ANT COLONY OPTIMIZATION FOR VOIP QUALITY ASSESSMENT

By

Rami Suleiman AL-Khawaldeh

Supervisor

Dr. Mousa Tawfiq AL-Akhras

Co-Supervisor

Dr. Khalil Mohammed Ali El-Hindi

This Thesis was Submitted in Partial Fulfilment of the Requirements for the Master degree of Science in Computer Science

Faculty of Graduate Studies

The University of Jordan

July, 2010

Committee Decision

This Thesis/Dissertation (ANT COLONY OPTIMIZATION FOR VOIP QUALITY ASSESSMENT) was Successfully Defended and Approved on Thursday 15/7/2010

Examination Committee	Signature
Dr. Mousa Tawfiq AL-Akhras (Supervisor) Assist. Prof. of Artificial Intelligence	
Dr. Khalil Mohammed Ali El-Hindi (Co-Supervisor) Assoc. Prof. of Artificial Intelligence	K. Hindi
Dr. Saleh Hosni Al-Sharaeh (Member) Prof. of Mobile Computing, Wireless Networks, High Performance Computing, Parallel Processing, and Computer Architecture	
Dr. Bassam Hasan Hammo (Member) Assoc. Prof. of Natural Language Processing	Bassam Hammo
Dr. Mohammad M. Banat (Member) Assoc. Prof. of Wireless Communications (Jordan University Of Science and Technology)	

Dedication

To Islam,

The greatest religion that gives us the motivation to work and achieve

To my dear father,

Who is playing a great role in my life, supporting, encouraging, and teaching me patience and challenge to face the difficulties and obstacles

To my merciful mother,

Who is still lightening my life as a candle by her warm spirit

To my wife and son ,

Who is never ending her moral support and prayers which always acted as a catalyst in my academic life

To my sisters and friends,

Who are still encouraging and inspiring me through my studying years

To the ones,

Whom I respect and never forget, who offer me so many great things

To all those,

I dedicate this work which I hope will have your satisfaction and attention.

Acknowledgement

Thank Allah the Lord of the Worlds, and May the prayers of blessing of Allah be upon Prophet Muhammad and upon his family and all his companions. May Allah help us to learn more, bless us with piety, and save us always. Thank Allah and his bounty for guiding me to the end of this study.

Thanks to all, thanks to the University of Jordan, which gave me the opportunity to complete this work. Thanks, appreciation and gratitude to the supervisors Dr. Mousa T. AL-Akhras and Dr. Khalil EL-Hindi who kindly supervised the studies and for their time. They gave me a great thing and did not hesitate at any moment to provide counseling, advice and opinion. For their wise and judicious guidance they gave me during my study. Special thanks for Dr. Mousa T. AL-Akhras for his patience in helping me through these years.

Big thanks go to the examination committee for the time they spent reading my thesis and giving constructive comments greatly improved the final output of the thesis. My thanks to my external examiner Dr Mohammad M. Banat for the depth of his questions, suggestions and comments which improved the thesis and gave me the opportunity to give parts of work extra thought and improvement. Internal Examiners Professor Saleh Hosni Al-Sharaeh and Dr Bassam Hasan Hammo also gave very useful comments on technical writing and structure of the thesis which I think helped improving the readability of the thesis.

Sincere gratitude and many thanks for my friends Eng. Mohammad K. Khawaldeh, Hamzah M. Ameen, Raed A. Qublan, Ahmmed Qandeel, Ahmmed Fraij, Mutasem Hamaidah, Reem Taharwah and Eng. Sara Aloqaily for their supporting, helping me at this work. I really appreciate their efforts.

Finally, special thanks and sincere gratitude for my mother Mrs. Fatima Khawaldeh for her indefatigable patience, My sisters Rania, Lana and Dana , My Wife Sammar and My Son Hamzah who is gift from Allah . My deep appreciation for their enormous support to review this work.

Rami S. Khawaldeh

Amman 2010

Contents

List of Tables

List of Figures

List of abbreviations

3SQM	**S**ingle-**S**ided **S**peech **Q**uality **M**easurement
ACO	**A**nt **C**olony **O**ptimization
ACR	**A**bsolute **C**ategory **R**ating
ACS	**A**nt **C**olony **S**ystem
AD	**A**uditory **D**istortion
ANN	**A**rtificial **N**eural **N**etwork
AS	**A**nt **S**ystem
ASD	**A**uditory **S**pectrum **D**istance
ASYM	**A**bsolute (**SYM**metric)
BSD	**B**ark **S**pectral **D**istortion
CCR	**C**omparison **C**ategory **R**ating
CD	**C**epstral **D**istance
CMOS	**C**omparison **M**ean **O**pinion **S**core
CTI	**C**omputer **T**elephony **I**ntegration
dB	**d**eci**B**el
DCR	**D**egradation **C**ategory **R**ating
DMOS	**D**egraded **M**ean **O**pinion **S**core
DoD	**D**epartment of **D**efense
EMBSD	**E**nhanced **M**odified **B**ark **S**pectral **D**istortion
ETSI	**E**uropean **T**elecommunications **S**tandards **I**nstitute
FFT	**F**ast **F**ourier **T**ransformation

FMNB	**F**requency **MNB**
GA	**G**enetic **A**lgorithm
GP	**G**enetic **P**rogramming
IETF	**I**nternet **E**ngineering **T**ask **F**orce
IOS	**I**nternational **O**rganization for **S**tandardization
IP	**I**nternet **P**rotocol
IRS	**I**ntermediate **R**eference **S**ystems
IS	**I**takura**S**aito
ITU-T	**I**nternational **T**elecommunications **U**nion-**T**elecommunication Standardization Sector
LAR	**L**og-**A**rea-**R**atio
LL	**L**og-**L**ikelihood
M2E	**M**outh to **E**ar
MBSD	**M**odified **B**ark **S**pectral **D**istortion
MMAS	**MAX-MIN A**nt **S**ystem
MNB	**M**easuring **N**ormalizing **B**locks
MOS	**M**ean **O**pinion **S**core
MOS_{CQE}	**MOS**Conversational **Q**uality **E**stimated
MOS_{LQE}	**MOS L**istening **Q**uality **E**stimated
MOS_{LQO}	**M**ean **O**pinion **S**core-**L**istening **Q**uality **O**bjective
MP-MLQ	**M**ultipulse excitation with a maximum-**L**ikelihood **Q**uantizer
MSE	**M**ean **S**quared **E**rror
ND	**N**oise **D**isturbance

OSI	**O**pen **S**ystem **I**nterconnection
PAMS	**P**erceptual **A**nalysis **M**easurement **S**ystem
PESQ	**P**erceptual **E**valuation **S**ignal **Q**uality
PSQA	**P**seudo-**S**ubjective **Q**uality **A**ssessment
PSQM	**P**erceptual **S**peech **Q**uality **M**easure
PSQM+	**P**erceptual **S**peech **Q**uality **M**easure plus
PSTN	**P**ublic **S**witching **T**elephony **N**etwork
QoS	**Q**uality **o**f **S**ervice
RMSE	**R**oot **MSE**
RNN	**R**andom **N**eural **N**etwork
RTCP	RTP **C**ontrol **P**rotocol
RTCP-XR	**RTCP** (e**X**tended **R**eport) control protocols
RTP	**R**eal-time **T**ransport **P**rotocol
SIP	**S**ession **I**nternet **P**rotocol
SNR	**S**ignal-to-**N**oise **R**atio
SSNR	**S**egmental **SNR**
TCP/IP	**T**ransmission **C**ontrol **P**rotocol/**I**nternet **P**rotocol
TMNB	**T**ime **MNB**
UDP	**U**ser **D**atagram **P**rotocol
VAD	**V**oice **A**ctivity **D**etector
VoIP	**V**oice **o**ver **I**nternet **P**rotocol
WB_{PESQ}	$\mathbf{W}ide\mathbf{B}and_{PESQ}$

ANT COLONY OPTIMIZATION FOR VOIP QUALITY ASSESSMENT

By

Rami Suleiman AL-Khawaldeh

Supervisor

Dr. Mousa Tawfiq AL-Akhras

Co-Supervisor

Dr. Khalil Mohammed Ali El-Hindi

This Thesis was Submitted in Partial Fulfilment of the Requirements for the Master degree of Science in Computer Science

Abstract

Voice signal is traditionally carried over Telephony Networks-Public Switched Telephone Networks (PSTN) network. In order to exploit the IP network, voice traffic can be transmitted over IP network, which is originally dedicated for transmitting data, therefore IP networks can be used for carrying two types of traffic data and voice as Voice over Internet Protocol (VoIP) application. Although such networks are not designed to support real-time voice communication, because of their variable characteristics (e.g. delay, delay variation and packet loss) lead to deterioration in voice quality. A major challenge in such networks is how to measure or predict voice quality accurately and efficiently for QoS (Quality of Service) monitoring and/or control purposes to ensure that legal, technical

and commercial requirements are met.

There are two methods in order to measure the transmitted voice quality subjective and objective methods; Subjective methods depend on user for estimating the voice quality. So as speech quality is subjective as it is determined by the listener's perception, the most reliable approach for assessing speech quality is through subjective tests. Objective speech quality assessment plays an important role in recent research to circumvent the limitations of subjective testing by simulating the opinions of human testers algorithmically or using machine automation estimation. Objective quality evaluation has received considerable attention. Majority of these objective methods are based on input/output voice signal comparisons, i.e. intrusive methods which estimate the speech quality by measuring the distortion between the input and the output voice signals or non-intrusive methods which estimate the speech quality depending only on degraded or output voice signals, and mapping the distortion values to the predicted quality metric.

The foraging behavior of real Ants modeled for solving some problems that require optimal solution, so a proposed Ant Colony Optimization (ACO) technique is used in this thesis to establish a novel Non-intrusive objective method to estimate the quality of transmitted voice over Internet Protocol (IP). The ACO technique that used for this purpose was Ant-Miner with three types original Ant miner, $_{c}AntMiner$ and $_{c}AntMiner$ type 2 which look into a training data to find the optimal rules. These optimal rules represent underlying relation in the training data and will be used as model for predicting the future unknown voice quality value from a set of observed attributes values. The training data is built through capturing a set of parameters from IP network and taking the computed voice quality as class label. As a result the optimal rules can be represented as IF-THEN rules that consist of two tiles antecedent rule (IF part) and consequent rule (THEN part) where the rule's antecedent includes the parameters with their values and the rule's consequent includes voice quality as class label. This thesis accomplished two types of experiments; initial experiments that study the effect of each captured parameters on voice quality and other experiments that study all captured parameters together on voice quality. The accuracy of the proposed techniques is compared with other classification tasks such as Linear and Non-linear regression and Artificial Neural Network (ANN) technique which showed that ANN method is most correlated method in comparison with empirical values with Pearson correlation of 0.955, non-linear regression, $_{C}AntMiner$, $_{C}AntMiner$ type 2, linear regression and Ant Miner accuracy come next with correlation of 0.917, 0.875, 0.866, 0.848 and 0.470 respectively, but the Ant miners' rules use IF-THEN pattern are important as they are more understandable and more readable (plausible) for users than the regression and ANN methods.

Chapter 1

Introduction

Telecommunication industry received great attention over the last 3 decades. One of the important applications of this industry is the transmission of Voice over Internet Protocol networks which is known as VoIP. Measuring the quality of this service is the focus of this thesis.

This chapter discusses the motivations behind this thesis and emphasize the main problems to be solved in order to achieve an attractive VoIP services. This chapter is organized into 4 sections: Section 1 discusses the problem statement, section 2 illustrates the important of the problem, main contributions of the thesis are emphasized in section 3 and the organization of the remaining chapters of the thesis is shown in section 4.

1.1 Problem Statement

Nowadays voice quality plays a major role and is one of the most important aspects of the telecommunication systems. The voice should be transmitted from the sender to receiver side over communication media to guarantee satisfactory service for subscribed customers. The transmitted voice may reserve a dedicated path (channel) for transmission as in traditional telecommunication networks such as PSTN or using data-centric network such as Internet Protocol (IP) network by dividing the voice into segments called packets, in IP networks each packet could be transmitted in different paths to the receiver end which act as a collector of packets to be played orderly.

The above two transmission approaches have different effect on voice quality, as voice in PSTN network is transmitted through a dedicated path and all the resources in that path are allocated, high quality for the voice is expected. On the other hand, in case of voice transmission over IP network different paths may be used to transmit voice packets, consequently, the quality of voice traffic in PSTN network will be greater than the quality of voice in IP network. However, the disadvantages of using PSTN network are that: it requires a dedicated line to complete a call; it has bandwidth limitation; optimum usage of bandwidth is not possible and monthly fees applicable for maintenance of the network implies higher call charges, which are one of main reasons behind the switch to VoIP network, also other reasons lead to switch to VoIP network are that: to save money in comparison with PSTN network which is the most important factor; and having one information service department and one standard network for voice signal and data which saves huge administration expenses; and many new applications can be created by unifying voice and data traffic; and reconstructs the original coded voice in digital form is more accurate than reconstruct it in analogy form because of using bit stream data (voice) and digital form of voice can be cleaned to its original condition by repeaters on the packet path.

Data networks were not designed to carry real-time traffic like voice packets, therefore sending voice packets over the data network may be affected by many impairments during their trip from the sender to the receiver such as packet loss due to excessive bit errors, or congestion in the IP network, or simply excessive delay that causes the receiver to ignore the late frames in the decoding process, also voice packets may suffer delay variance (jitter) over time from point to point that occurs when network congestion, improper queuing occur and/or low bandwidth situations. Such impairments decrease the quality of voice in IP networks when compared to the quality of voice in PSTN networks. This motivated the discovery of new techniques to assess the quality of VoIP traffic as customers need to be satisfied with quality, most importantly quality measurement is needed for legal, commercial and technical reasons where the service providers evaluates their own and their competitors' service using a standard scale.

1.2 Motivations

When combining the voice and data traffic in IP network, this combination helps in reducing the disadvantage of using two separate networks, however a technique is needed to compute the quality of the transmitted voice. Computing the voice quality help in managing and monitoring the network state and solve each problem that affect the voice quality by either increasing the bandwidth capacity or upgrading to new better devices. Many efforts were going on to obtain good quality estimator, so this problem motivated researchers to propose several models to measure the quality. Some of previous attempts from researchers to estimate the transmitted VoIP quality accurately will be discussed in a separate chapter and the main contribution of this thesis which is about producing a new applicable model to assess the VoIP quality accurately will be explained thoroughly.

1.3 Main Contributions

ACO is classified as one branch of Artificial Intelligence known as swarm intelligence. ACO is used to solve many NP-complete optimization problems which are a set of NP (Non-deterministic Polynomial time) problems where any solution to the problem can be verified quickly in polynomial time complexity, also they are a set of NP-hard problems where any NP problem can be converted into NP-hard problem by a transformation of inputs in polynomial time complexity. In this thesis the use of ACO as a proposed model to assess the VoIP quality is introduced. ACO uses a number of ants to interact with each other through the problem environment as indirect interaction, so the study used the training data set as an environment to find the optimal rules that represent underlying relation in the training data. These rules will nearly approximate the future unseen examples providing some studied parameters to estimate the outcome (quality of voice in this case) as will discussed in the next chapters. ACO estimator was appeared in the year 2000 as an Ant-miner which plays a major rule in data mining classification through discovering the optimum rules. The Ant-miner used in this thesis with its variations such as Ant-miner, $_{c}Antminer$ and $_{c}Antminer$ type 2 -where the Ant-miner deals with discrete attribute data (nominal) and $_{c}Antminer$ and $_{c}Antminer$ type 2 deal with continuous attributes values- as classifiers' model for estimating the transmitted voice quality. The accuracy of these classifiers are computed and compared with two derived regression classifier, mainly Linear and Non-Linear regression and with an ANN architecture. The experiments clarifies that the ANN is more accurate than the Ant-miner variation classifiers and the two regression classifiers as will illustrated in chapter 4 that will discuss the results in detail, but using rules classifiers patterns is important because they are more understandable and more readable (plausible) for users than the regression and ANN methods.

1.4 Organization of the Thesis

In order to follow the proposed study, it is important to clarify the organization of this thesis. This thesis is organized as follows:

1. Chapter 2 Literature Review: This chapter attempts to clarify: the concept of VoIP with its main steps as a way for transmission of voice traffic; the reasons to convert to this technology; the main application of VoIP; the challenges that affect the quality of VoIP traffic and the major protocols used in VoIP. Also a short brief analysis of quality of service measurement techniques that assess the quality of voice that transmitted over IP network is given in this chapter. Finally an overview of Optimization methods such as ACO, Data mining technique such as Ant Miner, $_{c}AntMiner$ and $_{c}AntMiner$ type 2 and Artificial Neural Network that will be used for assessment or evaluation of voice quality is given.

2. Chapter 3 Quality of Service and Assessment Methods: This chapter will discuss in more details the two methods of assessments of VoIP traffic: subjective method that uses the users or customers as estimator of voice quality and objective method that uses a computational or automated algorithms that evaluate the quality of voice far way of users and laboratories standards conditions efforts.

3. Chapter 4 The Proposed Technique for Estimating The Voice Quality: This chapter aims at: proposing a new parametric and non-intrusive objective technique for assessment of voice quality; studying the effects of several parameters on voice quality (Packet loss, Burst ratio, Language, Gender and Codec) and studying the effects and accuracy of the new techniques such as Linear and Non-Linear Regression, ANN and ACO in comparison with each other. This chapter also presents the performed simulation and the conducted experiments and results. The simulation was used to produce the training set for training classifiers, The experiments studies the effects of each parameters on voice quality independently; provides some statistical analysis methods and discusses in more details the effects of all parameters together on voice quality and analyses the accuracy of each of the used six classifiers through testing set and compare each classifier with ACO especially Ant-Miner technique.

4. Chapter 5 Conclusions And Future Work: This Chapter draws conclusions and presents points for future work.

Chapter 2

Literature Review

Internet applications have been increasing rapidly in the past few years, one of them is VoIP application that carries voice packets over Internet network to exploit this network for telecommunication. This chapter attempts to clarify: the concept of VoIP with its steps as an application for transmission of voice traffic, the reasons to convert to this technology, its applications, challenges and protocols with a short brief analysis of quality of service measurement techniques that assess the quality of voice that transmitted over IP network, because of many challenges that affect the transmitted packets voice. Finally an overview of Optimization methods such as ACO, Data mining technique such as Ant Miner $_cAntMiner$ and $_cAntMinertype2$ and ANN that will be used for assessment or evaluation of voice quality is given. A reader who is familiar with any of the presented section, may skip it to save his/her precious time.

2.1 Traditional Telephony Networks-Public Switched Telephone Networks (PSTN)

Since the invention of the telephone by A. G. Bell (Collins, 2003), and the subsequent development on the technology, circuit switching has been the dominant technology for voice communications, and remained so well into digital era. Traditional PSTN uses circuit switching technology for carrying voice signals over a dedicated channel or circuit selected over the most efficient route using intelligent switches between two terminal nodes, where the necessary resources across this path are allocated to the phone call communication from the begging to the end of call, consequently the dedicated channel or circuit is unavailable until the communication released by one of terminals, the allocated resources remain unavailable even when there is no real communication between the two terminals. The advantages of PSTN technology are good respond time, availability over time, utilizing resource in efficient manner, high quality of transferred voice signals because of less effect over path and increase capacity where anyone can subscriber with this technology. The disadvantages of PSTN is dedicated line required to complete a call, limited scalability, optimum usage of bandwidth not possible, monthly fees applicable for maintenance, and higher call charges. Additionally PSTN networks are very good for transmission of voice traffic but not good for anything else like the transmission of data traffic. These disadvantages motivated the use of Voice over IP network as an alternative network for voice transmission.

2.2 Voice over Internet Protocol

VoIP is an alternative technology for PSTN with many attractive features. This section discusses the main concept behind VoIP, the main steps in VoIP transmission, VoIP applications, protocols, and challenges.

2.2.1 VoIP Concept

One alternative technology to circuit switching telephone networks for carrying voice signals is to use data-centric packet switching networks such as IP networks. In packet switching technology, there is no dedicated path between two terminals of callers, where voice signals sent as packets over IP network and each packet may take different effective route from the sender node to the receiver node.

