S. Henning, “QoE Matters More Than QoS: Why People Stop Watching Cat Videos,” Conference: IEEE INFOCOM 2016 - IEEE Conference on Computer Communications, 2016.

[16] N. Hyunwoo and S. Henning, “YouSlow: What Influences User Abandonment Behavior for Internet Video?,” Columbia University, New York, NY., Tech. Rep., 2016.

[17] A. Toskala and H. Holma, “UMTS Services and Applications,” Jouni Salonen, John Wiley & Sons, Ltd, 2002.

[18] GSMA, “Mobile Technology,” 2018. Available: https://www.gsma.com/aboutus/gsm-technology/hspa /, [Accessed Jun. 11, 2019].

[19] D. Striccoli, Transmission Optimization of Digital Compressed Video in Wireless Systems-Scientific Figure on ResearchGate, 2010, [Online]. Available: https://www.researchgate.net/figure/The-3G-network-archtecture-The-Core-Network [Accessed Sep. 30, 2019].

[20] S. Junde and F. Ruijun, "Some QoS issues in 3G Wireless Networks," TENCON '02. Proceedings. 2002 IEEE Region 10 Conference on Computers, Communications, Control and Power Engineering, vol. 2, pp. 724 -727, 2002

[21] ITU Telecommunication Standardization Sector and OF ITU. “Definition of quality of experience (QoE), liaison statement”. ITU-T Recommendation P.10/G.100, Amd 1, 2007.

[22] European Telecommunication Standards Institute TC HF (Human Factors). “Quality of Experience (QoE) requirements for real-time communication services”. Technical report, European Telecommunication Standards Institute, 2010.

[23] C. Sieber, A. Blenk, M. Hinteregger, and W. Kellerer, ‘The cost of aggressive HTTP adaptive streaming: Quantifying YouTube’s redundant traffic’, in 2015 IFIP/IEEE International Symposium on Integrated Network Management (IM), 2015, pp. 1261–1267.

[24] A. Rao, Y. Lim, C. Barakat, A. Legout and D. Towsley. Network Charac-teristics of Video Streaming Traffic. CONEXT 2011, Tokyo, Japan. inria-00638063, Dec 2011.

[25] M.D. Walker, M. Nilsson, T. Jebb, and R. Turnbull, Mobile Video-Streaming. *BT Technology Journal* 21, 192–202, 2003. [Online]. Available: https://doi.org/10.1023/A:1025183704729. [Accessed: Oct. 02, 2019].

[26] YouTube, YouTube for Press, 2019. [Online]. Available: https://www.youtube.com/intl/en-GB/about/press. [Accessed: Aug. 02, 2019].

[27] A. Begen, T. Akgul and M. Baugher, "Watching video over the web: Part 1: Streaming protocols," IEEE Internet Computing, Volume 15, Issue 2, 2011.

[28] R. Ashwin, L. Arnaud, L. Yeon-sup, T. Don, B. Chadi, and D. Walid, "Network characteristics of video streaming traffic," in Proceedings of the Seventh Conference on emerging Networking Experiments and Technologies (Conext '11).ACM, New York, NY, USA, Article 25, 12 pages.

[29] H. Tobias, S. Raimund, B. Ernst and P. Louis, "Internet Video Delivery in YouTube: From Traffic Measurements to Quality of Experience," Data Traffic Monitoring and Analysis Lecture Notes in Computer Science Volume 7754, 2013, pp 264-301.

[30] T. Hoßfeld, M. Seufert, M.Hirth, T. Zinner and P. Gia, "Quantification of YouTube QoE via crowdsourcing,", University of W ̈urzburg, Institute of Computer Science, Wurzburg, Germany, 2011

[31] R. Ashwin, L. Arnaud, L. Yeon-sup, T. Don, B. Chadi, and D. Walid, "Network characteristics of video streaming traffic," in Proceedings of the Seventh Conference on emerging Networking Experiments and Technologies (Conext '11).ACM, New York, NY, USA, Article 25, 12 pages.

[32] M. Fiedler, T. Hoßfeld, and P. Tran-Gia, "A generic quantitative relationship between quality of experience and quality of service," Network, IEEE, vol. 24, no. 2, pp. 36–41, 2010.

[33] R. Serral-Gracia`, E. Cerqueira, M. Curado, M. Yannuzzi, E. Monteiro, and X. Masip-Bruin, "An overview of quality of experience measurement challenges for video applications in IP networks," in Proc. International Conference on Wired/Wireless Internet Communications (WWIC), 2010, pp. 252–263.

[34] Z. Wang, L. Lu, and A. Bovik, "Video quality assessment using structural distortion measurement," in Image Processing. 2002. Proceedings. 2002 International Conference on, vol. 3, 2002, pp. III–65–III–68 vol.3.