The reasons for switching to data-centric packet switching network include saving money in comparison with PSTN network, having one information service department with one group of administrative staff and having one standard unified network for voice and data traffic which allow creation of new and advanced services. Carrying voice packets over data-centric packet switching networks such as IP network is known as VoIP. Other terms can be used interchangeably as IP Telephony or Internet Telephony. VoIP is a technology that digitizes and compresses voice conversations into IP packets for transport over a public or a private IP-based data network rather than transmit voice signals as signal waves and supports real-time, two-way transmission of conversations using IP network, hence the voice information is sent in digital form using discrete packets rather than via dedicated connections as in the PSTN. PSTN and IP networks can be connected through a network gateway that act as translator between two different networks, thus reducing the need for a separate voice and data infrastructures, VoIP offer the promise of streamlined network management and operation. However, full migration to VoIP technology will require industry-wide adoption of open standards, processes and requirements, both technical and operational in nature but the reasons to convert to VoIP network is cost reduction. From that base, VoIP is able to provide some compelling features which make switching even more attractive such as: Eliminating long distance charges, number portability, and Computer Telephony Integration (CTI).

2.2.2 Process of Voice transfer over IP network

The main steps for transferring VoIP traffic are as follows:

1. Converting of voice signal from analogue to digital form, this has the advantage that the digital form of voice can be cleaned to its original condition by repeaters on the packet path.

2. Compressing of digital voice by using some of coding technique at 8 kb/s bit rate to reduce the bandwidth that used to send packet.

3. Packetising the transmitted packet by adding IP protocol header to ensure sending of packets to receiver and determining the terminated packets host.

4. Transmission packets over IP network.

5. Depacketisating the received packets by removing the network headers to take the payload voice.

6. Decompressing that detaches the compressed digital voice to original voice.

7. Playback that sends the voice to playout devices.

These steps are depicted in Figure 2.1.

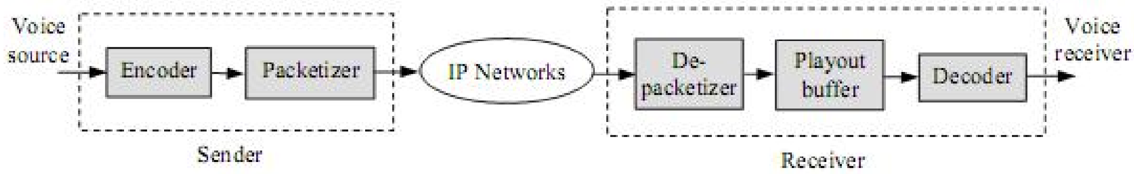

Figure 2.1: VoIP Steps (Sun,L. , 2004)

2.2.3 VoIP Applications

VoIP can be used in many applications, including (AL-Akhras, 2007):

1. Call Centre Integration: Integration of one unified network for carrying both voice and data reduces the infrastructure and administration costs. This integration may also serve as a way of integrating the Internet and IP networks with the PSTN and cellular networks all together. Such integration allows transmission of voice and data traffic over the same network, thus reducing the cost. Additionally by integrating voice and data over one network, various types of messages can be retrieved on a single device which enables applications such as an IP-based call center.

2. Directory Services over Telephones: By integrating voice and data into one network, users can search the phone book by providing the name s/he is looking for, then after the search process, the user will have the option to ring that number.

3. IP Video Conferencing: Video conferences could be held over IP networks which reduces the cost for running business. Skype is a tool that enables Skype users to speak to other Skype users for free, call traditional telephone numbers for a fee (SkypeOut), receive calls from traditional phones for a fee (SkypeIn), and receive voice mail messages for a fee. It is estimated there are more than 100 million Skype subscribers around the world.

4. Fax over IP: Facsimile services could be used over IP networks by converting the data into IP packets and sending them over an IP network which lead to great saving especially for long destinations. Fax services are more delay tolerable than voice services.

5. Radio/ TV Broadcasting: Using IP networks to carry multimedia signals enables live broadcasting of radio/TV channels; therefore the reachability of TV stations is taken to a new level.

2.2.4 VoIP Protocols

Two main telecommunication standardization organizations, International Telecommunications Union-Telecommunication Standardization Sector (ITU-T) and Internet Engineering Task Force (IETF) proposed specialized VoIP protocols to control how VoIP subscribers can interact with each other in a VoIP session. The two organizations produced many types of protocols that can be used for utilizing VoIP technique. VoIP protocols are used for initializing a IP voice call, therefore this section discusses briefly the major VoIP protocols.

To send a message from a network to another, a protocol should exist to ensure successful sending of a message. There are two standards: the Open System Interconnection (OSI) model and Transmission Control Protocol/Internet Protocol (TCP/IP) protocol suite that is the most common one usually used over current networks. OSI model was developed by the International Organization for Standardization (ISO) contains seven layers, while TCP /IP which was developed by US Department of Defense (DoD) is a protocol with five layers.

In the TCP/IP protocol there are 4 layers where the layers near the top are logically closer to the user application, while those near the bottom are logically closer to the physical transmission of the data, these layer from top to bottom are application , transport, internet and link layer, either TCP or User Datagram Protocol (UDP) are used in the transport layer. TCP offers reliability, ensures ordered delivery without loss of data, in contradiction UDP is an inherently unreliable protocol with no mechanism for ensuring ordered delivery and with no mechanism to retransmit packets in case of loss. Because of sensitivity of VoIP to delay when voice is to be carried over IP, UDP is used rather than TCP protocol for transmission. As UDP offers no mechanism for ordering packets and as packets traversing IP network could experience different delay times due to taking different routes, consequently they may arrive out-of-order. To solve this problem, a set of protocols are needed and used above UDP protocol in the protocol stack to manage this mechanism which clarified as follows (Collins, 2003; Karapantazis and Pavlidou, 2009):

1. Real-time Transport Protocol (RTP): RTP is a protocol that provides basic services and runs on top of existing transport protocols, typically UDP. RTP provides real-time applications with end-to-end delivery services such as payload type identification and delivery monitoring and provides transport of data with a notion of time to enable the receiver to reconstruct the timing information of the sender. In addition, RTP messages contain a message sequence number to allow applications to detect packet loss, packet duplication, or packet reordering. RTP message contains an RTP header followed by an RTP payload, the payload type (in the header) specifies the format of the RTP payload following the fixed header.

2. RTP Transport control protocol (RTCP): RTCP extends RTP protocol by exchanging information about an on-going session and monitors the data delivery and provides the users with some statistical functionality. The receivers can use RTCP as a feedback mechanism to notify the sender about the quality of an on-going session.

3. H.323: H.323 is an ITU-T Recommendation that defines "Packet-based multimedia communications systems". H.323 defines a distributed architecture for creating multimedia applications, including VoIP and call-setup signaling.

4. Session Internet Protocol (SIP): SIP is a protocol that establishes a calling session between two subscribers, manages a session and ends a session to finish a call.

5. Media Gateway Control Protocol (MGCP): MGCP or a similar protocol called MEGACO is a protocol that helps in creating multimedia applications by providing translation of control and multimedia services between two different networks.

2.2.5 VoIP Challenges

Although VoIP is an attractive technology, nevertheless there are several challenges that need to be tackled to achieve comparable quality to that of traditional telephony. These challenges come from the fact that IP network was not designed originally to carry voice traffic. These challenges may occur before, during or even after a VoIP session and they include (AL-Akhras, 2007):

Challenges before VoIP Session

- **Availability of Dial Tone :** Users who are familiar with circuit switching systems are used to hear a dial tone, because it gives the user the impression that the called host is ready to start the session (voice call). Circuit Switching users expect the availability of the dial tone in the VoIP networks as well.

- **Availability of Resources to Start the Call:** Some resources must be initialized to achieve acceptable quality between callers by determine routes, bandwidth and check if the call request is accept.

- **Total Amount of Time Required to Setup the Call**.

Challenges After VoIP Session

Among the challenges arise after the VoIP session is finished, is maintaining detailed call records for the purpose of billing, testing, diagnosis, network capacity planning and traffic engineering. The major issue in accounting is the selection of a suitable billing model. A number of billing models have been proposed:

- **Time-based:** The billing is metered by flow duration, time-of-day, day-of week with possible flat price regardless of the destination or the offered quality.

- **Distance-based:** The billing is based on the distance between the caller and the called users. The current IP protocol (IPv4) is not designed to support region-based IP. The newer version of IP protocol (IPv6) is designed in a manner that the user is assigned an IP address based on his/her geographical location .

- **Quality of Service-based:** The billing is based on the service quality offered.

- **Congestion-Based:** The billing depends on the congestion at the gateway which is measured as the percentage of trunks in use.

- **Hybrid Model:** A combination of two or more of the above models.

Challenges during VoIP Session

The challenges that affect a VoIP quality during an ongoing session are as follows:

- Packet Loss: During transmission of packets in IP networks, speech stream may suffer from packet loss due to excessive bit errors, or congestion in the IP network, or simply excessive delay that causes the receiver to ignore the late frames in the decoding process. In IP networks (especially in public non-managed networks), queues might overflow in network nodes between the sender and the receiver, resulting in loss of packets. Retransmission mechanisms can be applied in case of lost delay-tolerant data packets that use the reliable TCP protocol to retransmit the lost packets. These retransmission mechanisms cannot be applied to real-time applications (such as voice) because the time needed to detect the lost packets and retransmit them is long enough to the degree that lost packets become useless for decoding process.

- Delay or latency : Delay is caused when packets of data (voice) take more time than expected to reach their destination. This causes some disruption in the voice quality. There are five components that contribute to the overall delay (from the speaker's Mouth To the listener's Ear, M2E)delay:

  1. **Encoding Delay**: Which is the time spent for encoding signal which depends on the employed voice codec.
  2. **Packetization Delay**: This is the time interval to packetize the encoded voice stream.
  3. **Network Delay**: Network delay can be summarized as the sum of transmission, propagation and queuing delay.
  4. **Playback Delay**: This term refers to a playback buffer resides at receiver side to play consecutive packets smoothly.
  5. **Decoding Delay**: Decoding is the opposite operation of encoding which is the time spent to reconstruct the voice signal.

- Jitter: Jitter is one of parameters that effect the quality of the voice signal which defines as a measure of delay variance over time from point to point that occurs when network congestion, improper queuing occur and/or low bandwidth situations in VOIP. So the delay is not a constant value; rather it varies with time depending primarily on network load. The amount of jitter tolerable on the network is affected by the depth of the jitter buffer (temporary memory) on the network equipment in the voice path. The more jitter buffer available, the more the network can reduce the effects of jitter and the small buffer size causes discarding the delayed packets that leads to an audible gap.

- Ease of Use.
- Reliability (availability): The VoIP should be available most of the time.
- Scalability: The network should be able to handle large number of calls simultaneously.
- Security: The need for security is compounded through protect two invaluable assets, data and voice from illegal hacking.

2.3 Quality of Service (QoS) and Assessment

QoS is defined as a collective effect of service performance as perceived by the user (ITU Recommendation P.800.1, 1994). In VoIP networks the quality of a session is affected by the previously discussed challenges, therefore a method is needed to estimate the quality and choose the best route to achieve that best quality. Several methods have been proposed to estimate the quality: subjective and objective methods. As speech quality is a subjective as it is determined by the listener's perception, the most reliable approach for assessing speech quality is through subjective tests. The most widely used subjective test is the Absolute Category Rating (ACR) method where speech quality is estimated by averaging the opinions of a set of suitably listener's perceptions, each of the testers assign an opinion score on an integral scale of 1 (bad) to 5 (excellent) after hearing received speech signal under test. The opinion scores of the testers are averaged into a Mean Opinion Score (MOS) that provides numerical indication of the perceived quality of the degraded speech. Although MOS is a reliable technique but it is expensive, time-consuming and needs strict lab conditions (Rango et al., 2006).

Objective speech quality assessment plays an important role in recent research to circumvent the limitations of subjective testing by simulating the opinions of human testers algorithmically or using a machine automation estimation. Objective quality evaluation has received considerable attention. Majority of these objective methods are based on input/output comparisons, i.e. intrusive methods which estimate the speech quality by measuring the distortion between the input and the output signals, and mapping the distortion values to the predicted quality metric. Perceptual Evaluation of Signal Quality (PESQ) that standardized in ITU-T Recommendation P.862 (2001) is an example of such method. PESQ output can be mapped into a MOS score. Chapter 3 discusses each assessment method in details as this the core of this thesis.

2.4 Optimization Methods

2.4.1 Optimization problem

Optimization is the process of finding the optimal (minimum or maximum) output or result according to met some criteria. The optimum can be achieved by finding the best combination of input parameters. The main difficulty in optimization problems is determining if a given minimum is the best (global) minimum or a suboptimal (local) minimum according to some criteria (Haupt and Haupt, 2004).

An adequate model for optimization problem can be described as follows:

A combinatorial optimization problem

A model P = (S , Ω, f) of a combinatorial optimization problem consists of:

1. a search space S defined over a finite set of discrete decision variables X_i, $i = 1,\ldots,n$; which consists of set of components **C**.

2. a set Ω of constraints among the variables; and

3. an objective function f : $2^{\mathbf{C}} \rightarrow R^+$ to be minimized.

The generic variable X_i takes values in $D_i = v_i^1, \ldots, v_i^{|Di|}$. A feasible solution s ∈ S is a complete assignment of values to variables that satisfies all constraints in Ω. A solution s* ∈ S is called a global optimum if and only if: $f(s*) \leq f(s)\ \forall s \in S$.

2.4.2 Ant Colony Optimization techniques

In the 20^{th} century a French entomologist called Grassé (1959) discovered that some of species of termites interact with themselves and its environment in a significant stimuli which referred by them as stigmergy. Stigmergy is indirect, non-symbolic communication mediated with environment with a local information that only accessed by these insects. Deneubourg and Goss (1990) attempted to clarify this communication through the pheromone of ants which is a substance they lay on their way toward food by conducting an experiment called "double bridge experiment"; this experiment contained two paths with the same length from nest of Ants (Colony) to food, they observed that one of the paths has concentration of Ants which result of more substance of pheromone on this path. Deneubourg and Goss (1990) also conducted another experiment which is variant of "double bridge experiment" by using two paths with different lengths, this experiment illustrated that the shortest path has converges of ants.

From previous experiments and observations, it can be concluded that ants can give us the shortest path from its nest to searched food, which can be exploited for solving some problems that required optimal solution, this foraging behavior of Ants was the main source of inspiration for the development of ACO.

ACO is a paradigm that originally developed for combinatorial optimization problems as metaheuristic algorithm. it is general-purpose framework which can be applied for optimization problems (Dorigo et al., 1996; Stützle and Hoos, 2000; Dorigo and Gambardella, 1997). Real ants deposit a chemical substance called pheromone on their path between nest and colony to be marked for favorable path for other ants to follow in the future. Ant colony optimization exploits a similar mechanism for solving optimization problems.

They work blindly with dynamic memory behavior and on parallel to search near-optimal solution. These ants interact with the environment through pheromone. Artificial ants modify the amount of pheromone during their trip for finding solution. Different Ant Colony optimization algorithms have been proposed. The original one was introduced in 1990s by Marco Dorigo and colleagues (Dorigo et al., 1991), then a number of other ACO algorithms were introduced (Dorigo et al., 1996; Stützle and Hoos, 2000; Dorigo and Gambardella, 1997).

The combinatorial optimization problem defines the pheromone model of ACO, where a pheromone value is associated with each possible solution component; that is, with each possible assignment of a value to a variable. Formally, the pheromone value τ_{ij} is associated with the solution component c_{ij}, which consists of the assignment $X_i = v_i^j$. The set of all possible solution components is denoted by **C**.

Each artificial ant builds a solution through traversing across complete construction graph G(V,E) where V: is a set of vertices and E: a set of edges between these vertices where the set of possible component solution **C** can be a set of sequence of edges or vertices that have values viewed as pheromone based on some constraints, these solution component constructed incrementally that can be viewed as partial solutions constructing when each ants move on the graph from a vertex to a vertex and deposits a certain amount of pheromone $\Delta\tau$.

The ACO metaheuristic is shown in following program. After initialization, the metaheuristic iterates over 3 phases: at each iteration, a number of solutions are constructed by the ants; these solutions are then improved through a local search (this step is optional), and finally the pheromone is updated.

The following is a more detailed description of the initialization and the 3 phases.

1. Set parameters, initialize pheromone trials by initialize all parameters and pheromone values for each component that will be traversed next; initialized to constant value, that required by ants to start their search.

Program 1 Framework of the Ant Colony Optimization metaheuristic

```
Set Parameters initialize Pheromone Trials
While Termination Condition not met do
 Construct Ant Solutions
 Apply Local Search (Optional)
 Update Pheromone
End While
```

2. Construct Ant solution, each ant starts building their own solution by traversing the search space S, adding elements incrementally from a finite set of components, **C**. This construction is like walking through graph edges from vertex to vertex, where the process of selecting a component depends on the following probabilistic Equation (2.1) (Dorigo et al., 1991).

$$P_{ij}^{k} = \begin{cases} \frac{[\tau_{ij}(t)]^{\alpha}.[\eta_{ij}]^{\beta}}{\sum_{k \in allowed_k}[\tau_{ik}(t)]^{\alpha}.[\eta_{ik}]^{\beta}}, & if\, j \in allowed_k; \\ 0, & \text{otherwise.} \end{cases} \tag{2.1}$$

where

τ_{ij} is the amount of pheromone on arc i,j.

α is a parameter to control the influence of τ_{ij}.

η_{ij} is the desirability of arc i,j (a priori knowledge, typically $1/d_{ij}$),d_{ij} represents the distance between i,j.

β is a parameter to control the influence of η_{ij}.

The two parameters α and β may be viewed as positive parameters whose values determine the significant of the pheromone information and heuristic information.

This equation favors components with large pheromone values. In most ACO algorithms the probabilities for choosing the next solution component, also called the transition probabilities.

3. Apply Local Search: A local search procedure may be applied for improving the solutions constructed by the ants. The use of such procedure is optional, though experimentally it has been observed that, if available, its use improves the algorithms overall performance.

4. Update pheromones: Increases the pheromone values associated good solutions and decreases those associated with bad ones. Usually, this is achieved by decreasing all the pheromone values through pheromone evaporation, and by increasing the pheromone levels associated with a chosen set of good solutions. Different ACO algorithms such as, for example, Ant System (AS), Ant Colony System (ACS) and MAX-MIN Ant System (MMAS) mainly differ in the update of the pheromone values they apply as discussed in the following paragraphs.

The Main Variants of ACO

1. Ant System

   The original algorithm in ACO (Dorigo et al., 1996) where each ant builds a solution update the pheromone values where the pheromone τ_{ij}, associated with the edge joining vertex i and j, is updated as follows:

$$\tau_{ij} \leftarrow (1-\rho).\tau_{ij} + \sum_{k=1}^{m} \Delta\tau_{ij}^{k} \tag{2.2}$$

where

ρ is the evaporation rate.

m is the number of ants.

$\Delta\tau_{ij}^{k}$ is the amount of pheromone laid on edge (i, j) by ant k.

$$\Delta\tau_{ij}^{k} = \begin{cases} Q/L_k, & \text{if ant k used edge (i,j) in its tour;} \\ 0, & \text{otherwise.} \end{cases} \tag{2.3}$$

where

Q is a constant.

L_k is the length of the tour constructed by ant k.

2. MAX-MIN Ant System (MMAS)

   Another improvement, proposed by Stützle and Hoos (2000), over the original ant system idea. Its characterizing elements are that only the best ant updates the pheromone trails and that the value of the pheromone is bound. The pheromone update is implemented as follows:

$$\tau_{ij} \leftarrow [(1-\rho).\tau_{ij} + \Delta\tau_{ij}^{best}]_{\tau_{min}}^{\tau_{max}} \tag{2.4}$$

   where τ_{max} and τ_{min} are respectively the upper and lower bounds imposed on the pheromone; the operator $[x]_a^b$ is defined as

$$[x]_a^b = \begin{cases} a, & \text{x>a;} \\ b, & \text{x<b;} \\ x, & \text{otherwise.} \end{cases} \tag{2.5}$$

   where $\Delta\tau_{ij}^{best}$ is

$$\Delta\tau_{ij}^{best} = \begin{cases} 1/L_{best}, & \text{if (i, j) belongs to the best tour;} \\ 0, & \text{otherwise.} \end{cases} \tag{2.6}$$

3. Ant Colony System

   The first major improvement over the original ant system was Ant Colony System (ACS), introduced by Dorigo and Gambardella (1997). The first important difference between ACS and AS is the form of the decision rule used by the ants during the construction process. Ants in ACS use the so-called pseudo random proportional rule: the probability for an ant to move from vertex i to vertex j depends on a random variable q uniformly distributed over [0,1] , and a parameter q_0; if $q < q_0$, then, among the feasible components, the component that maximizes the product $\tau_{il}.\eta_{il}^{\beta}$ is chosen, otherwise the same equation as in AS is used. This rather greedy rule, which favors exploitation of the pheromone information, is counterbalanced by the introduction of a diversifying component: the local pheromone update. The local pheromone update is performed by all ants after each construction step. Each ant applies it only to the last edge traversed using the following formula (2.7):

$$\tau_{ij} = (1-\varphi).\tau_{ij} + \varphi.\tau_0 \tag{2.7}$$

where

$\varphi \in$ (0,1] is the pheromone decay coefficient (The evaporation rate).

τ_0 is the initial value of the pheromone.

The main goal of the local update is to diversify the search performed by subsequent ants during one iteration. In fact, decreasing the pheromone concentration on the edges as they are traversed during one iteration encourages subsequent ants to choose other edges and hence to produce different solutions. This makes it less likely that several ants produce identical solutions during one iteration. Additionally, because of the local pheromone update in ACS, the minimum values of the pheromone are limited. As in AS, the ACS performs at the end of the construction process a pheromone update, called offline pheromone update. ACS offline pheromone update is performed only by the best ant, that is, only edges that were visited by the best ant are updated, according to the equation (2.8):

$$\tau_{ij} \leftarrow .(1-\rho).\tau_{ij} + \rho.\tau_{ij}^{best} \tag{2.8}$$

where $\tau_{ij}^{best} = 1/L_{best}$ if the best ant used edge (i,j) in its tour, $\tau_{ij}^{best} = 0$ otherwise (L_{best} can be set to either the length of the best tour found in the current iteration

iteration-best, L_{ib} or the best solution found since the start of the algorithm, best-so-far, L_{bs}). It should be noted that most of the innovations introduced by ACS were introduced first in Ant-Q, a preliminary version of ACS by the same authors.

2.5 Data Mining

Nowadays, databases are rich with hidden patterns that play an important role in intelligent decisions making. Consequently a discipline called Data Mining has emerged to be used to extract or discover knowledge accurately and comprehensibly to aid the users to make reliable decisions (Witten and Frank, 2005).

To extract or discover knowledge, one of data analysis that used by data mining is classification which is a data mining technique used to find a model (or function) that describes and distinguishes data classes or concepts for the purpose of predicting the class of objects whose label is unknown. The derived model is based on the analysis of a set of training data, which is a database that consists of a set of attributes one of them is referred to as class label. The derived model can be presented in various forms such as classification (IF-THEN) rules, Decision trees, Mathematical formulae, or ANN. Although each data mining technique has a bias (i.e. each task in the data mining can be regarded as a kind of problem to be solved by a data mining algorithm). The classification (IF-THEN) rules has received an important attention, because of plausibility to the human mind more than other classification techniques. This section discusses Ant miner variants. Ant miner is an ACO data mining technique represented as (IF-THEN) rules which is used in this thesis to estimate the voice quality that transmitted over IP network.

The classification (IF-THEN) rule formed as follows:

IF $< conditions >$ THEN $< Class >$ where

Conditions refer to antecedent part of the rule (IF part) that consists of a set of terms each of them act as a tuple $< attribute, operator, value >$, value belongs to the domain of attribute, hence terms are connected by a logical operator usually (AND) operator.

Class refers to the rule consequent (THEN part); it defines the classifier whose predictor attribute satisfies all terms specified in the rule antecedent.

2.5.1 Ant Miner

Parpinelli et al. (2002) proposed an ACO algorithm for the classification task of data mining, called Ant-Miner (Ant-Colony-based data miner) that extracts classification rules from data. The method uses a sequential covering algorithm to discover the classification rules, it uses ACO to find the best antecedents (conditions) for each rules. This method was inspired by the behavior of real ants colonies and some of data mining principles and concepts. The process of Ant-Miner is illustrated as pseudocode in program 2.

At first, the list of discovered rules is empty and the training set consists of all the training cases. Each iteration of the WHILE loop discovers one classification rule. This rule is added to the list of discovered rules and the training cases that are covered correctly by this rule (i.e., cases satisfying the rule antecedent and having the class predicted by the rule consequent) are removed from the training set. This process is performed iteratively while the number of uncovered training cases is greater than a user-specified threshold, called Max_uncovered_cases. Each iteration of the REPEAT-UNTIL loop of Ant-Miner algorithm consists of four steps, comprising initialization pheromone, rule construction, rule pruning, and pheromone updating, detailed as follows:

1. **Pheromone initialization:** Initialize all trial with the same amount of pheromone; each term has an initial pheromone value as follows:

$$\tau_{ij}(t=0) = \frac{1}{\sum_i^a b_i} \tag{2.9}$$

Where

a is the sum of attributes.

b is the sum of values in the domain of attribute a.

Program 2 A High-Level Description of Ant-Miner (Parpinelli et al., 2002)

```
Training set={all training cases};
DiscoveredRuleList=[ ]; /*rule list is initialized with an emptylist*/
WHILE( TrainingSet > Max_uncovered_cases)
 t=1; /* ant index */
 j=1; /* Convergence test index */
 REPEAT
  Initialize all trials with the same amount of pheromone;
  Ant_t start wit an empty rule and incrementally constructs
  a classification rule R_t by adding one term at a time to the
  Current rule;
  Prune rule R_t ;
  update the pheromone of all trials by increasing pheromone
  In the trial followed by Ant_t (proportional to the quality of R_t)
  and decreasing pheromone in the other trials ( Simulating
  pheromone evaporation );
  IF( Rt is equal to R_t-1 ) THEN /* update convergence test */
    j = j + 1;
  ELSE
   j=1;
  END IF
  t = t + 1 ;
 UNTIL (t > No_of_ants) OR (j > No_rules_converg)
 Choose the best rule R_best  among all rules Rt constructed by all the ants;
 Add rule  R_best to DiscoveredRuleList;
 TrainingSet = TrainingSet - {set of cases correctly covered by R_best};
END WHILE
```

2. **Rule Construction:** Each ant starts with an empty rule and terms are added by sequentially one term at a time according to a probability distribution that takes into account the pheromone and the heuristic values of each term ; The probability distribution is defined as follows:

$$P_{ij}(t) = \frac{\eta_{ij}.\tau_{ij}(t)}{\sum_{i=1}^{a} x_i.\sum_{j=1}^{b_i}(\eta_{ij}.\tau_{ij}(t))}, \tag{2.10}$$

where

η_{ij}: is a value of problem-dependent heuristic functions for $term_{ij}$, its value determines the ants choices, this value depends on information theory; that is the value of η_{ij} for $term_{ij}$ involves a measure of the Entropy (or a mount of information) associated with that term.

The entropy of each term is computed as follows:

$$H(W \backslash A_i = V_{ij}) = -\sum_{w=1}^{k}(p(w \backslash A_i = V_{ij}).\log_2 p(A_i = V_{ij})), \tag{2.11}$$

Where

$term_{ij} \equiv A_i = V_{ij}$.

w: is the class attribute whose domain consists of the classes to be predicted.

$P(A_i = V_{ij})$: is the conditional probability of observing class w, having observed $A_i = V_{ij}$.

The uniformly distributed classes and the smaller probability that the current ant chooses to add $term_{ij}$ to it is partial rule depends on the higher value of $H(W\ A_i = V_{ij})$. Finally the heuristic for preferring those terms can be computed as the following equation (2.12):

$$\eta_{ij} = \frac{\log_2(k) - H(W \backslash A_i = V_{ij})}{\sum_{i}^{a} x_i.\sum_{j=1}^{b_i}(\log_2(k) - H(W \backslash A_i = V_{ij})}, \tag{2.12}$$

τ_{ij}: amount of pheromone associated with $term_{ij}$ at iteration t; corresponding to the amount of pheromone currently available in position ij of a path being followed by the current ant. The better the quality of the rule constructed by an ant, the higher the amount of pheromone added to the trial segments visited by the ant, therefore as time goes on, the best trial segments will be added to the rule.

a : total amount of attributes.

x_i : set to one if the attribute A_i has not yet used by the current ant or zero, otherwise.

$term_{ij}$: is chosen to be added to the current partial rule with probability proportional to the value of p_{ij}.

3. **Rule Pruning:** After constructing a rule, the rule pruning process begins to increase the comprehensibility and accuracy of the constructed rule. It attempts to determine and remove the term that will cause a maximum increase in the quality of the rule according to the following formulae:

$$Q = Sensitivity \times specificty \tag{2.13}$$

Also defined as:

$$Q = \frac{TP}{TP+FN} \cdot \frac{TN}{FP+TN} \tag{2.14}$$

Where

TP: True Positive, the number of cases covered by the rule that have the class predicted by the rule.

FP: False Positive, the number of cases that are not covered by the rule but that have the class predicted by the rule.

TN: True Negative, the number of cases that are not covered by the rule and that do not have the class predicted by the rule.

So the Q's value is within the range $0 \leq Q \leq 1$ and the larger the value of Q, the higher the quality of the rule.

4. **Pheromone Update Rule:** The Pheromone update is carried out after the rule construction step as follows:

$$\tau_{ij}(t+1) = \tau_{ij}(t) + \tau_{ij}(t).Q, \; \forall \, term_{ij} \in rule \tag{2.15}$$

Finally pheromone evaporation associated with each $term_{ij}$ which does not occur in the constructed rule should be decreased by dividing the value of each τ_{ij} by the summation of all τ_{ij}.

2.5.2 Continuous Ant Miner

As Ant-Miner does not cope well with continuous attributes directly, one may use a discrete technique in a preprocessing step. An extension to Ant-Miner called *c*Ant-Miner was proposed by Otero et al. (2008) which incorporates an entropy-based discretization method in order to cope with continuous attributes during the rule construction process dynamically and consequently to avoid the need for running a discretization method in a preprocessing step. The algorithm is applied as follows: First of all, an extension was added to the original Ant-Miner to support continuous attributes in the rule antecedent taking the form of ($attribute_c < value$) or ($attribute_c \geq value$), where value is a value belonging to the domain of the continuous attribute $attribute_c$; Incorporate an entropy-based discretization method into Ant-Miner's rule construction process to dynamically create thresholds on continuous attributes domain values. In order to compute the entropy of a continuous attribute, a threshold value v is selected to dynamically partition the continuous attribute a_i into two intervals: $a_i < v$ and $a_i \geq v$. The best threshold value is the value v that minimizes the entropy of the partition, given by:

$$ep_v(a_i) = \frac{|S_{ai} < v|}{|S|}.entropy(a_i < v) + \frac{|S_{ai} \geq v|}{|S|}.entropy(a_i.v) \tag{2.16}$$

Where

$|S_{ai} < v|$ is the total number of examples in the partition $a_i < v$ (partition of training examples where the attribute a_i has a value less than v).

$|Sai \geq v|$ is the total number of examples in the partition $a_i \geq v$ (partition of training examples where the attribute a_i has a value greater or equal to v).

$|S|$ is the total number of training examples.

After the selection of the threshold v_{best}, the entropy of the continuous attribute corresponds to the minimum entropy value of the two partitions is computed as:

$$entropy(continuousattribute) = min(entropy(a_i < v_{best}), entropy(a_i.v_{best})) \tag{2.17}$$

The selection of the lowest entropy value corresponds to select of the "purist" partition (the partition with more examples belonging to the same class) and it represents the expected predictive power (quality) of the continuous attribute (when continuous attribute is added to the rule). It should be noted that the entropy of every continuous attribute, i.e. every term having a continuous attribute, is always the same as the entropy value of every discrete attribute; every term representing an attribute-value pair of a nominal attribute. Therefore, the entropy of all attributes are computed as a preprocessing step to save computational time. In the rule construction, when an ant chooses a node that represents a continuous attribute a_i to add to its current partial rule, a relational operator and a threshold value are selected as follows. First, the best threshold value for attribute a_i is selected as in Equation (2.16) then after selecting the threshold value v_{best}, a relational operator

op is selected based on the entropy values of the two partitions generated. If the partition $a_i < v_{best}$ has a lower entropy, then the operator $<$ (less-than operator) is selected. If the partition $a_i \geq v_{best}$ has a lower entropy, then the operator $\geq$ (greater-than or equal operator) is selected. The operator selection has a bias of selecting the "purist" partition, given that lower entropy values are favored over higher entropy values.
Once the threshold value v_{best} and the operator op are selected, a term in the form of a triple (a_i, op, v_{best}) is added to the ant's current partial rule and the rule continues to undergo the Ant-Miner's rule construction process.

2.6 Artificial Neural Network

The foundation of neural networks in a scientific sense begins with biological nervous systems. The human brain consists of an estimated 10 billion neurons (nerve cells) and 6000 times as many synapses (connections) between them (Haykin, 1994). All information taken in by a human is processed and assessed in this particular part of the body. ANN is an information processing paradigm that consists of a set of highly interconnected processing elements called neurons working in a coordinate manner to solve specific problems as depicted in Figure 2.2. Therefore both biological and artificial neurons are elementary information processing units and fundamental building blocks of a neural network (Grothmann, 2002).

As in biological systems, ANN learning involves adjustments to the weights connections that exist between the neurons (Nygren, 2004) where Figure 2.2 illustrate that the connections (synapses) w_{ij} transfer the signals (stimulus) x_i into the neuron; w_{ij} can be interpreted as a weight representing the importance of that specific input x_i. Inside the neuron the sum of the weighted inputs $w_{ij} * x_i$ is taken as a transfer function. Given that this sum net_j is greater than an externally applied threshold, the neuron emits an output o_j, o_j is either continuous or binary valued, depending on the activation function (or squashing function).

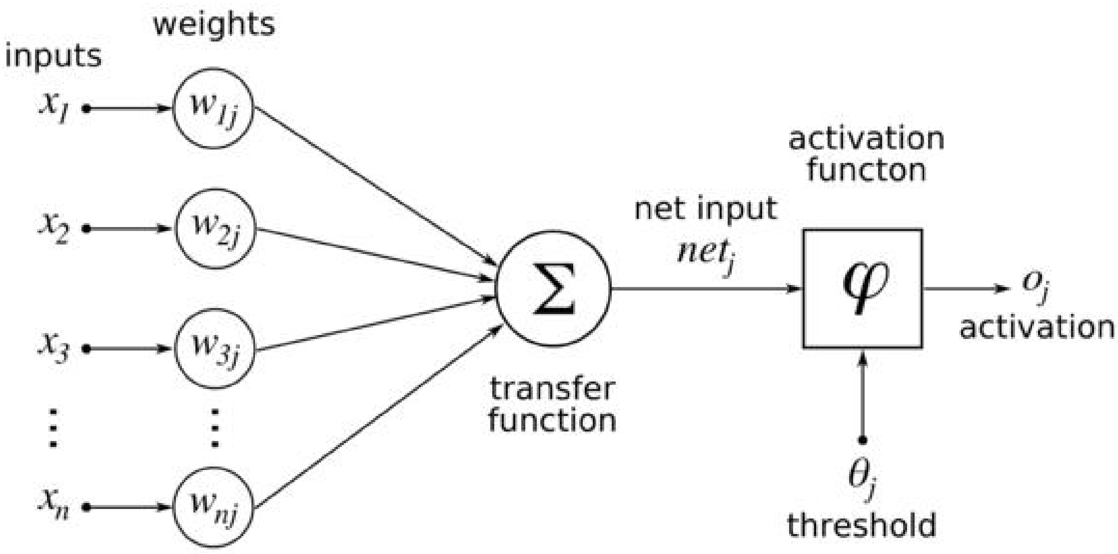

Figure 2.2: The Structure of Neuron Model

The principal strength with the network is its ability to find patterns and irregularities as well as detecting multi-dimensional non-linear connections in data (Fausett, 1994; Aleksander and Morton, 1990) which is a function approximation that predict the output according to input data. ANN may contain complex structure with an input layer, set of hidden layers and an output layer where each layer consists of a set of neurons that works together to achieve a specific goal.

Chapter 3

Quality of Service and Assessment Methods

Sending voice packets over the IP network suffers from many challenges that affect packets during their trip from the sender to the receiver which decrease the quality of VoIP traffic, as such it is necessary to assess the quality of these service to make sure of user's satisfaction.

QoS is a term used to refer to the capacity of the network to provide the level of service expected/agreed by the customer or the user (Rango et al., 2006). Traditional PSTN provide high voice quality than packet-switched network such as VoIP network which motivated researchers to proposed new techniques to assess the quality in VoIP networks for legal, commercial and technical reasons; measuring the quality also helps VoIP service providers to evaluate their own and their competitors' service using a standard scale and to achieve customers' satisfaction. Quality assessment techniques are divided into two categories depending on the role of the user in the assessment processes, these categories are: subjective and objective methods where in subjective methods the users estimate the voice quality and in the objective either computational or automated algorithms that evaluate the quality of voice far way of users and laboratories standards conditions efforts are employed. Figure 3.1 depicts different classifications. The following sections will discuss in more details the two methods of assessments, readers who are familiar with any of the presented sections, may skip it.

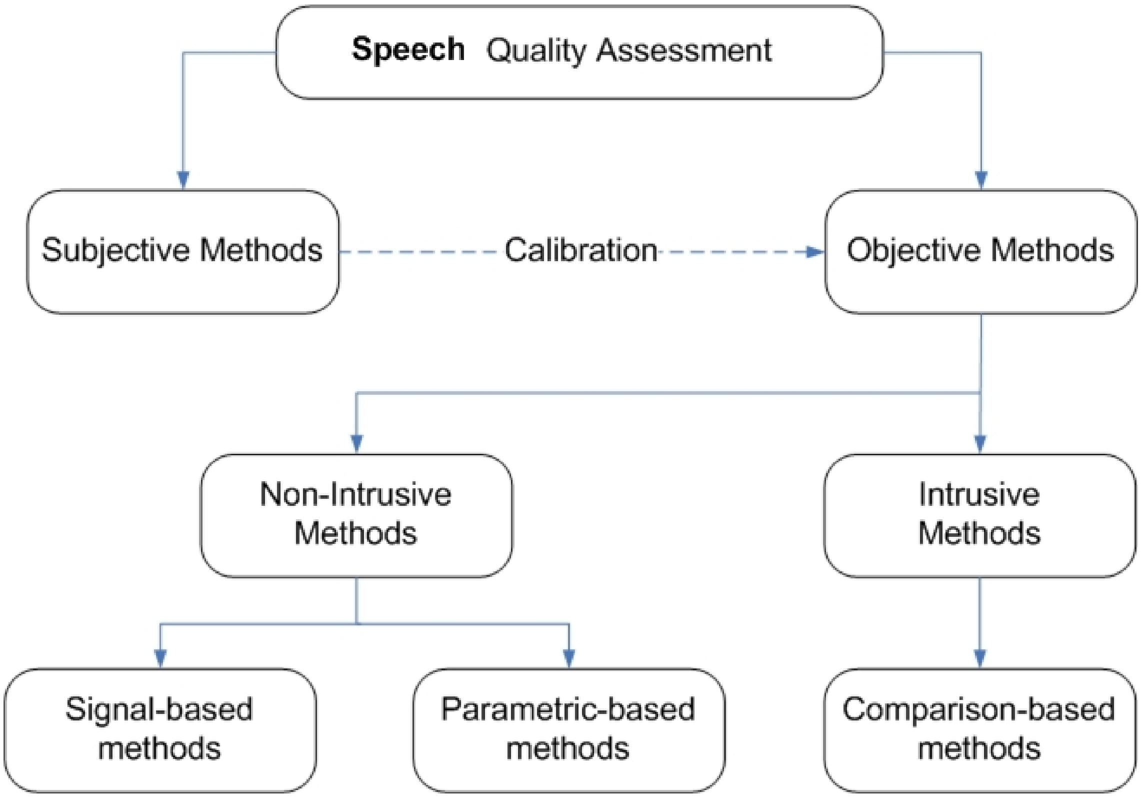

Figure 3.1: Overview of VoIP measurement methods (Sun,L. , 2004)

.