[35] B. Staehle, M. Hirth, F. Wamser, R. Pries and D. Staehle, "YoM: A YouTube Application Comfort Monitoring Tool.". Berlin, Germany: Inst. Informatik, Tech.Rep. 467, 2010.

[36] A. Dalal and E. Perry, "A New Architecture for Measuring and Assessing Streaming Media Quality," Hewlett-Packard Laboratories, Palo Alto, CA, USA, 2003.

[37] R. Saraiva, "Wireless Communications and Mobile Computing," Volume 2017, Article ID 2315036, 17 pages, Campos Department of Computer Engineering, CEFET/RJ, Campus Petrópolis, Petrópolis, RJ, Brazil

[38] Huawei, “NetProb3010 SEQ Analyst KQI&KPI Modeling summary,” Version: 05, Oct. 2015.

[39] ITU, Regulatory & Market Environment, Quality of Service Regulation Manual, ITU, 2017.

[40] A. Reesom, “Quality of Experience Evaluation for Addis Ababa UMTS Enterprise Data Customers,” MSc. dissertation, Dept. Elect. and Comp. Eng, Addis Ababa Univ., Addis Ababa, 2018.

[41] Index, Cisco Visual Networking. "Global Mobile Data Traffic Forecast Update, 2012–2017, Cisco White Paper, Feb. 6, 2013."

[42] F. Wamser, M. Seufert, P. Casas, R. Irmer, P. Tran-Gia, and R. Schatz,“Yomoapp: A tool for analyzing qoe of youtube http adaptive streaming in mobile networks,” in European Conference on Networks and Communications (EuCNC), 2015.

[43] Yomoapp, Chair of Communication Networks University of Wurzburg, Version 1.6.6, https://play.google.com/store/apps/details?id=org.speedspot.speedanalytics [accessed 30 Jul, 2019].

[44] N. Wehner, “YoMoApp – Documentation,” pp. 1-9, Apr. 3, 2018.

[45] Speedcheck Organization, “Operation Mannual,” [online]. Available: https://www.speedcheck.org/. [Accessed Jul. 12, 2019].

[46] M. Tolstrup, Indoor Radio Planning - A Practical Guide for GSM, DCS, UMTS, HSPA and LTE, 2nd ed. John Wiley & Sons, Ltd., 2011.

[47] C. Pedro and S. Raimund, “Quality of Experience in Cloud Services: Survey and Measurements,” Computer Networks, vol. 68, Aug. 2014.

[48] T. Hoßfeld, S. Egger, R. Schatz, M. Fiedler, K. Masuch, and C. Lorentzen, “Initial Delay vs. Interruptions: Between the Devil andthe Deep Blue Sea,” in Proceedings of the 4th International Work shopon Quality of Multimedia Experience (QoMEX), Yarra Valley, Australia,2012.

[49] H. Nam and H. Schulzrinne, “YouSlow: What Influences User Abandonment Behavior for Internet Video? ” , Columbia University, New York, 2016

[50] S. Shunmuga, K. Ramesh and K. Sitaraman, “Video Stream Quality Impacts Viewer Behavior: Inferring Causality Using Quasi-Experimental Designs,” Akamai Technologies, University of Massachusetts, Amherst & Akamai Technologies, Boston, Massachusetts, USA, 2012.

[51] A. Bentaleb, B. Taani, A. Begen, C. Timmerer and R. Zimmermann, “A Survey on Bitrate Adaptation Schemes for Streaming Media over HTTP,”. IEEE Communications Surveys & Tutorials, 2018.

[52] O. Wang, H. Dai, D. Wu, and H. Xiao, “Data analysis on video streaming QoE over mobile networks,”. Journal on Wireless Communications and Network, 2018, [Online]. Available: https://doi.org/10.1186/s13638-018-1180-8

[53] R. Mok, E. Chan, X. Luo, and R. Chang, “Inferring the QoE of HTTP Video Streaming from User-Viewing Activities,”. W-MUST '11 Proceedings of the First ACM SIGCOMM Workshop on Measurements up the stack, Toronto, Canada, 19 Aug 2011, p. 31-36, [Online]. Available: http://hdl.handle.net/10397/37570

[54] R. K. P. Mok, E. W. W. Chan and R. K. C. Chang, "Measuring the quality of experience of HTTP video streaming," 12th IFIP/IEEE International Symposium on Integrated Network Management (IM 2011) and Workshops, Dublin, 2011, pp. 485-492.