3.1 Subjective assessment methods

Subjective assessment methods need a panel of users to perform the subjective tests of voice quality (test subjects) by listening to a set of recorded or live conversational speech samples in different network conditions under special laboratory settings. Each subject gives his\her opinion on voice quality sample by assigning a score value to each sample ranging from 1 to refer to bad quality up to 5 to refer to excellent quality, these scores will result with metric that called Mean Opinion Score (MOS) that is calculated by averaging the scores obtained from the panel of users for a given sample, the dominant method is ACR (Absolute Category Rating) that will be described shortly. Subjective assessment methods seemed to be more accurate, useful and reliable in measuring perceived voice quality than objective methods, this is because it reflects the opinion of ultimate receivers of voice service quality and they are crucial to be a benchmarks for objective methods, nevertheless it has several disadvantages such as it needs to special laboratory conditions for testing, it is expensive to perform the test because it needs many materials and resources to perform this operation, cost time to performs the test (time consuming) and finally it is not suitable in real-time or online applications as in non-managed networks such as Internet. International Telecommunication Union-Telecommunication standardization sector ITU-T Recommendation P.800 (1996) defines many techniques as a standards for subjective quality assessment in two-way conversional test and in one-way listening only test and a strict lab conditions, such conditions concerns the room size, noise level, and the use of sound-proof cabinet in a room with a volume not less than 20 m^3. In source recording the test room must be a volume of 30 m^3 up to 120 m^3, with an echo duration lower than 500 ms (200-300 ms is preferred) and a background noise lower than 30 decibel (dB). All receiving systems, local phones or an Intermediate Reference Systems (IRS) must be calibrated and the sensibility test of the system must be performed at the beginning and at the end of test. The recording system must be of high quality (2 ways tape, 2 digital channels audio microprocessor or computer-driven recording system) and recorded voice signals must consist of few short phrases, taken from papers or non-technical lectures, casually ordered (phrases of 3-6 seconds of length or conversations of 2-5 minutes of length), all the used material must be recorded with a microphone at a distance of 140-200 mm from the speakers mouth. Also the sound pressure level should be measured from a vertical position above the subjects seat while the furniture in place and take into account other conditions regarding the subjects who participate in the test such as they have not been directly involved in work connected with assessment of the performance of telephone circuits, or related work such as speech coding, also they have not participated in any subjective test whatever for at least the previous six months, and not in a conversational/listening test for at least one year. Also in case of listening-test they have never heard the same sentence lists before, the next subsection will illustrate two techniques of subjective assessment methods briefly.

3.1.1 Conversional test

Conversational opinion tests are laboratory or two-way tests that aim to reproduce the real conditions experienced by customers. It is important that the simulating conditions are correctly specified, reproduced and accurately measured before and after experiments, Two subjects who perform the test have to listen and talk interactively via the transmission system under test, where they are placed in separated and isolated rooms to report their opinion on the opinion scale recommended by ITU-T and the arithmetic mean of these opinions are calculated where the result called Mean Conversational Opinion Score (MOS_c).

3.1.2 Listening test

Listening only test is a one-way test in which a group of subjects listen to a conversation or prerecorded phrases and give their opinions to be averaged to produce Mean Opinion Score-Listening Quality (MOS_{LQ}), where this method lakes realistic conversations, the goal of this test is the evaluation of single performances of terminals and algorithms under different conditions; some of the well known listening tests are: ACR, DCR and CCR.

1. Absolute Category Rating (ACR)

   Absolute Category Rating is the most common listening quality subjective method where the two subjects rate their opinions with five-point scale as shown in Table 3.1 on signal samples that are heard, these scales or scores will gathered and averaged to yield a final score referred to as the mean opinion score, or MOS as commonly known.

2. Degradation Category Rating (DCR)

Table 3.1: Grades in the MOS scale. Listeners express their opinion on the quality of the perceived speech signal (no reference presented) (ITU-T Recommendation P.800, 1996)

Category	Speech Quality
Excellent	5
Good	4
Fair	3
Poor	2
Bad	1

ACR tends to lead to low sensitivity in distinguishing among good quality circuits. A modified version of the ACR procedure, called the Degradation Category Rating (DCR) procedure, affords higher sensitivity and used with high-quality voice samples where ACR results are inappropriate to discover quality variations where the reference sample will perform a good evidence about the quality of degraded sample. In this listening test two samples (A and B) are present: A represents the reference sample with the reference quality, while B represents the degraded sample; listeners must compare the B sample on a degradation scale according to his\her acoustic model and the results (opinions) are averaged as Degraded MOS (Degraded Mean Opinion Score). Each configuration is evaluated by almost 4 talkers; the samples must be composed of two periods, separated by silence (for example 0.5 seconds), firstly sample A then sample B. DCR is different from ACR for the types of used samples; at this point, subjects use a degradation scale, composed of five points to indicate the degradation of sample B, referring to sample A as degradation shown in the following Table 3.2.

3. Comparison Category Rating (CCR)

The CCR method is similar to the DCR method. Listeners are presented with a pair of speech samples on each trial (A and B) sample. In the DCR procedure, a reference (unprocessed) sample is presented first (A) sample, followed by the same

Table 3.2: Grades in the DMOS scale. Listeners are asked to describe degradation in the second signal in relation to the first signal (ITU-T Recommendation P.800, 1996)

Degradation level	Score
inaudible	5
audible but not annoying	4
slightly annoying	3
annoying	2
very annoying	1

speech sample (B), which has been processed by some technique. In the DCR method, listeners always rate the amount by which the processed (B) sample is degraded relative to the unprocessed (A) sample. In the CCR procedure, the order of the processed and unprocessed samples is chosen at random for each trial. On half of the trials, the unprocessed sample is followed by the processed sample. On the remaining trials, the order is reversed. Listeners use the scale shown in Table 3.3 to judge the quality of the second sample relative to that of the first. In this technique listeners provide two judgements with one response where the advantage of the CCR method over the DCR procedure is the possibility to assess speech processing that either degrades or improves the quality of the speech. The quantity evaluated from the scores (Comparison Mean Opinion Score) is represented by the symbol CMOS.

3.2 Objective assessment methods

As subjective speech quality tests are hard-to-conduct, consequently objective tests received great interest from researchers and engineers for a long time, who have attempted to evolve and develop new algorithmic, mathematical and computational machine models to automatically evaluate the transmitted speech quality over IP-based network that replace the human panel, where the aim is to predict MOS values that are as close as

Table 3.3: Grades in the CMOS scale. Listeners grade the perceived quality of a speech signal in relation to a reference speech signal (ITU-T Recommendation P.800, 1996)

Much Better	3
Better	2
Slightly Better	1
About the Same	0
Slightly Worse	-1
Worse	-2
Much Worse	-3

possible to the rating obtained from subjective test and to make up or circumvent the limitations of subjective assessment methods, since subjective assessment methods are the most accurate, therefore the accuracy, effectiveness and performance evaluation of objective methods are determined by their correlation with the subjective MOS scores, usually using the Pearson (linear) correlation shown in Equation (3.1), if it has a high correlation, typically greater than 0.8 it is deemed to be effective measure, but sometimes their correlation can be low because of network parameters.

$$r = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2 \sum(y_i - \bar{y})^2}}, \tag{3.1}$$

where x_i is the subjective condition MOS for condition i, and $\bar{x}$ is the average over the subjective condition MOS values, x_i.
y_i is the mapped condition averaged-score of a given objective model for condition i, and $\bar{y}$ is the average over the condition-averaged values y_i.

Objective speech quality assessment methods base their measures on objective metrics of physical parameters and properties of the speech signal, different objective measures based on various speech analysis models are available. Objective measures can be classified into two classes: intrusive and non-intrusive as shown in Figure (3.2), where intrusive measures, often referred to as input-to-output measures, base their measurement on comparing the original (clean or input) speech signal with the degraded (distorted or output) speech signal which are comparison-based methods, on the other hand, non-intrusive measures (also known as output-based or single-ended or passive measures) use only the degraded signal without access to the original signal, non-intrusive methods are also divides into to subcategories, signal-based and parametric-based methods. In the following subsections different categories will be explained in more detailed.

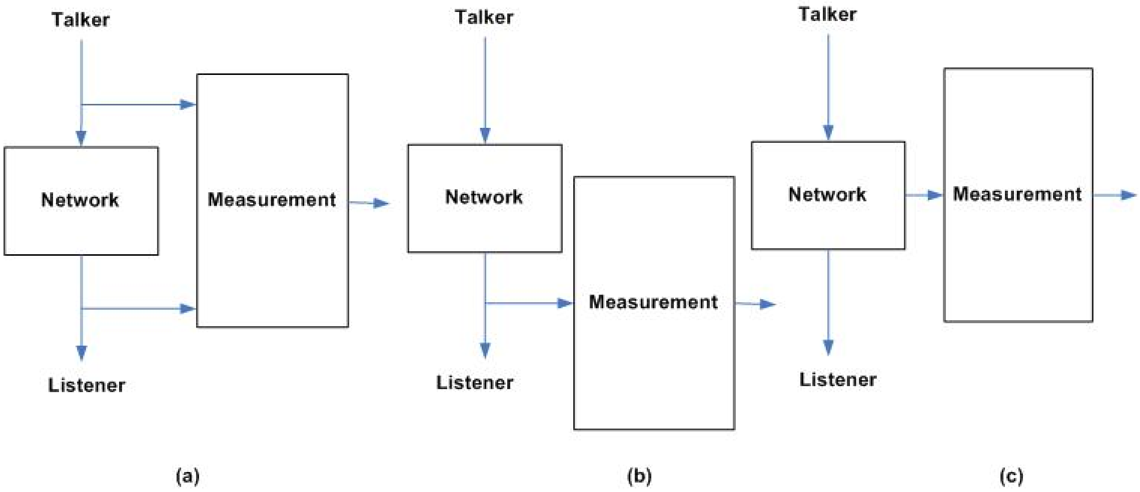

Figure 3.2: Three main categories of objective quality measurement: (a) Comparison-based intrusive method, (b) Signal-based non-intrusive method, (c) Parametric-based non-intrusive method (Sun,L. , 2004)

3.2.1 Intrusive methods

Intrusive objective-based methods or comparison-based, or full-reference model need the existed original signal in sending side to be reference for comparison with the presence of degraded, output or received signals as reconstructed by the decoder in receiving side to compute the MOS measure that expected to be as close as possible to MOS of subjective methods. Despite of existing many different types of the intrusive (or input-to-output) objective speech quality measures all share the same measurement structure which consists of two main process, as shown in Figure 3.3 where the first process involves pre-processing of the speech signals and extraction of the relevant speech parameters; i.e.

both the original (input) speech signal and the signal degraded by the system under test (output signal) are transformed into a specific domain such as temporal (timeable), spectral or perceptual domain, the second process involves a distance measure that computes the appropriate quantitative measure.

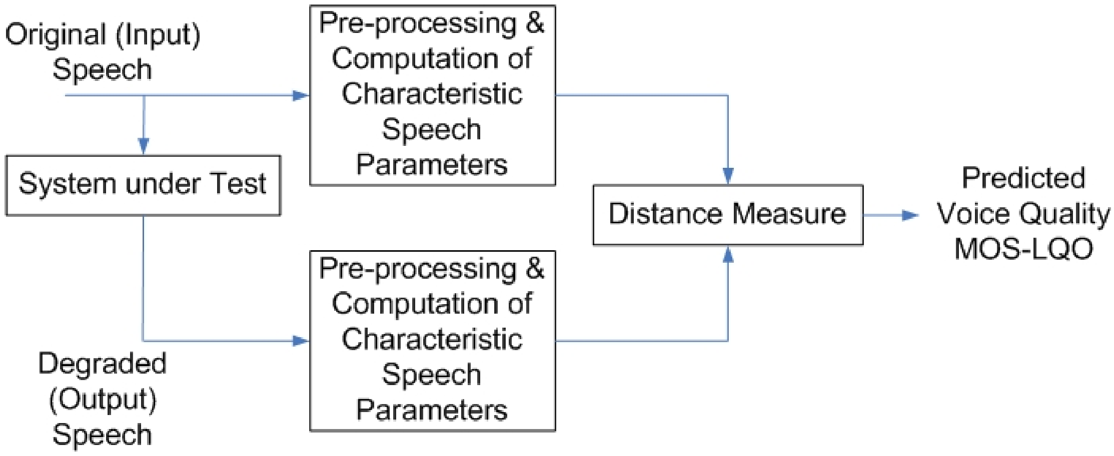

Figure 3.3: Basic Structure of an intrusive (input-to-output) objective voice quality measure (Mahdi and Picoviciv, 2009)

Objective intrusive measures can be classified into 3 kinds of transformation domain as shown in Figure 3.4.

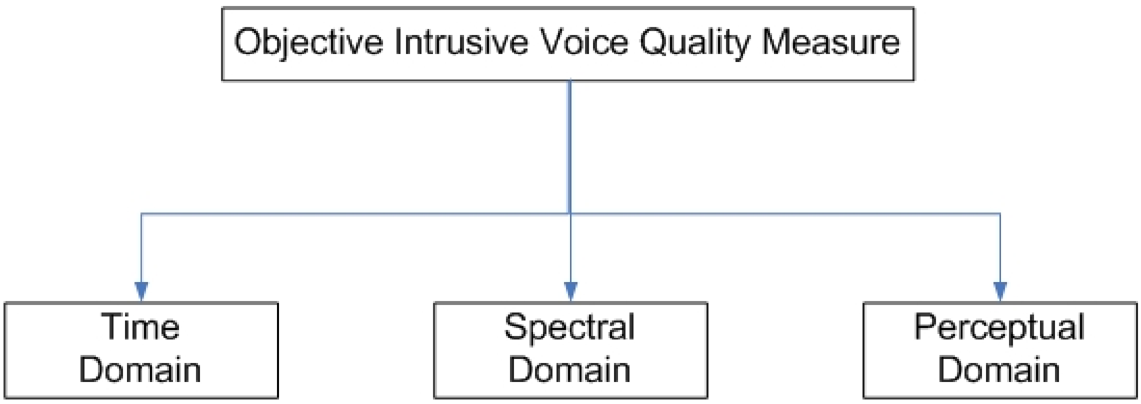

Figure 3.4: Classification of objective intrusive voice quality measures based on the transformation domain (Mahdi and Picoviciv, 2009)

- **Time domain measures**

Time domain measures is the simplest class of intrusive objective quality measures consists of an analogue or waveform-comparison algorithms in which the target is to reproduce a copy of input waveform such that the original and distorted signals can be time aligned and noise can be accurately calculated, the most important methods of this category are the Signal-to-Noise Ratio (SNR) and Segmental SNR (SSNR) (Quackenbush et al., 1988). Signal refers to useful information conveyed by some communications medium, and noise to anything else on that medium (Mahdi and Picoviciv, 2009). Classical SNR is computed as follows:

$$SNR = 10\log_{10}\frac{\sum_n x^2(n)}{\sum_n (x(n)-d(n))^2}, \tag{3.2}$$

where x(n) represents the original (undistorted) speech signal, d(n) represents the distorted speech reproduced by a speech processing system and n is the sample index (determined points on time domains).

Segmental signal-to-noise (SSNR), on the other hand, represents one of the most popular classes of the time-domain measures. The measure is defined as dividing two signals into smaller segments and calculating SNR value for each of these segments where the result is an average of the SNR values of segments, and can commonly be computed as follows:

$$SSNR = \frac{10}{M}\sum_{m=0}^{M-1}\log_{10}\left(\frac{\sum_{n=Nm}^{Nm+N-1} x^2(n)}{\sum_{n=Nm}^{Nm+N-1}[d(n)-x(n)]^2}\right), \tag{3.3}$$

where x(n) represents the original speech signal, d(n) represents the distorted speech signal, n is the sample index, N is the segment length, and M is the number of segments in the speech signal. Classical windowing techniques are used to segment the speech signal into appropriate speech segments. Performance measure in terms of SSNR is a good estimator of voice quality of waveform codecs, although its performance is poor for vocoders where the aim is to generate the same speech sound rather than to produce the speech waveform itself.

These two algorithms are easy to implement, have low computational complexity, and can provide indications of perceived speech quality for a specific waveform-preserving speech system. Unfortunately, when used to evaluate coding and transmission systems in a more-general context, SNR and SSNR show little correlation to perceived speech quality, these measures are also sensitive to a time shift, and therefore require precise signal alignment such that to achieve the correct time alignment it may be necessary to correct phase errors in the distorted signal or to interpolate between samples in a sampled data system. The time domain measure needs an accurate synchronization between reproduced waveform and original signals even the results obtained by these measures reflect the distortions introduced by the system under test. These measures are of little use nowadays because current Low-Bit-Rate (LBR) speech coders use parametric model to approximate short segments of the speech by estimating a set of source model parameters for each segment and converting them into bit stream, in opposite of conventional waveform speech coders which attempt to produce a reconstructed signal whose waveform is as close as possible to the original speech.

- **Spectral domain measures**

  Spectral domain measures or frequency-domain measures are known to be significantly better correlated with human perception, but still relatively simple to implement. One of their critical advantages is that they are less sensitive to signal misalignment and phase shift between the original and the distorted signals than time domain measures. Most spectral domain measures are related to speech codecs design and use the parameters of speech production models. Their capability to effectively describe the listeners auditory response is limited by the constraints of the speech production models. Some of the most popular frequency domain techniques are the ItakuraSaito (IS) (Itakura and Saito, 1978), the Cepstral Distance (CD) (Kitawaki et al., 1988) and the Log-Likelihood (LL) (Itakura, 1975).

- **Perceptual domain measures**

  Many perceptual domain measures are based on mimicry of the human auditory system, they have shown to be a highly accurate objective performance measures because many modern codecs are non-linear and non-stationary making the shortcomings of the previous objective measures even more evident. In these measures, speech signals are transformed into a perception-based domain that approximates human perception or simulate the psychophysics of hearing, such as the critical-band spectral resolution, frequency selectivity, the equal-loudness curve and the intensity-loudness power law to derive an estimate of the auditory spectrum, then the relevant information or features that are sufficient and necessary for accurate assessment of perceived speech quality are extracted, where the cognition/judge model can map the difference between original (reference) and distorted (degraded) signals into estimated perceptual distortion or further to Mean Opinion Score (MOS) scale as illustrated in Figure 3.5.

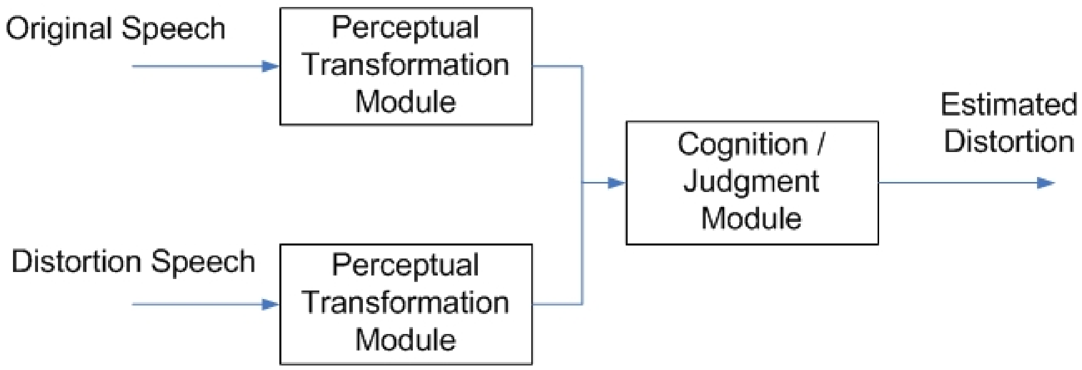

Figure 3.5: Basic structure of perceptual speech quality measurement (Sun,L. , 2004)

The perceived quality of the coded speech will, therefore, be independent of the type of coding and transmission (Campbell et al., 2009), when estimated by a distance measure between perceptually transformed speech signals which face the problems of time and spectral domain measure. Schroeder et al. (1979) was one of the first to develop an algorithm to include aspects of the human auditory system and Karjalainen (1985) was one of the first to use an auditory model to assess the quality of sound where his proposed model is based on comparison of auditory transforms of the original and processing signals, he introduced a more general technique for estimating error audibility based on a comparison of audible time-frequency loudness representations using the Auditory Spectrum Distance (ASD). According to Karjalainen's work, several new perceptual models for evaluating the quality of speech and audio coders emerged in the early 1990s such typical perceptual measure methods are Bark Spectral Distortion (BSD) measure, Modified and Enhanced Modified Bark Spectral Distortion measures (MBSD and EMBSD), Perceptual Speech Quality Measure and Perceptual speech quality measurement plus (PSQM and PSQM+), Measuring Normalizing Blocks (MNB), Perceptual Analysis Measurement System (PAMS) and Perceptual Evaluation of Speech Quality (PESQ) which is the latest ITU standard for assessing speech quality for communication systems and networks as will describe as follows.

1. Bark Spectral Distortion measure (BSD)

   BSD is a simplest perceptual measurement method based on the assumption that speech quality is directly related to speech loudness, which is a quantifiable property of auditory perception or a psychoacoustical term defined as the magnitude of auditory sensation. In order to calculate loudness, the speech signal is processed using the results of psychoacoustic measurements, which include critical band analysis, equal-loudness preemphasis, and intensity-loudness power law. The overall BSD measurement represents the average squared Euclidean distance of loudness vectors of the reference and the distorted speech. The main aim of the measure is to emulate several known features of perceptual processing of speech sounds by the human ear, especially frequency scale warping, as modeled by the bark transformation, and critical band integration in the cochlea; changing sensitivity of the ear as the frequency varies; and difference between the loudness level and the subjective loudness scale where the Figure 3.6 illustrates the approach steps for measure signal quality (Wang et al., 1992).

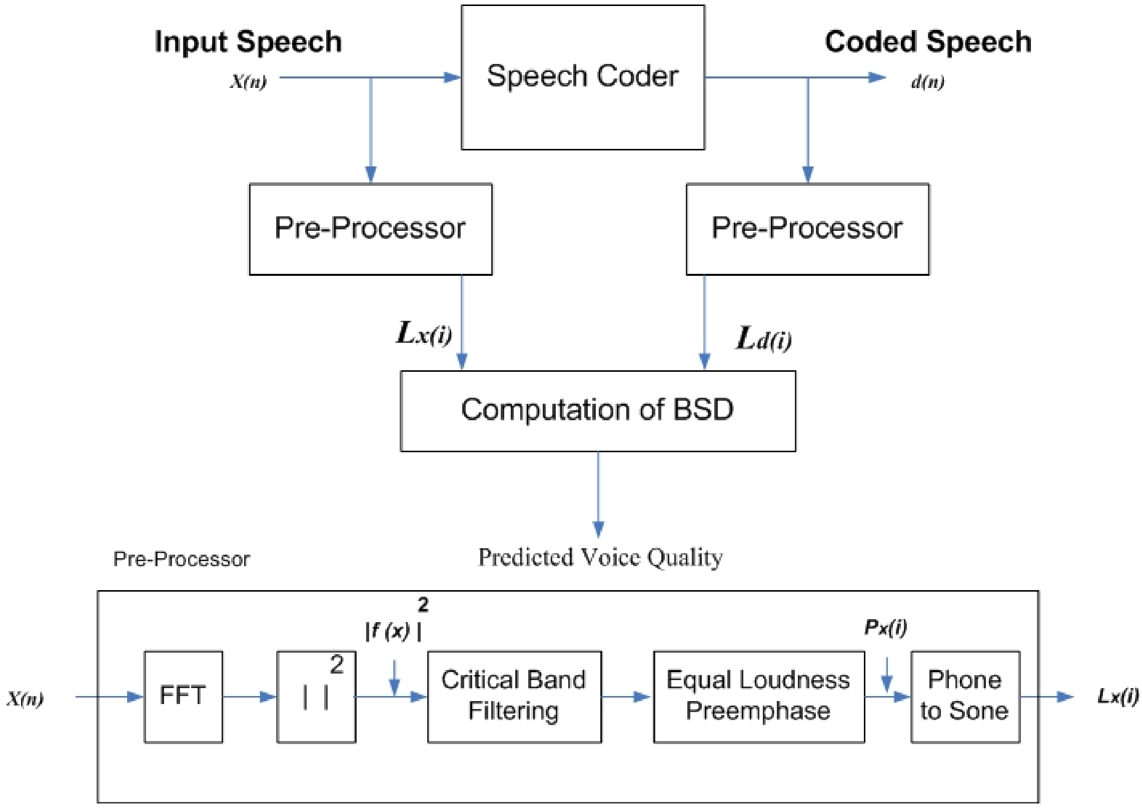

Figure 3.6: Block diagram representation of the BSD measure (Wang et al., 1992)

Both the original speech, x(n), and the distorted speech (coded version of the original speech), d(n), are pre-processed separately by same operations to obtain their bark spectra, L_x(i) and L_d(i), respectively. The starting point of the pre-processing operations is the computation of the magnitude squared Fast Fourier Transformation (FFT) spectrum to generate the power spectrum, $|X(f)|^2$. This is followed by critical-band filtering to model the nonlinearity of the human auditory system, which leads to a poorer discrimination at high frequencies than at low frequencies, and the masking of tones by noise. The spectrum available after critical band filtering is loudness equalized so that the relative intensities at different frequencies correspond to relative loudness in phones rather than acoustical levels. Finally, the processing operation ends with another perceptual nonlinearity: conversion from phone scale into perceptual scale of sones. By definition a sone represents the increase in power which doubles the subjective loudness. The ears nonlinear transformations of frequency and amplitude, together with important aspects of its frequency analysis and spectral integration properties in response to complex sounds, are represented by the bark spectrum L(i). By using the average squared Euclidean distance between two spectral vectors, the BSD is computed as (Wang et al., 1992)

$$BSD = \frac{\frac{1}{M}\sum_{m=1}^{M}\sum_{i=1}^{O}[L_x^{(m)}(i) - L_d^{(m)}(i)]^2}{\frac{1}{M}\sum_{m=1}^{M}\sum_{i=1}^{O}[L_x^{(m)}(i)]^2}, \tag{3.4}$$

where M is the number of frames (speech segments) processed, O is the number of critical bands, $L_x^{(m)}(i)$ is the bark spectrum of the m^{th} critical frame of original speech, and $L_d^{(m)}(i)$ is the bark spectrum of the m^{th} critical frame of coded speech. BSD works well in cases where the distortions in voice regions represent the overall distortion because it processes voiced regions only. Hence, voiced regions have to be detected (Wang et al., 1992).

2. Modified and enhanced modified bark spectral distortion measures (MBSD and EMBSD)

   MBSD is a modification of the BSD in which the concept of a noise-masking threshold is incorporated, that differentiates audible and inaudible distortions (Yang et al., 1997). It uses the same noise-masking threshold as that used in transform coding of audio signals. MBSD assumes that loudness differences below the noise masking threshold are not audible and therefore are excluded from the calculation of the perceptual distortion. There are two differences between the conventional BSD and MBSD. First, noise-masking threshold for determination of the audible distortion is used by MBSD, while the conventional BSD uses a power threshold. Second, the way in which the distortion is computed, while the BSD defines the distortion as the average squared Euclidean distance of estimated loudness, the MBSD defines the distortion as the difference in estimated loudness. Figure 3.7 describes the MBSD measure, The MBSD computes the distortion frame by frame, with the frame length of 320 samples using 50 percent overlap. MBSD uses a simple cognition model to calculate the distortion value. The distortion value for an entire test speech utterance was obtained by averaging over all non-silence frames.

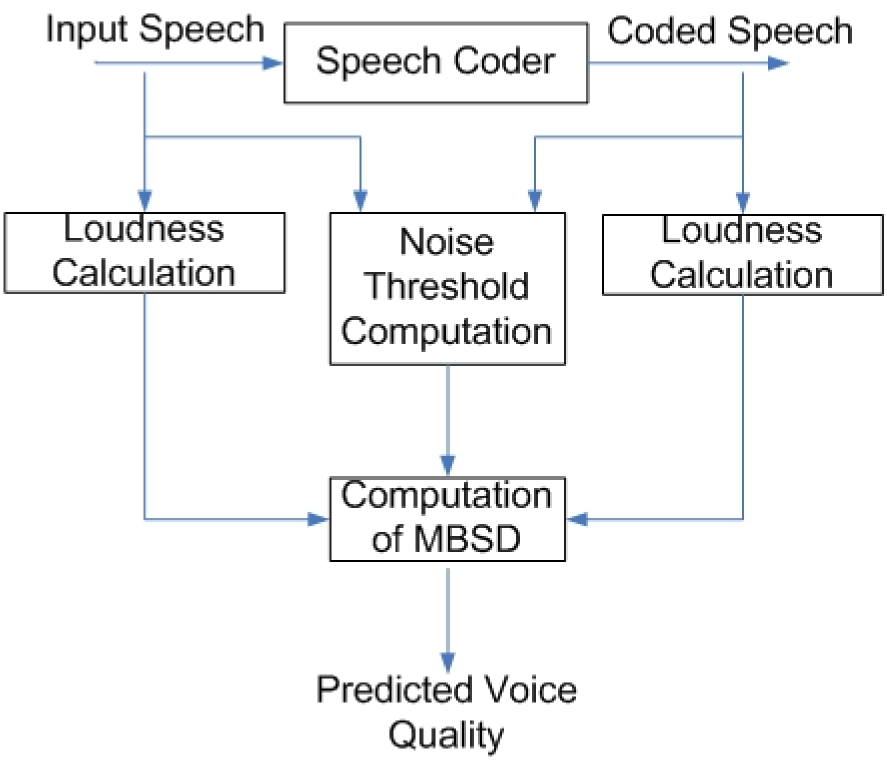

Figure 3.7: Block diagram of MBSD measure (Yang et al., 1997)

The EMBSD, on the other hand, is a development of the MBSD measure where some procedures of the MBSD have been modified and a new cognitive model has been used (Yang, 1999). These modifications involve the following: the amount of loudness components used to calculate the loudness difference, the normalization of loudness vectors before calculating loudness difference, the inclusion of a new cognition model based on post masking effects, and the deletion of the spreading function in the calculation of the noise masking threshold.

3. Measuring Normalizing Blocks (MNB)

   MNB algorithm comprises two stages: a simple perceptual transformation, and a distance measure that uses hierarchies of measuring normalizing blocks as illustrated in Figure 3.8. For perceptual transformation, the time-aligned, normalized signals, original and received, are divided into 50 percent overlapping frames of 128 samples. Each frame is multiplied by a Hamming window and transformed using Fast Fourier transform (FFT). Only the squared magnitudes of the FFT coefficients are preserved (Voran and States., 1998).

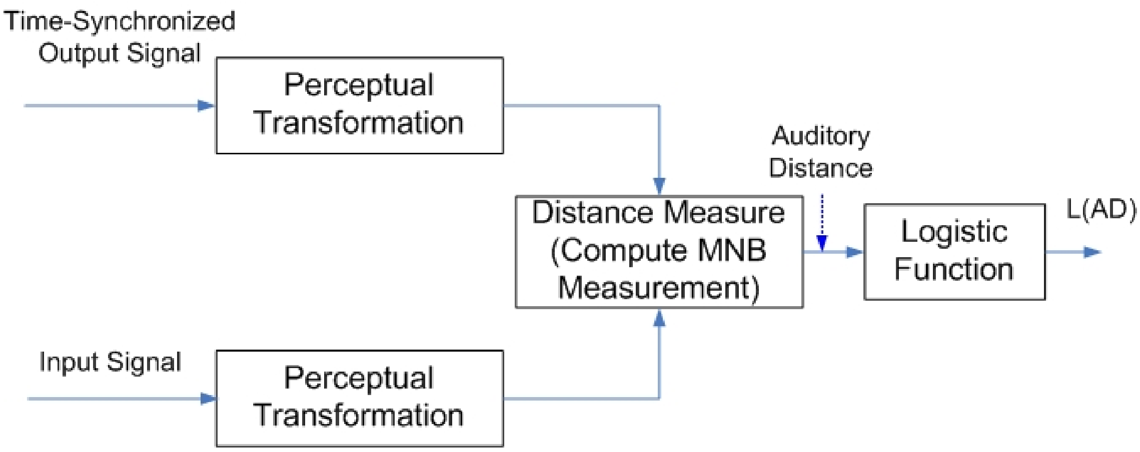

Figure 3.8: The MNB Model (Voran and States., 1998)

The coefficients are transformed to the Bark scale, a psychoacoustic frequency scale (Schroeder et al., 1979) where

$$b = 6.\sinh^{-1}(\frac{f}{600}) \tag{3.5}$$

defines the transformation. This is achieved by clustering the squared FFT coefficients into bins of equal width on the Bark scale. The total energy of each frame is computed, and frames below an energy threshold in either the original or received signals are discarded. All samples in remaining frames are transformed using a logarithm to model perceived loudness.

The distance measure used is a linear combination of the distances computed in the time and frequency MNBs. There is one Frequency MNB (FNMB) for each power spectrum coefficient. A frequency MNB averages the difference at that coefficient between the original and received signals across all frames that exceed the above-mentioned energy thresholds. Four measurements covering the lower and upper band edges of telephone band speech are saved in measurement vector m. There are two different Time MNB (TMNB) structures using different frequency scales, producing either eight or seven measurements saved in m, depending upon which is used. The measurement vector m is multiplied by a weight vector w to compute a single Auditory Distortion (AD) number.

$$AD = \sum_{i=1}^{12} weight_i . m(i) \tag{3.6}$$

These values are then passed through a logistic function to create L(AD). The logistic function is:

$$l(z) = \frac{1}{1 + e^{\alpha . AD + \beta}} \tag{3.7}$$

to map it to (0,1).

4. Perceptual Speech Quality Measure and Perceptual speech quality measurement plus (PSQM and PSQM+)

   Perceptual Speech Quality Measurement or PSQM was approved in 1996 by ITU-T and published by as ITU-T Recommendation P.861 (1996). PSQM, as shown in Figure 3.9, is a mathematical process that provides an accurate objective measurement of the subjective voice quality. PSQM transforms the speech signal into the loudness domain, modifying some parameters in the loudness calculation through mapping the physical signals constituting the source and coded speech onto psychophysical representations that match the internal representations of the speech signals (the representations inside our heads) as closely as possible. These internal representations make use of the psychophysical equivalents of frequency (critical band rates) and intensity (Compressed Sone), where the Masking is modeled in a simple way: only when two time-frequency components match in both the time and frequency domains, masking is taken into account.

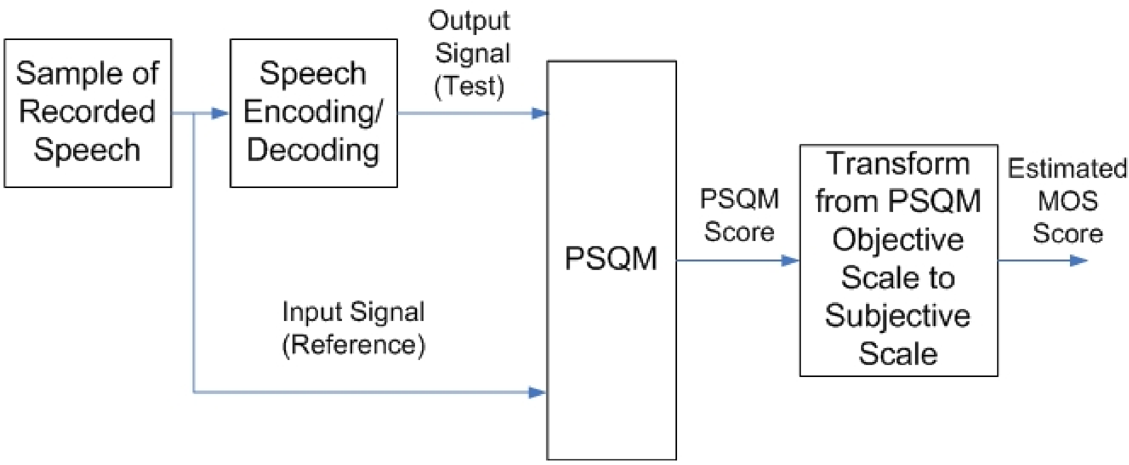

Figure 3.9: Block Diagram of PSQM (Rango et al., 2006)

To optimize performance, PSQM applies a nonlinear scaling factor to the loudness vector of distorted speech, where the scaling factor is obtained using the loudness ratio of the reference and the distorted speech. The difference between the scaled loudness of the distorted speech and loudness of the reference speech is called noise disturbance as a function of time and frequency. The final estimated distortion is an average Noise Disturbance (ND) over all the frames processed. Silence portions have only a small weight in the calculation of distortion. PSQM computes the distortion frame by frame, with the frame length of 256 samples (8 KHz sampling) with 50 percent overlap. The result is shown in noise disturbance as a function of time and frequency. The average noise disturbance is directly related to the quality of coded speech.

In order to improve the performance of PSQM, PSQM+ was proposed for loud distortions and temporal clipping. PSQM+ uses the same perceptual transformation module as PSQM. PSQM+ also introduces an additional scaling factor when the overall distortion is calculated. This scaling factor makes the overall distortion proportional to the amount of temporal clipping distortion. Otherwise, the cognition module is the same as PSQM.

5. Perceptual Assessment of Speech Quality (PAMS)

   Psytechnics, a UK-based company associated with British telecommunications (BT), developed an objective speech quality measure called perceptual analysis measurement system (PAMS) (Rix and Hollier, 2000). PAMS uses a model shown in Figure 3.10 based on factors of human perception to measure the perceived speech clarity of an output signal as compared with the input signal.

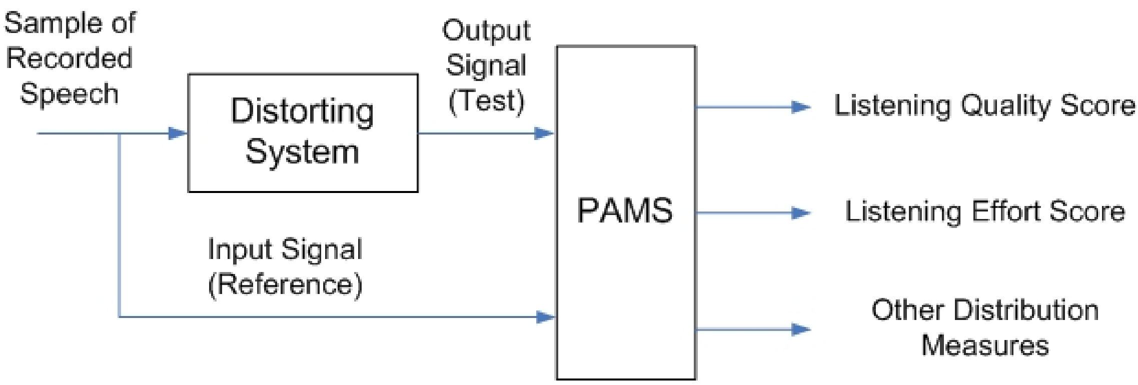

Figure 3.10: PAMS testing process (Rix and Hollier, 2000)

Two signals are inserted in a model that aligns themselves and remove the delay effects. Although similar to PSQM in many aspects, PAMS uses different signal processing techniques and a different perceptual model. PAMS extract and select parameters describing speech degradation addressed by damaging factors such as time clipping, packets loss, delay and distortion due to the codec usage and constrained mapping to subjective quality. A parameter set, in which each parameter increases with increasing degradation, is generated. The best set of parameters is selected with a training procedure, in summery PAMS uses a concept of mapping from the parameter domain to subjective quality domain. It is flexible in adopting other parameters if they are perceptually important. The performance of PAMS depends upon the designer's intuition in extracting candidate parameters as well as selecting parameters with a training data set. Since the parameters are usually not independent of each other, it is not easy to optimize both the parameter set and the associated mapping function. So, extensive computation is performed during training.

6. Perceptual evaluation of speech quality (PESQ),

   The classical PSQM was improved to correlate better with subjective tests under network conditions. This resulted in a new measure known as PSQM99. The main difference between the PSQM99 and PSQM concerns the perceptual modeling where they are differentiated by the asymmetry processing and scaling. PSQM99 provides more accurate correlations with subjective test results than PSQM and PSQM+. Later on, ITU-T recognized that both PSQM99 and PAMS had significant merits to be combined into a new measurement technique that is beneficial to industry, so ITU-T in May 2000 submitted a new measurement technique for intrusive objective speech quality assessment called Perceptual Evaluation of Speech Quality (PESQ). In February 2001, ITU-T approved the PESQ under ITU-T Recommendation P.862 (2001). PESQ is directed at narrowband telephone signals and is effective for measuring the impact of the following conditions: waveform and no waveform codecs, transcodings, speech input levels to codecs, transmission channel errors, noise added by system (not present in input signal), and short and long term warping.

   PESQ combines the time-alignment techniques of PAMS with the accurate perceptual modeling of PSQM99. It is designed for use with intrusive methods where an input signal is injected into the system under test, and the distorted output is compared with the input (reference) signal. The difference is then analyzed and converted into a quality score. The structure of PESQ model is shown in Figure 3.11 as illustrated.