Appendix A - Experiment Instruction

1. Download and install YouTube Monitoring Application (YoMoApp) and Speedcheck Pro on your android smartphone from google play
2. Setups your Smartphone devices as follows
    - Make the brightness of your device adaptive
    - If there is an advertisement pop up, click skip Immediately
    - lock your android device on 3G or UMTS network before watching the video sess ions
3. Restart your device before starting the evaluation this is used to stop other applications consumin g or sharing the download throughput of your device at the background.
4. Before starting to watch the sample videos through YoMoApp, please configure the Speedcheck Pro Application

    4.1 Open your speedcheck pro application, go to schedule menu then configure as below
    - click network and select Cellular/Mobile
    - set the beginning and ending date of the test
    - click Repeat test on and set to every 5 minutes duration to repeat the test
5. Open your YoMoApp application, make location service on and accept rating request

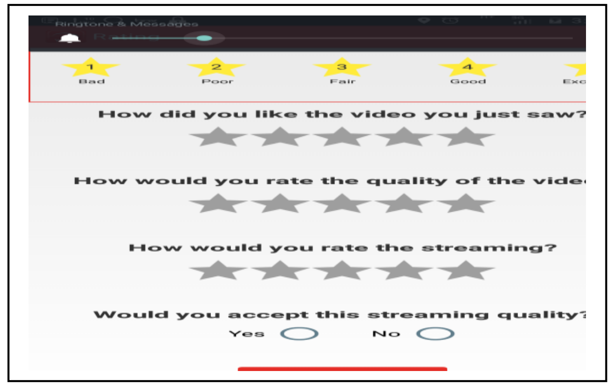
    5.1 Write zel fit, that is my YouTube channel name, in the search button and watch each video immediately. Rate the online questionnaire on the application and hard copy questionnaire after each video

    5.2 You can make the videos screen display size option as per you're watching (viewing) habit of you tube videos during the evaluation experiment
6. After finishing watching of each video sample, please come back to the hard copy perception questionnaire to fi rate your annoyance level on the streaming:
7. Please make sure to fill in your correct location where you watched the videos at the end of the Survey and your Device ID on YoMoApp
8. After 30 minutes of the starting time of the evaluation experiment you will get 6 results of the speedcheck pro application, then go to setting menu and export the result to email address zelfit@gmail.com

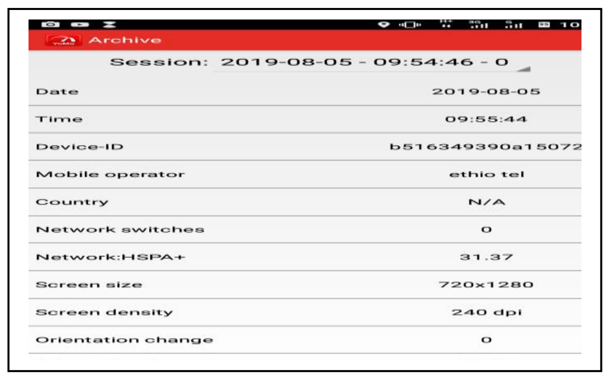
Thanks for your support

Appendix B – YoMoApp & Speedcheck Pro Tools Sample Interface & Results

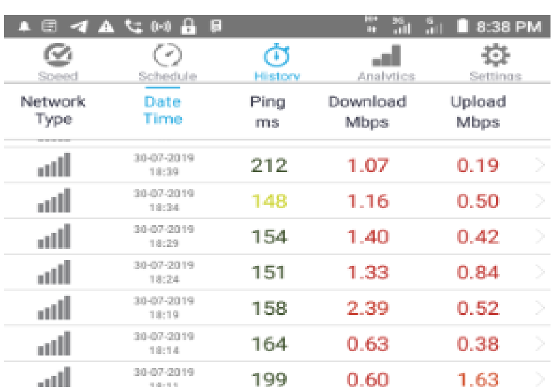
1. YoMoApp Application User Interface and Rating Scale

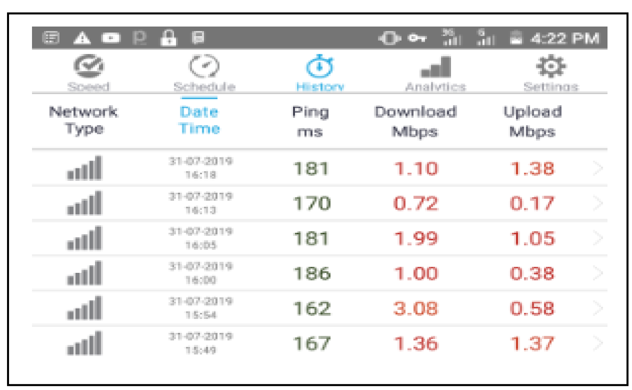


2. YoMoApp Application Sample Results



3. Speedcheck Pro Application Sample Results





Appendix C - Correlation results of subjective and objective metrics

Correlations

			MOS	Average Acceptability rating	Average Download _Mbps	Avrage_Initial delay	Average_Stalling events	Average Total stalling length	Average Video Resollution Watched	Average of Average buffer level
Spearman's rho	YoMoApp MOS	Correlation Coefficient	1.000	.852	.398	-.489	-.457	-.414	.776	.669
		Sig. (2-tailed)	.	.000	.018	.003	.006	.013	.000	.000
		N	35	35	35	35	35	35	35	35
	Average Acceptability rating	Correlation Coefficient	.852**	1.000	.543**	-.707**	-.624**	-.618**	.891**	.797**
		Sig. (2-tailed)	.000	.	.001	.000	.000	.000	.000	.000
		N	35	35	35	35	35	35	35	35
	Average Download _Mbps	Correlation Coefficient	.398*	.543**	1.000	-.607**	-.313	-.422*	.532**	.601**
		Sig. (2-tailed)	.018	.001	.	.000	.067	.011	.001	.000
		N	35	35	35	35	35	35	35	35
	Avrage_Initial delay	Correlation Coefficient	-.489**	-.707**	-.607**	1.000	.552**	.656**	-.747**	-.750**
		Sig. (2-tailed)	.003	.000	.000	.	.001	.000	.000	.000
		N	35	35	35	35	35	35	35	35
	Average_Stalling events	Correlation Coefficient	-.457**	-.624**	-.313	.552**	1.000	.888**	-.486**	-.523**
		Sig. (2-tailed)	.006	.000	.067	.001	.	.000	.003	.001
		N	35	35	35	35	35	35	35	35
	Average Total stalling length	Correlation Coefficient	-.414*	-.618**	-.422*	.656**	.888**	1.000	-.567**	-.658**
		Sig. (2-tailed)	.013	.000	.011	.000	.000	.	.000	.000
		N	35	35	35	35	35	35	35	35
	Average Video Resollution Watched	Correlation Coefficient	.776**	.891**	.532**	-.747**	-.486**	-.567**	1.000	.891**
		Sig. (2-tailed)	.000	.000	.001	.000	.003	.000	.	.000
		N	35	35	35	35	35	35	35	35
	Average of Average buffer level	Correlation Coefficient	.669**	.797**	.601**	-.750**	-.523**	-.658**	.891**	1.000
		Sig. (2-tailed)	.000	.000	.000	.000	.001	.000	.000	.
		N	35	35	35	35	35	35	35	35

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).

Appendix D - Subjective Metrics Questionnaire

Pre watching questionnaire part I

1. Where are you?

- ▢ On vehicle transportation
- ▢ Office
- ▢ Home
- ▢ Outside home and office (outdoor)

2. In which Age group are you?

- ▢ Below 18
- ▢ 35 – 44
- ▢ 18 - 24
- ▢ 45 – 54
- ▢ 25 - 34
- ▢ 55 & above

3. What is your current Level of Education?

- ▢ High school
- ▢ Second degree
- ▢ Diploma or Certificate
- ▢ Third degree
- ▢ First degree

4. What is your job status?

- ▢ Investor
- ▢ Government or Private employee
- ▢ Business Owner
- ▢ Job seeker

5. Gender

- ▢ Female
- ▢ Male

6. Are you currently using 3G mobile network to watch mobile YouTube videos?

- ▢ Yes
- ▢ No

7. How is your video habits or experience?

- ▢ 1 (bad)
- ▢ 4 (Good)
- ▢ 2 (poor)
- ▢ 5 (Excellent)
- ▢ 3 (Fair)

8. Do you have eye disease history?

- ▢ Yes
- ▢ No

9. What is your mobile phone brand?

10. What is your Mobile phone model?

--   --------------------------------------

Post watching questionnaire part II

10. How would you evaluate the perception of annoyance level of initial loading time of the video during video playback?

Video Clips	1 (perceptible and very annoying).	2 (noticeable, annoying)	3 (noticeable, slightly annoying)	4 (noticeable, not annoying)	5 (not perceptible)
Movie					
Sport					
News					
Advertisement					
Slow-motion					
Lecture					
Music					
Cartoon					

11. How would you evaluate your perception on annoyance level of frequency of interruptions (buffering/stalling) events, excluding initial delay interruption, as annoying during video viewing session?

Video Clips	1 (perceptible and very annoying).	2 (noticeable, annoying)	3 (noticeable, slightly annoying)	4 (noticeable, not annoying)	5 (not perceptible)
Movie					
Sport					
News					
Advertisement					
Slow-motion					
Lecture					
Music					
Cartoon					

12.How would you evaluate your perception for the duration time of each interruptions (buffering/stalling) event, excluding initial delay interruption, as annoying during video viewing session?

Video Clips	1 (perceptible and very annoying).	2 (noticeable, annoying)	3 (noticeable, slightly annoying)	4 (noticeable, not annoying)	5 (not perceptible)
Movie					
Sport					
News					
Advertisement					
Slow-motion					
Lecture					
Music					
Cartoon					

13. Where is your specific location during video watching session? ______________________

14. Date of experiment ___________________________