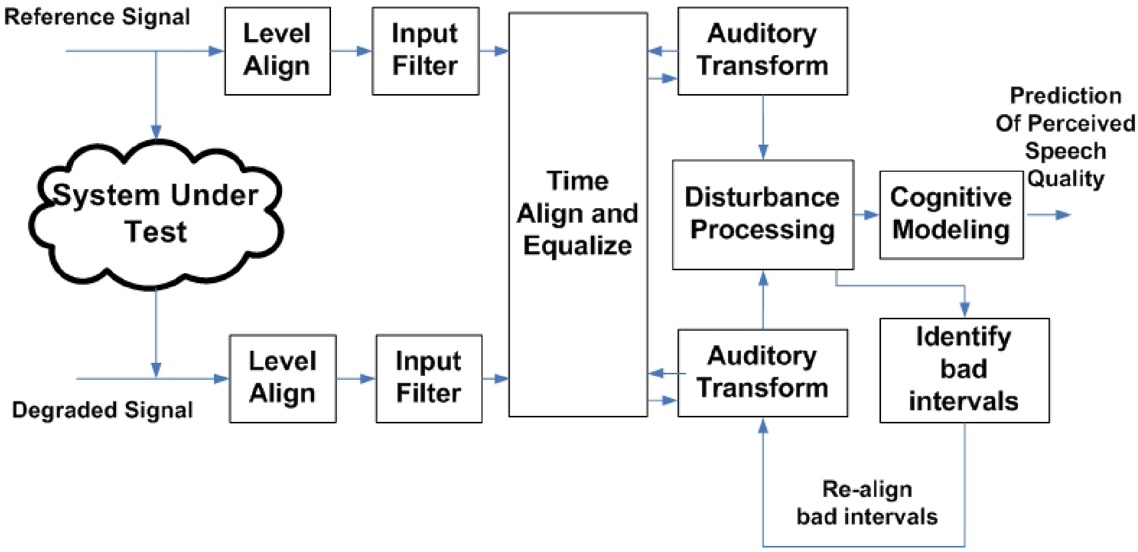

Figure 3.11: Structure of PESQ (Rix et al., 2001)

PESQ algorithm involves the following processing stages. First, the model aligns both the reference signal and the degraded signal to the same constant power level that corresponds to the normal listening level used in subjective tests to process a series of delay between two signals to determine the start and end points. Based on the set of delays that are found, PESQ compares the original (input) signal with the aligned degraded output of the device under test using a perceptual model. Then the signals are filtered using an FFT-based input filter to model and compensate or mimic for the filtering that takes place in the telephone handset and in the network. Both signals are aligned in time and then processed through an auditory transform similar to that used in PSQM and PAMS.

PESQ auditory transform is a psychoacoustic model which maps the signals into a representation of perceived loudness in time and frequency as mimicked in human auditory taking account of perceptual frequency (Bark) and loudness (Sone) and removes those parts of the speech that are inaudible to the listener which achieved in several stages such as time alignment, level alignment to a calibrated listening level, time-frequency mapping, frequency warping, and compressive loudness scaling. In the disturbance processing stage, two distortion parameters are extracted from the difference between the auditory transforms of the two signals: the absolute (symmetric) disturbance, which is a measure of absolute audible error, and the additive (asymmetric) disturbance, which is a measure of audible errors that are significantly louder than the reference. The two distortion parameters are aggregated in frequency and time over several time-frequency scales using a nonlinear averaging method designed to take optimal account of the distribution of error in time and amplitude.

The final PESQ score is a linear combination of the average symmetric disturbance value and the average asymmetric disturbance value, computed using the following formula (ITU-T Recommendation P.862, 2001) :

$$MOS_{PESQ} = 4.5 - 0.1 * d_{SYM} - 0.0309 * d_{ASYM}, \tag{3.8}$$

where MOS_{PESQ} represents the P.862 PESQ MOS, d_{SYM} is the average symmetric disturbance value and d_{ASYM} is the average asymmetric disturbance value. The range of the PESQ score is between -0.5 and 4.5 as opposed to the ACR listening quality MOS which is on a 1-5 scale. It is therefore desirable to provide an objective listening quality score from the P.862 that allows a linear comparison with subjective MOS. In order to achieve that and to align the PESQ MOS with the new MOS terminology as defined in ITU Recommendation P.800.1 (1994), ITU-T published their recommendation P.862.1 (2003). This recommendation defines a mapping function and its performance for a single mapping from raw P.862 PESQ MOS scores to PESQ MOS-LQO. The mapping function is defined by P.862.1 (2003):

$$z = 0.999 + \frac{4.999 - 0.999}{1 + e^{-1.4945y + 4.6607}} \tag{3.9}$$

where

y is the P.862 PESQ MOS score

z is the corresponding PESQ MOS_LQO score

In 2005, the ITU-T issued P.862.2 (2005), which describes another extension to the P.862 PESQ algorithm. The P.862.2 provides recommendation to extend the application of P.862 PESQ to wideband audio systems (50-7000 Hz). The definition of a new output mapping function, which is a modification to that recommended in P.862.1, to be used with wideband applications is as follows:

$$z = 0.999 + \frac{4.999 - 0.999}{1 + e^{-1.3669y + 3.8224}} \tag{3.10}$$

where y and z are as in Equation (3.9)

3.2.2 Non-intrusive methods

The intrusive methods described in section 3.2.1, based on input-output approach needs the input signal as a reference for comparison with the output signal which is a problem in live network as it is difficult to obtain the original speech signal from the sender side, in addition intrusive methods have few other problems. First of all is time-alignment between input and output signal, which is achieved by automatic synchronization, is an important factor in deciding the accuracy of the measure. In practice, perfect synchronization is difficult to achieve due to fading or error burst that are common in wireless systems (Mahdi and Picoviciv, 2009), and hence degradation in the performance of the measure is inevitable. Second, absent of the original signal in many applications, as in cases of wireless and satellite communications which makes intrusive methods not applicable for live traffic monitoring although it is more accurate. Furthermore, in some situations the input speech may be distorted by background noise, and hence, measuring the distortion between the input and the output speech does not provide a true indication of the speech quality of the communication system. hence non-intrusive or passive speech quality assessment methods that depend only on the degraded, test or output signal for evaluation with absent of reference, input or original signals that would therefore heal all the above problems and provide a convenient non-intrusive measure for monitoring of live networks, in other word, real-time quality assessment which is important for instance if the application is to perform some form of dynamic quality control, e.g. by changing encoding or redundancy parameters to optimize quality when network conditions worsen.

Non-intrusive models are classified into two category, signal-based and parametric-based methods as illustrated in Figure 3.2, where signal-based methods as the name suggests based on digital signal processing that estimates the speech quality when the envelope of the speech signal may have suffered from degradation overtime due to low-bit rate coding or transmission over noisy wireless links, in other words signal-based methods process the audio stream that are decoded after buffer playout to extract relevant information for estimating the voice quality. ITU-T Recommendation P.563 or 3SQM (Single-Sided Speech Quality Measurement) that achieves a correlation coefficient with listening tests of around 0.8 defined to be a standard for this type of measure.

On the other hand, parametric-based methods based their results on various properties relevant to telecommunication network parameters for example jitter, delay and packet loss, this leads parametric model to be more specific for a particular type of communications network by depending their prediction on the parameters of that network, which makes parametric-based methods to be more accurate than signal-based methods for that network which are more suitable for general prediction for a wider variety of networks and conditions. The E-model which is one of the most widely used parametric-based methods defined according to ITU-T Recommendation G.107. The details of the ITU-T

Recommendation P.563 or 3SQM and the E-Model are discussed in the next paragraphs.

- **ITU-T Recommendation P.563 or 3SQM**: In May 2004, the ITU-T approved a new non-intrusive voice quality assessment algorithm under its recommendation P.563: single ended method for objective speech quality assessment in 3.1 kHz (narrow-band) telephony applications (ITU-T Recommendation P.563, 2004). The P.563 approach is the first recommended method for single-ended non-intrusive measurement applications that takes into account the full range of distortions occurring in PSTN and that is able to predict the speech quality on a perception-based scale MOS_{LQO} according to ITU-T Recommendation P.800.1. This recommendation is not restricted to end-to-end measurements; it can be used at any arbitrary location in the transmission chain. The basic block diagram of P.563 is shown in Figure 3.12.

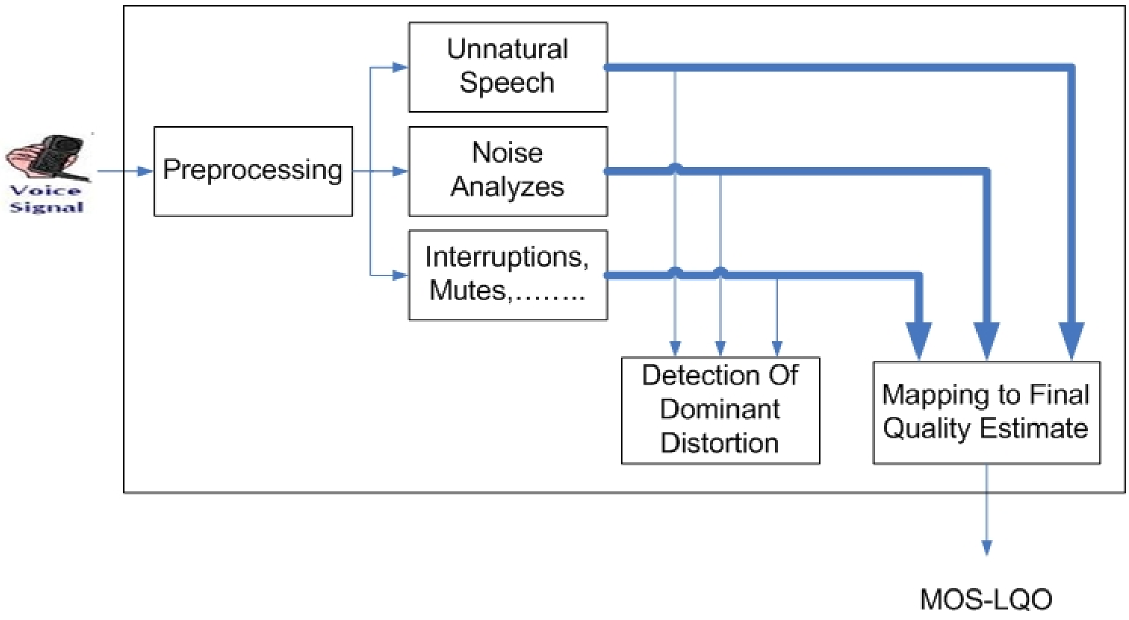

Figure 3.12: Basic block diagram of P.563 overall structure (ITU-T Recommendation P.563, 2004)

The P.563 approach could be visualized as an expert who is listening to a real call with a test device like a conventional handset into the line in parallel. This visualization explains also the main application and allows the user to rate the scores

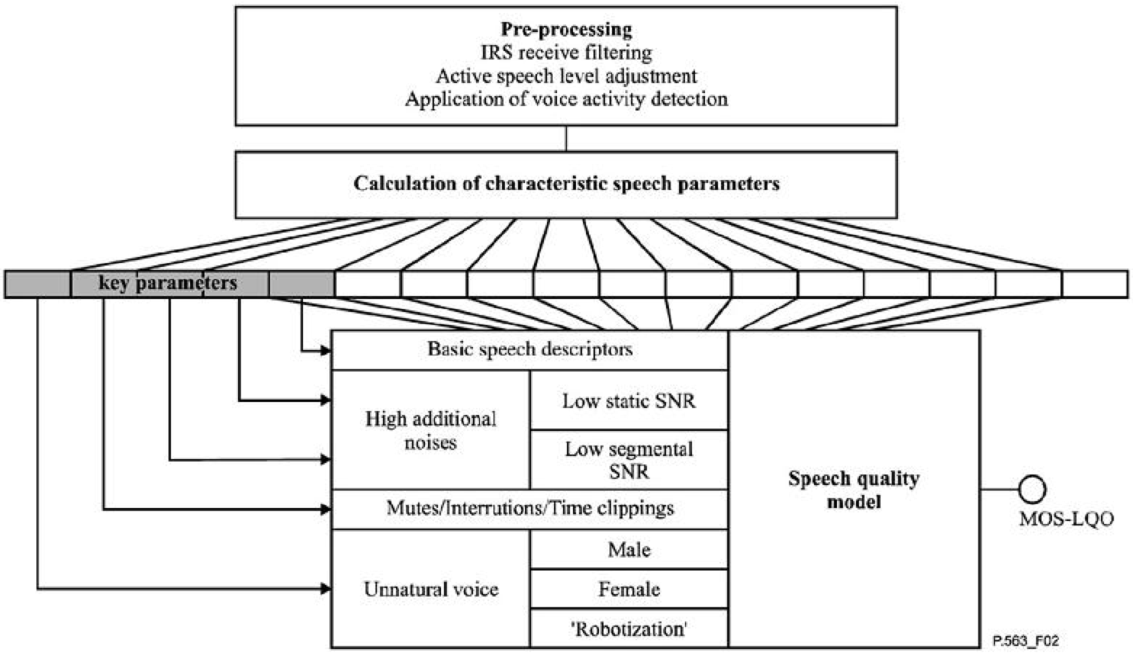

Figure 3.13: Block diagram of P.563 algorithm detailing the various distortion classes used (ITU-T Recommendation P.563, 2004)

gained by P.563. The quality score predicted by P.563 is related to the perceived quality by linking a conventional handset at the measuring point. Hence, the listening device has to be part of the P.563 approach. To achieve this, the algorithm combines 4 processing stages as illustrated in Figure 3.13: preprocessing; basic distortion classes and speech parameters extraction; detection of dominant distortion; and mapping to final quality estimate. Brief overview of the main steps is given here:

- **Preprocessing**: the first step in this stage is the intermediate reference system (IRS) filtering, where the speech signal to be assessed is filtered using a filter that simulates a standard receiving telephone handset. This is followed by a Voice Activity Detector (VAD) to identify and separate portions of the signal that contain speech. The speech level is then calculated and adjusted to -26 dBov.
- **Extraction of basic distortion classes and speech parameters**: The pre-

processed speech signal to be assessed will be investigated by several separate analyzes, which detect a set of characterizing signal parameters. These parameters are divided up into 3 independent functional blocks corresponding to 3 main classes of distortion, namely: vocal tract analysis and unnaturalness of speech; analysis of strong additional noise; and speech interruptions, mutes and time clipping. In total, 51 distortion parameters are computed. All of these distortion classes are based on very general principles which make no assumptions about the underlying network or distortion types occurring under certain conditions. The only prerequisite is the scientific knowledge on how human speech is generated and how it is perceived by human beings. In addition, a set of basic speech descriptors like active speech level, speech activity and level variations are used, mainly for adjusting the pre-processing and the VAD. Some of the signal parameters calculated within the pre-processing stage are used in these 3 functional blocks.

– **Detection of dominant distortion**: This analysis will be applied at first to all signals. Based on a restricted set of key parameters, an assignment to a main distortion class will be made. The key parameters and the assigned distortion class are used for the adjustment of the speech quality model. This provides a perceptual based weighting where several distortions are occurring in one signal but one distortion class is more prominent than the others. The process models the phenomenon that any human listener focuses on the foreground of the signal stream. That is the listener would not judge the quality of the transmitted voice by a simple sum of all occurred distortions but because of a single dominant noise artifact in the signal.

– **Final quality estimate**: in this stage, a speech quality model is used to map the estimated distortion values into a final quality estimate equivalent to a Mean Opinion Score-Listening Quality Objective (MOS_{LQO}) according to P.800.1. The speech quality model is composed of 3 main blocks:

* decision on a distortion class.
* speech quality evaluation for the corresponding distortion class.
* overall calculation of speech quality.

First, the assigned dominant distortion class calculated in the key parameters block is used for the adjustment of the speech quality model. In the case of several distortions occurring in the signal, a prioritization is applied on the distortion classes according to the distortions relevance with respect to the average listeners opinions. This is followed by estimation of an intermediate speech quality score for each class distortion. Each class distortion uses a linear combination of parameters to generate the intermediate speech quality. The final speech quality estimate is calculated by combining the intermediate quality results with some additional signal features

- **The E-model**: the E-model is a computational model originally developed, for transmission network planning tool to help operators design the network or to live monitoring network, by a working group called "Voice Transmission Quality from Mouth to Ear" (ITU-T Recommendation G.107, 2000), within European Telecommunications Standards Institute (ETSI), In the late 1990s, the E-model was standardized by the ITU-T as Recommendation G.107 (ITU-T Recommendation G.107, 2000). Today, the E-model is a widely used transmission planning tool that describes several parametric models of specific network impairments, such as packet lose rate that contained in RTP and RTCP protocols, and terminal impairments which exploited to predict voice quality non-intrusively for VoIP applications. The basic paradigm of the E-model is that the quality is perceived psychologically and thus the psychological factors are additive on the psychological scale. This concept used to perceptually describe the effect of various impairments on the session calling, where the effect of these impairments will combined into transmission rating scale called R rating, where all impairments are -by definition- additive and thus independent of each other. The E-model is a function of 20 input parameters where Table 3.4 show these parameters with their default values, where the reference connection, as shown in Figure 3.14, is split into a send side and a receive side. The model estimates the conversational quality from mouth to ear as perceived by the user at the receive side, both as listener and talker.

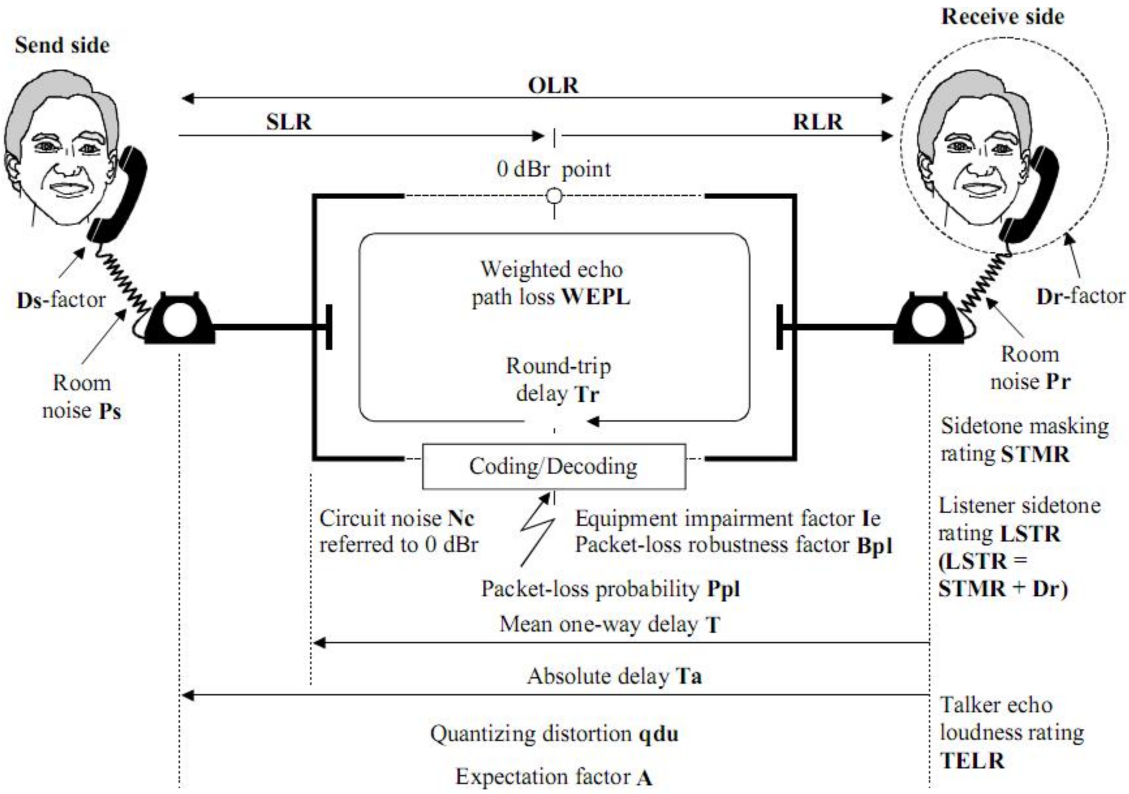

Figure 3.14: Reference connection of the E-model (ITU-T Recommendation G.107, 2000)

A transmission rating factor R is obtained from the impairment factors by

$$R = R_0 - I_s - I_d - Ie\text{-}eff + A; \tag{3.11}$$

where

R_0 Describes a base factor representative of the signal-to-noise ratio and an advantage (or expectation) factor.

Table 3.4: Default values and permitted ranges for the E-model's parameters (ITU-T Recommendation G.107, 2000)

Parameter	Default value	permitted range
Send Loudness Rating	8	0...+18
Receive Loudness Rating	2	-5...+14
Sidetone Masking Rating	15	10...20
Listener Sidetone Rating	18	13...23
D-value of Telephone, Send Side	3	3...+3
D-value of Telephone, Receive Side	3	-3...+3
Talker Echo Loudness Rating	65	5...65
Weight Echo Path Loss	5	5...110
Mean one-way Delay of the Echo path	0	0...500
Round Trip Delay in a 4-wire Loop	0	0...1000
Absolute Delay in echo-free Connections	0	0...500
Number of Quantization Distortion Units	1	1...14
Equipment Impairment Factor	0	0...40
Packet-loss Robustness Factor	1	1...40
Random Packet-loss Probability	0	0...20
Burst Ratio	1	1 2
Circuit Noise Referred to 0 dBr-point	-70	-80...-40
Noise Floor at the Receive Side	-64	
Room Noise at the Send Side	35	35...85
Room Noise at the Receive Side	35	35...85
Advantage Factor	0	0...20

I_s Represent speech transmission impairment factors (e.g., impairments due to quantization distortion).

I_d Delay impairment factors (e.g., impairments due to echoes).

Ie-eff Effective equipment impairment factors (e.g., impairments due to packet loss for different codec types).

A Advantage factor serves as an offset that accounts for user expectations of the quality of service, as example, for wireless communications, A = 10 is used. In turn, for satellite communications in remote locations where a minimum of two satellite hops are warranted, an advantage factor of A = 20 is recommended.

Each of the parameters in Equation (3.11) except the advantage factor (A) is further decomposed into a series of equations. When all parameters set to their default values, R-Rating Factor has the value of 93.2 which is mapped to an MOS value of 4.41. The estimated quality according to the E-Model is conversational, i.e. MOS Conversational Quality Estimated (MOS_{CQE}), because the effect of delay is taken into consideration. When the effect of delay is ignored and Id is set to its default value the estimation is listening only, i.e. MOS Listening Quality Estimated (MOS_{LQE}). The computed R-Rating Factor can be mapped into an MOS value. Equation (3.12) gives the mapping function between the computed R-Rating factor and the MOS value.

$$MOS-CQE = \begin{cases} 1, & \text{R<0;} \\ 1+0.035R+R(R-60)(100-R)*7*10^{-6}, & \text{0<R<100;} \\ 4.5, & \text{R>100.} \end{cases} \tag{3.12}$$

This formula can convert R Rating factor in the range 6.5$<$ R $<$100 to calculate R from MOS_{CQE}.

Appendix I of ITU-T Recommendation G.107 defines a set of equations that used to compute back the R factor from MOS value as follows:

$$R = \frac{20}{3}(8 - \sqrt{226}\cos(h + \frac{\pi}{3})) \tag{3.13}$$

where

$$h = \frac{1}{3}\arctan 2(18566 - 6750MOS_{CQE}, 15\sqrt{-903522 + 1113960MOS_{CQE} - 202500MOS_{CQE}^2}) \tag{3.14}$$

and:

$$\arctan 2(x,y) = \begin{cases} \arctan(\frac{y}{x}), & \text{for x} \geq 0; \\ \pi - \arctan(\frac{y}{-x}), & \text{for x} < 0. \end{cases} \tag{3.15}$$

The function $\arctan 2(x,y)$ is implemented in ANSI C as the function atan2(y, x). Users should note that the order of the two parameters differs in this case.

- **Other methods for speech quality assessments:**
  - A non-intrusive speech quality assessment depend on Genetic Algorithm (GA) approach that are readable and wedding out irrelevant information proposed by Raja et al. (2006), which automatically predicts the effect of packet loss on speech quality, where a set of parameters are extracted from received signals and entered into a scaled Mean Squared Error (MSE) equation as the fitness criterion to chose the best individuals that tests on problems by PESQ score that was used as a quality comparison estimation model and to optimize the coefficient of the model that will estimate the quality as close as PESQ, since the most frequent occurrence of input parameters in Genetic Programming

(GP) simulation or the best coefficient of structured model that appeared in genome of the best individuals of all of simulation is mean lose rate for talk-spurt frame distinguished by Voice Activity Detection (VAD)-mbl_{VAD}- that fed into a simple equation to compute the $MOS-LQO_{GP}$ which used as a model with one parameter for future estimation.

– Because GP has a main advantages of producing human-readable results in the form of analytical expressions and deals with significant input parameters which aids in automatic pruning of irrelevant information, Raja et al. (2007) in another work estimated the effect of burstiness on speech quality by using GP, that turn out that burst length least used by the best individuals of various runs of GP simulation, so according to optimizing the input parameters by MSE equation to compute the best individual that carry the best parameters that occur more times anew equation for real-time estimator that depend on frequent occurring parameters was developed, which is a good approximation to PESQ value which also more efficient.

– A signal-based method for non-intrusive evaluation of speech quality proposed by (Raja and Flanagan, 2008), that uses the parameter extraction model of ITU-T P.563 algorithm and a subjective test as a reference target value, the GP based symbolic regression was used to connect between the various features and estimated of speech quality from subjective test. In that novel approach two experiments were used, the first was based on GP with scaled MSE as the fitness function, while the second experiment employed a hybrid approach in which the coefficient of the selected individual where tuned using a GA. From the two experiments the best individual in term of fitness over testing data has been proposed as a model for quality estimation.

– The equipment impairment factors for a mixed Narrow band / Wide band (NB/WB) ($I_{e,WB,eff}$) which depends on the human listening can be predicted, where Raja et al. (2008) attempted to develop three formulae which tied the VoIP network traffic parameters as input variables, where the GP approach was used to reduce the parameters to best one, with the $I_{e,WB,eff}$ values as the target output for GP which perform the symbolic regression during evolution. In the proposed approach the WB_{PESQ} was used for deriving reference values of $I_{e,WB,eff}$ instead of subjective test

- Using ANN may give good indication on the accuracy of the perceived quality, Sun,L. and Ifeachor,E. C. (2002) proposed a ANN model to predict the speech quality for VoIP, where the extracted parameters perform the input parameter for the model to predict speech quality directly from network parameters and PESQ as target value for the same model, Sun,L. and Ifeachor,E. C. (2002) determined the parameters where they took into account talkspurt-based conditional and unconditional packet loss rate (instead of the network packet loss rates because they are perceptually more relevant), Codec type and the gender of the talker (extracted from decoder). The correlation coefficient accuracy of the test and validation data sets of the model are 0.952 and 0.946 respectively.
- A new novel methodology proposed by Da Silva et al. (2008), called Pseudo-Subjective Quality Assessment (PSQA) which is based on merging subjective assessment with a statistical learning tool (A Random Neural Network, or RNN that allows to produce subjective-like quality estimations. This method has 3 main steps: (a) a set of quality-effecting parameters must be selected, where it should select few parameters to reduce the method run time, the parameters may be measurable network parameters such as loss bursts, one-way delay and the delay jitter and application parameters such as bit rate, (b) a set of subjective session tests must be performed according to parameters of network conditions that were selected to capture the MOS values for each session, (c) A RNN should be used to map the selected parameters with the MOS subjective target by training more time and validating it, it was reported that

the chosen RNN has very good generalization capabilities.

– As shown in Equation 3.11 and mentioned in ITU-T Recommendation G.107 (2000) the *Ie-eff* which is impairment due to codec distortion, packet loss and jitter depends on 4 parameters to compute its value as follows:

  I_e Codec-specific equipment impairment factor

  Bpl Codec-specific packet loss robustness factor

  Ppl packet-loss probability

  BurstR Burst ratio (to count for Burstness in packet loss)

  The first two parameters depends on subjective test to compute their values, where the last two values calculated according to 2-state Gilbert Model which is packet-loss simulator. An extension to E-model where proposed by AL-Akhras et al. (2009) to face the problem of using subjective test to calibrate the E-model parameters through using an ANN model that maps the value of Bpl and BurstR that simply computed by simulator to *Ie-eff* Using the model, the PESQ score was used to compute MOS score for each Bpl and BurstR, this score is then converted to R rating, which from simple formulae the *Ie-eff* can be computed the formulae as follows:

$$R = R_0 - Ie\text{-}eff \tag{3.16}$$

  where R_0 has default value where all parameters are set to their values. The 3 values Bpl, BurstR and *Ie-eff* can be tied with ANN model, where the other values that depend on subjective test are melted as weighted on the ANN model that will be estimator. If new conditions or new coders are to be used, the simulation of above method are re-derived without using subjective tests.

– A parametric non-intrusive VoIP speech quality assessment algorithm was proposed by Ding et al. (2007) that consists of 3 steps. First, a particular impairment is detected, then its effects on speech quality are quantified, finally, an overall assessment model is developed. The proposed algorithm used 3 critical impairments: packet loss , temporal clipping and noise where the algorithm can use many impairments for assessment. The RTP and RTP eXtended Report (RTCP-XR) control protocols were used to detect the packet loss information according to E-model packet loss effect modeling can be computed which are transformed into MOS domain. Temporal clipping can be detected by analysing the decoded voice payload and its effects on speech quality was computed by a developed model. The last impairment was noise where the noise power was determined by degraded speech which is determined from active speech level, since its effect is modeled by E-model. Finally the overall assessment algorithm combines the effect of packet loss and temporal clipping into intermediate MOS score, then the effect of noise is incorporated using the E-model result. The performance results of the algorithm showed that the algorithm is effective and accurate. For the overall model, the correlation between prediction and measurement is 0.90; the Root Mean Square Error (RMSE) is 0.27 mean opinion score (MOS).

Chapter 4

The Proposed Technique For Estimating The Voice Quality

Measuring the transmitted voice quality is a challenge for current communication researchers especially in a packet_based network compared with traditional PSTN that has an established high voice quality through providing a dedicated path established especially for transmitting voice, so all hardware and software requirements on that path will be allocated to the voice call to guarantee receiving very satisfied voice quality at the receiver end. On the other hand, packet_based networks that fragment the transmitted voice signal into bulk of packets that transmitted randomly in non-dedicated paths. As a result some of these packets may get lost in any path which reduces the quality for transmitted voice signal. In this chapter the main contributions of this thesis will be introduced to estimate the voice quality. This chapter proposes a non-intrusive and objective technique for assessment of voice quality that helps in managing and monitoring the network state and solve each problem that affect the voice quality by either increasing the bandwidth capacity or upgrading to new better devices. The main aims behind this chapter are as follows:

1. Building a new non-intrusive and objective method for assessment of voice quality, to face the current challenges where the reference voice signal is not available at

the receiver end because of the far distance between the two ends.

2. Studying the affects of some parameters on voice quality, these parameters may be parametric depends on network statistical analysis or signal parameters that depends on the voice itself.

3. Studying six different classifiers to estimate the PESQ Mean Opinion Score -Listening Quality Estimation (PESQ MOS_{LQO}), these six classifiers are Regression with two types Linear and non-linear, Artificial Neural Network (ANN) and Ant-Miner with three types original Ant-miner, $_{c}Antminer$ and $_{c}Antminer$ type 2. These classifiers were trained and tested using 10 folds experiments as will discussed later.

4. Proposing a technique that depends on ACO to estimate the voice quality and the accuracy of the proposed technique is studied in comparison with other methods such as Linear Regression, Non-Linear Regression and ANN.

5. Presenting result of the conducted experiments to measure the speech quality.

4.1 Experiment Parameters

To build the proposed technique, a set of previous knowledge should be prepared to be a starting point for the process of the proposed technique as this study depends on ACO especially Ant-miner as illustrated in chapter 2 to estimate the voice quality. In Ant-miner a set of rules can be resulted from these Ants according to training data that used by ants in the input process. The resulted rules contain a set of attributes (parameters) with their values and a class label which is a Perceptual Evaluation of Speech Quality (PESQ) MOS Score (MOS_{PESQ}) that can be mapped into PESQ MOS_{LQO} explained in Figure 4.1.

The proposed technique uses 5 parameters that are packet loss, language, burst ratio, gender and codec, although there are several parameters that have effects on voice quality. This thesis focuses it's study on mentioned 5 parameters as will be illustrated in details as follows:

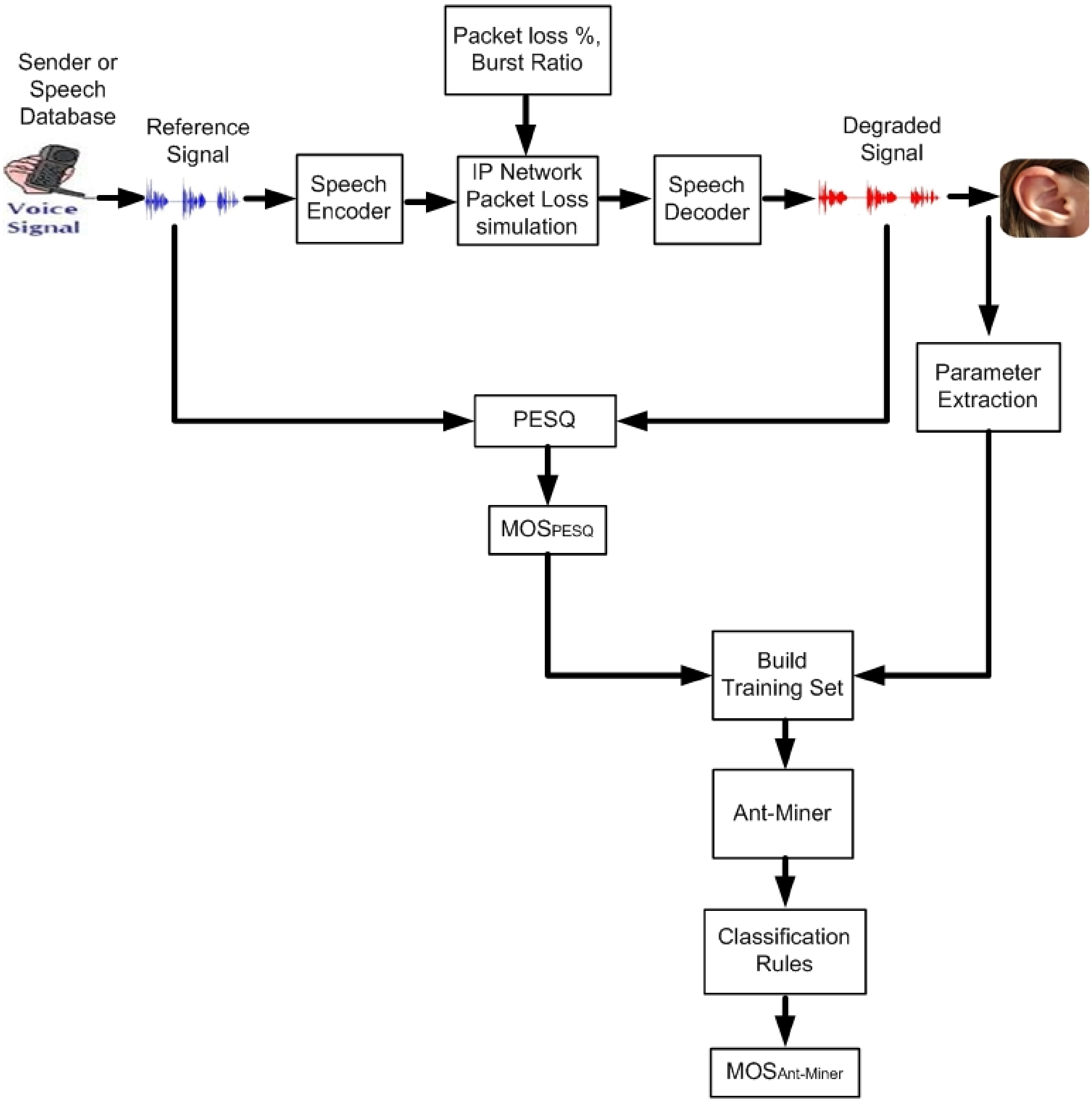

Figure 4.1: The proposed technique for measuring voice quality

4.1.1 Packet Loss

Transmitted voiced packets over IP network encountered many problems that lead to loss of some of packets, packet loss problem may result from different scenarios: hardware can not meet bandwidth demands for receiving packets on that hardware such as Routers; a packet loss may occur at the receiver end due to excessive delay or packet loss may occur due to the interference observed on a wireless link. Many techniques attempted to solve these problems, but packet loss in packet_based network is inevitable, so packet loss still has an effect on the received quality. To study the effect of packet loss on voice quality, a mathematical simulator model is needed.

A sent packet may be in two state found state or lost state depends on the time of playout these packets on receiver end. A Markov model is used to simulate a n-transition state of any system, it can be used to simulate the 2-state transition of packet loss of packet_based network which called a 2-state Markov model (Kekre et al., 1977) as showed in Figure 4.2.

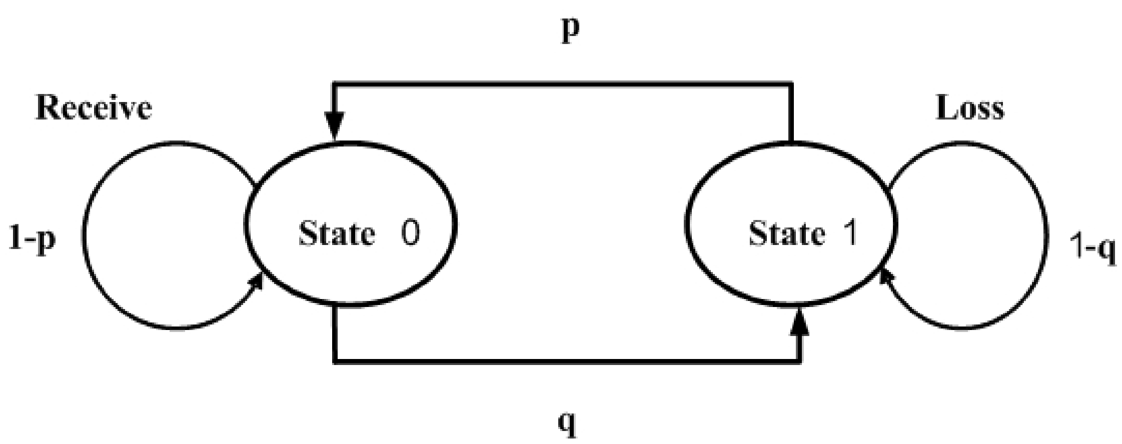

Figure 4.2: 2-State Markov Model

The 2-state Markov model depends on two parameters p and q where p is the probability of transition from found state to lost state and q is the probability of transition from lost state to found state.

Because the proposed method simulate network through providing packet loss using an equation from ITU-T Recommendation G.107 (2000) to derive p and q from the following equation 4.1:

$$Burst\_Ratio = \frac{1}{p+q} = \frac{Packet\_Loss}{p} = \frac{1-Packet\_Loss}{q} \tag{4.1}$$

From this equation p and q can be computed according to provided packet loss, these two values are used to simulate packet loss in a voice signal to simulate the network behavior.

4.1.2 Language

There are many languages in the world that possibly used over an IP network, different languages may have different effects on voice quality. To study this effect standard artificial voices from ITU-T Recommendation P.50 (1999) were used, where all available languages were utilized; i.e. a 20 languages of 16 male and 16 female, this study used one of both sex from each languages as illustrated in Table 4.1.

An artificial voice is a signal that is mathematically defined to reproduce the time and spectral characteristics of a human speech - over the bandwidth 100 Hz-8 kHz - which significantly affect the performances of linear and nonlinear telecommunication systems. The following time and spectral characteristics of real speech are reproduced by the artificial voice:

- long-term average spectrum.
- short-term spectrum.
- instantaneous amplitude distribution.
- voiced and unvoiced structure of speech waveform.
- syllabic envelope.

The artificial voice described in the recommendation is mainly used for objective evaluation of speech processing systems and devices, in which a single-channel signal with continuous activity (i.e. without pauses) is sufficient for measuring characteristics. e.g. evaluation of speech codec. The advantage of artificial voice is that it is more easily generated and have smaller variability than samples of real voice. The following are several examples of sentences from different languages taken from from ITU-T Recommendation P.50 (1999) appendix I:

Table 4.1: Test signals (ITU-T Recommendation P.50, 1999)

Directory		Filename	
Number	Language	Male	Female
1	*AMERICAN_ENGLISH*	*A_ENG_M*1.16*P*	*A_ENG_F*1.16*P*
2	ARABIC	*AR_M*2.16*P*	*AR_F*1.16*P*
3	*BRITISH_ENGLISH*	*B_ENG_M*1.16*P*	*B_ENG_F*1.16*P*
4	CHINESE	*CH_M*1.16*P*	*CH_F*1.16*P*
5	DANISH	*DA_M*1.16*P*	*DA_F*1.16*P*
6	DUTCH	*DUTCH_M*1.16*P*	*DUTCH_F*1.16*P*
7	FINNISH	*FI_M*1.16*P*	*FI_F*1.16*P*
8	FRENCH	*FR_M*1.16*P*	*FR_F*1.16*P*
9	GERMAN	*GER_M*1.16*P*	*GER_F*1.16*P*
10	GREEK	*GR_M*1.16*P*	*GR_F*1.16*P*
11	HINDI	*HIN_M*1.16*P*	*HIN_F*1.16*P*
12	HUNGARIAN	*HU_M*1.16*P*	*HU_F*1.16*P*
13	ITALIAN	*ITA_M*1.16*P*	*ITA_F*1.16*P*
14	JAPANESE	*JA_M*1.16*P*	*JA_F*1.16*P*
15	NORWEGIAN	*NOR_M*1.16*P*	*NOR_F*1.16*P*
16	POLISH	*PO_M*1.16*P*	*PO_F*1.16*P*
17	PORTUGUESE	*PORT_M*1.16*P*	*PORT_F*1.16*P*
18	RUSSIAN	*RU_M*1.16*P*	*RU_F*1.16*P*
19	SPANISH	*SP_M*1.16*P*	*SP_F*1.16*P*
20	SWEDISH	*SWE_M*1.16*P*	*SWE_F*1.16*P*

عَلى قَدرِ أَهلِ العَزمِ تَأتي العَزائِمُ ·

ala qadri ahli l'azmi ta'tī l'azā'imu

Figure 4.3: Arabic Text and Spelling

我去无锡市，他到黑龙江。沈阳旅大市，广州内蒙古。

Figure 4.4: Chinese Text

I saw it with my own eyes. The note was immediately dispatched. He wanted to leave college.

Figure 4.5: English Text

Mittwoch kommt uns der Besuch ja passend, denn dann bin ich noch nicht wieder naß geworden.

Figure 4.6: German Text

Alors la bise se mit à souffler de toutes ses forces; mais plus elle soufflait, plus le voyageur serrait son manteau autour de lui, et à la fin la bise renonça à le lui faire enlever.

Figure 4.7: French Text

Esa señora venía mucho a mi casa y, a veces, me ayudaba a blanquear y limpiar los armarios de la cocina.

Figure 4.8: Spanish Text

4.1.3 Gender

The study tested the genders male and female from the 20 languages where the effect of gender on voice quality is taken into account as will be seen in initial experiments in the section 4.3.

4.1.4 Burst_Ratio

A new parameter has an effect on voice quality has been appeared defined as a measure of burstiness in packet_based network called Burst_Ratio (J. W. McGowan, 2005) which is also defined as a ratio of the average length of observed bursts in a packet arrival sequences over the average length of bursts expected for a random loss in packet_based network, where the burst length refer to the number of packets in a single loss bursts. The burst ratio can be calculated from following equation (J. W. McGowan, 2005):

$$Burst\_Ratio = \frac{MBL_B}{MBL_R} = \frac{\left(\frac{\sum_{b=1}^{Number-of-Burst} Busrt-length_b}{Number-of-Burst}\right)}{\left(\frac{1}{1-L}\right)} \tag{4.2}$$

where

MBL	Mean Burst Length.
$L = \frac{t}{T}$	The proportion of lost packets that is a lost rate.
t	Number of Packets that are lost at the receiving end.
T	The sending packets in a given period of time.

It has long been observed that loss on packet_based network is bursty which means that the loss of a packet depends on lost previous packets. This study used equation provided in ITU-T Recommendation G.107 (2000) to compute the burst ratio according to p and q as follows:

$$Burst\_Ratio = \frac{1}{(p+q)} \tag{4.3}$$

Packet Loss in the experiments utilizes burst ratio percentage to compute p and q values in order to simulate degradation due to packet loss in a transmitted voice over packet_based network. The proposed simulator takes packet loss and burst ratio percentage as input parameters for equations that are illustrated later to compute p and q that used to insert loss in some positions in transmitted voice stream.

4.1.5 Codec

In computers, encoding is the process of putting a sequence of characters (letters, numbers, punctuation, and certain symbols) into a specialized format for efficient transmission or storage. Decoding is the opposite process the conversion of an encoded format back into the original sequence of characters. En**Co**ding and **dec**oding (Codec) are used in data communications, networking, and storage and they are the second and sixth steps in transmitting Voice over Internet Protocol (VoIP) network as illustrated in Figure 2.1. The ITU-T Recommendation G.723.1 (1996) CS-ACELP (Conjugate Structure Algebraic Codebook Excited Linear Prediction, 8 Kbps) and ITU-T Recommendation G.729 (1996) MP-MLQ/ACELP (Multipulse excitation with a maximum-likelihood quantizer/ Algebraic Codebook Excited Linear Prediction, Dual rate: 5.3/6.3 Kbps) are both standardized by ITU and have been used in this thesis for Codec, the two codec types belong to CELP (Codebook Excited Linear Prediction) analysis by synthesis hybrid codec. At each speech analysis frame, the speech signal is analyzed to extract the parameters of the CELP model (Linear Prediction, or LP filter coefficients, adaptive and fixed codebooks indices and gains). For stability and efficiency, LP filter coefficients are transformed into Line Spectral Frequencies, or LSFs for transmission. These parameters are then encoded and transmitted. At the decoder, the parameters are decoded and speech is synthesized by filtering the reconstructed excitation signal through the LP synthesis filter. The major differences between the two codecs lie in the excitation signals, the partitioning of the excitation space (the algebraic codebook), delay and the way in which the coefficients of the filter are represented. The frame information for G.729 and G.723.1 is shown at Table 4.2. The delay induced at encoder is referred as algorithmic delay.

Table 4.2: Frame information for G.729/G.723.1

Codec	Algorithm	Bit Rates (Kb/s)	Frame length	Look-ahead	Algorithmic delay
G.729	CS-ACELP	8	10 ms	5 ms	15 ms
G.723.1	MP-MLQ/ACELP	5.3/6.3	30 ms	7.5 ms	37.5 ms

Two codecs have voice activity detection and silence suppression processing. The frames are classified as normal speech frame, SID (Silence Insertion Description) frame and null frame (non-transmitted frame).

As discussed in this section voice quality can be affected by many parameters. In the next sections a series of experiments are conducted to study the effect of each parameter or combination of parameters on voice quality with the aim of finding accurate classifiers to predict voice quality for observed parameter values. These classifiers should be trained to be accurate where the attributes values act as inputs and the voice quality value expressed as PESQ MOS_{LQO} act as output or class label to the classifier system.

4.2 The Proposed Technique Simulation

The proposed technique as discussed in section 4.1 has many processes to build the training set that will be input to the classification methods for producing classifier to be used as model for predicting the future unknown voice quality value from a set of attributes values. These processes came from the step of VoIP network as in figure 2.1, these processes are as follows:

1. Encoding Process

   The transmitted signal needs to be encoded in IP-network in order to reduce the quantity of data that transmitted over the network, in other words reduce the channel bandwidth by compressing the data packets. Hence the encoding technique was used in the proposed simulator. Two encoding methods were used which are ITU-T Recommendation G.723.1 (1996) and ITU-T Recommendation G.729 (1996), these two coding technique written in C++ program that called with the original signal for compression as in the next calling:

   Compressed_Signal = EncoderG7321(Reference_Signal).

   Compressed_Signal = EncoderG729(Reference_Signal).

   These two calling methods will result with Compressed_Signal to be an input for

Packet Loss simulator as in the next process.

2. Simulate Packet Loss

   Providing packet loss simulator required deleting some packets from the voice signal with suitable probability, even though the p and q values should be calculated from provided packet loss and Burst Ratio probability as depicted in equations 4.1 and 4.3, the p and q used to put zeros to indicate packet loss in places through specific threshold which is the packet loss percentage, so the insertloss function called as follows:

   Degraded_Signal = insertloss(Compressed_Signal,p,q,Packet_Loss)

   The simulator of packet loss used values from 1 to 50 percent and five values from 1 to 5 percent for burst ratio to input loss in some place inside data stream.

3. Decoding process

   As known the received degraded signal at the end-side needs to be decompressed to retrieve the required original signal to playout at receiver buffer for listening, so the same two coder that were used for coding have an opposite process of encoding process that is a decoding process, hence the two decoder ITU-T Recommendation G.723.1 (1996) and ITU-T Recommendation G.729 (1996) were used for decompression as follows:

   Decompressed_Signal = DecoderG7321(Degraded_Signal).

   Decompressed_Signal = DecoderG729(Degraded_Signal).

   The Decompressed_Signal is a signal that will be listened to by the user at the receiver end, this decompressed and degraded signal due to packet loss signal will be compared with the original signal to compute the voice quality as objective estimation in the next step.

4. Perceptual Evaluation of Speech Quality (PESQ)

   PESQ that was standardized in ITU-T Recommendation P.862 (2001) is used to compute the voice quality as MOS value, the method needs two signals as input to make the comparison and compute the quality of degraded (decompressed) signal. The PESQ method written in C++ program with two input and one output, the output are MOS_{PESQ}, that is computed according to equation (3.8), that converted to MOS_{LQO} by using Equation (3.9) for narrow band audio systems Equation (3.10) for wide band audio systems, the calling function is as follows:

MOS_{PESQ} = PESQ(+8000,Reference_Signal,Decompressed_Signal).

The calling function contains three parameters two mentioned signal and +8000 this number means compute the quality at 8000 samples of streams.

5. Convert MOS_{PESQ} to MOS_{LQO}

   The PESQ result which is MOS_{PESQ} should be converted to MOS_{LQO} that is a quality estimation for perspective user, because MOS_{PESQ} as defined in P.862 was calibrated against an essentially arbitrary objective distortion scale and, hence, was not designed to be on exactly the same scale as MOS. For normal subjective test material, however, the PESQ output range will be a listening quality MOS-like score between 1.0 (bad) and 4.5 (no distortion) by using Equation (3.9).

These five processes were repeated 20,000 often to produce 20,000 records, i.e the combination of parameters which are 50 packet loss values, five burst ratio values, 20 language, two genders and two codecs. Each value of each parameter with MOS_{LQO} act as a record to training set where the programs in Appendix A explain the pseudocode for all previous processes.

4.3 Study The Effect of Each Parameter on Voice Quality

The building training set that was produced from the previous five processes produced 20,000 records that are used to study the effect of each parameters on voice quality independently. This section aims to find the effect of each parameter by determining the value of this parameter and fixing the other parameters values and using the regression methods which are the linear and non-linear regression (SPSS v.17, n.d.) to analyze each parameter effect on empirical voice quality.

There are two figures for each parameter that determine the relationship between the expect and empirical MOS values. The first figure shows the relation where the parameter value is shown in the x axis and the MOS value is shown in the y axis. The second figure determines the boxplot of difference in values between the expected and the empirical values, the boxplot presents five sample statistics - the minimum, the lower quartile, the median, the upper quartile and the maximum - in a visual display. The box of the plot is a rectangle which encloses the middle half of the sample, with an end at each quartile. The length of the box is thus the interquartile range of the sample. The other dimension of the box does not represent anything in particular. A line is drawn across the box at the sample median. Whiskers sprout from the two ends of the box until they reach the sample maximum and minimum. The crossbar at the far end of each whisker is optional and its length signifies nothing. Figure 4.9 shows a boxplot of a sample of 20 observations (actual sample values used in the display) together with a boxplot of the same data.

The next subsections study in detail the effect of each parameters on voice quality.

4.3.1 The Effect of Packet Loss

This subsection aims to study the effect of of different values of packet loss on voice quality, the parameter values of packet loss as mentioned are discrete values from 1 to 50 percent, these values are use to show the effect of packet loss on voice quality, it is expected that voice quality increases when the packet loss value decreases and visa versa i.e. inversely proportional relation. This inverse relationship can be concluded by using linear and non-linear regression methods where Equations (4.4) and (4.5) declare the equations to connect the packet loss values with MOS values with Arabic male and female languages through using g.729 codec.

$$MOS_{LQO} = -0.040 * PacketLoss + 2.969 \tag{4.4}$$

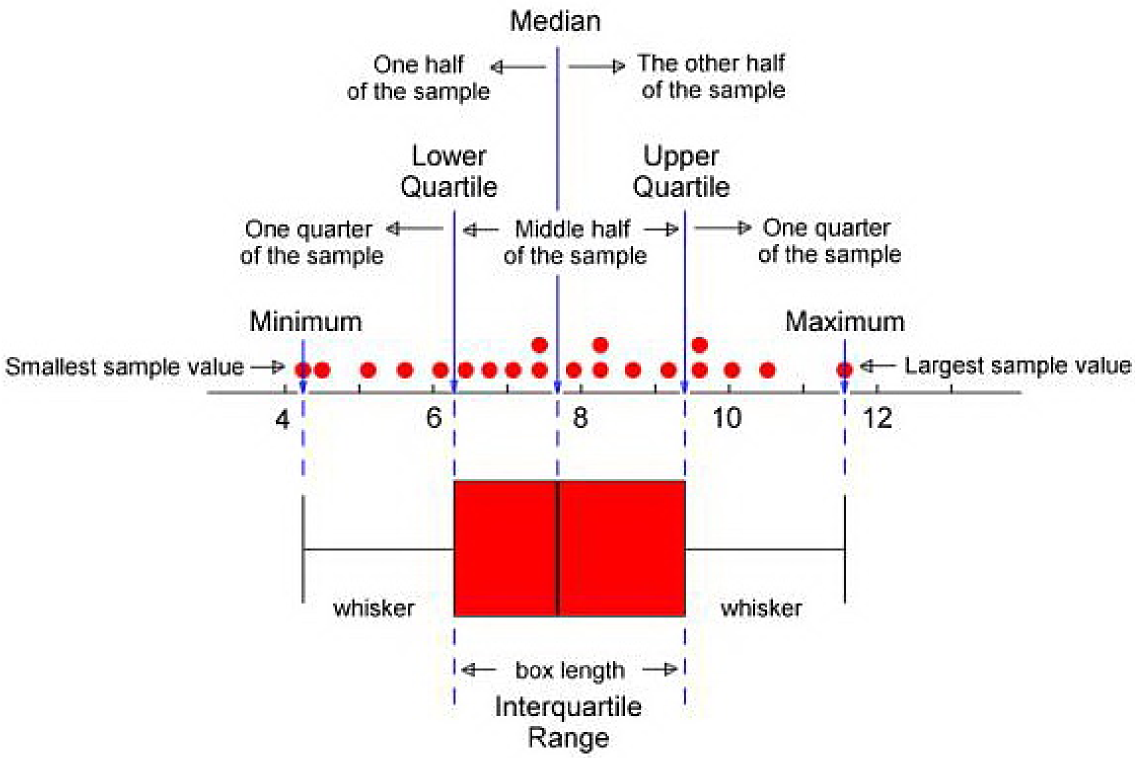

Figure 4.9: Example on boxplot

$$MOS_{LQO} = -0.091 * PacketLoss + 0.001 * PacketLoss^2 + 3.412 \tag{4.5}$$

Figure 4.10 illustrates the chart relationship between two values for linear and non-linear equations which shows that empirical MOS quality value decrease, when packet loss percentage values increase; i.e. inverse relation. From 2% to 12% packet loss percentage linear MOS is lower than empirical and non-linear MOS, but from 12% to 40 % packet loss percentage linear MOS is greater than empirical and non-linear MOS which is concluded that linear MOS has no effect in this interval. Finally from 40% to 50 % , packet loss linear and non-linear MOS values start to increase from empirical MOS values which explains the inverse relation between packet loss and empirical MOS values in this interval. Figure 4.10 shows that as packet loss increase, Empirical MOS values decrease. This inverse relation because of the increase lost of the packets on voice signal which has an effect on voice quality.

The difference absolute value between MOS empirical and MOS resulted from equa-

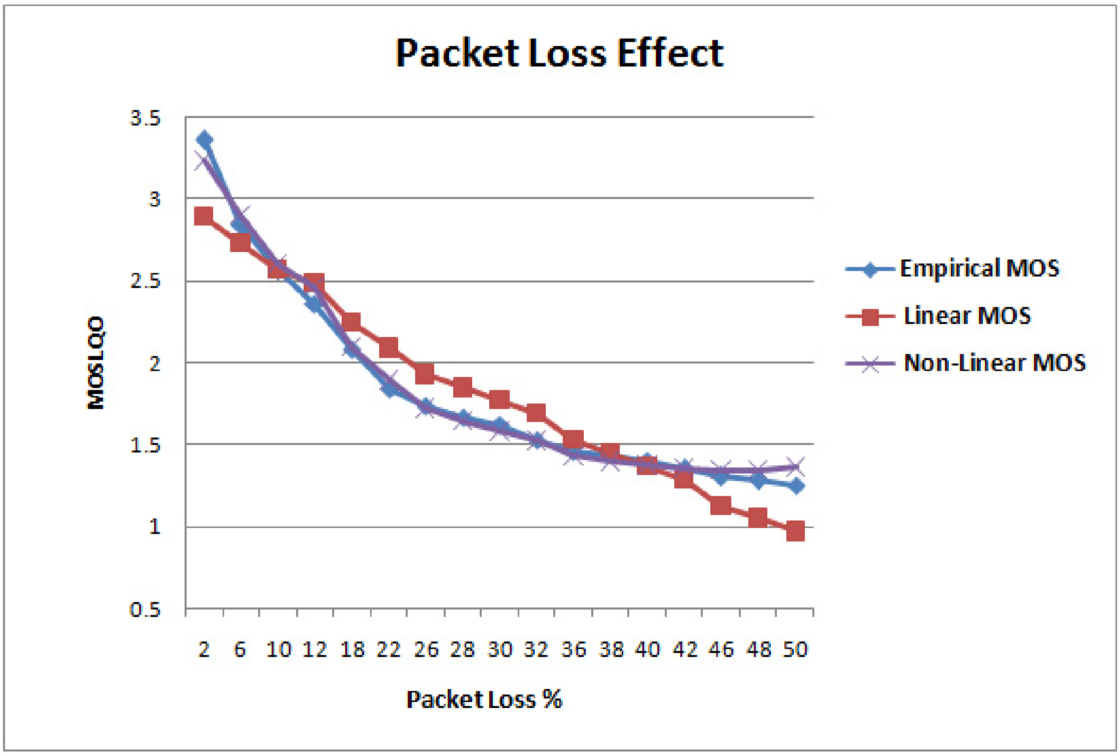

Figure 4.10: Packet Loss Effect Using Regression Methods

tions (4.4) and (4.5) is viewed in the boxplots shown in Figure 4.11.

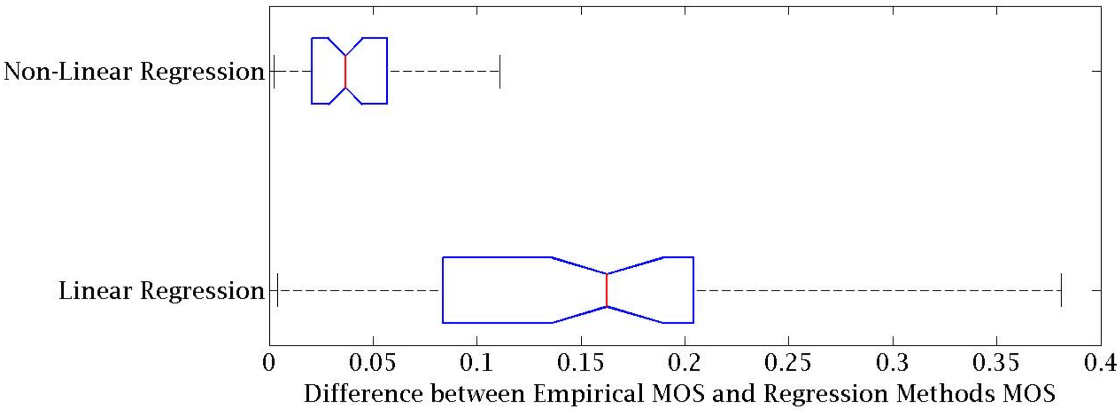

Figure 4.11: Boxplot of absolute values between MOS empirical and MOS resulted from equations of Packet Loss

The boxplots for linear and non-linear regression methods clarify the absolute differences between the Empirical MOS values and MOS values resulted from the two equations where the absolute differences are distributed in the range of 0 to 0.39 MOS for linear regression and 0 to 0.13 for non-linear regression. The first quartile (25% of data) of linear regression lie in the range of 0 to 0.09 and from 0 to 0.02 for non-linear regression, the two middle interquartiles (50% of data) of linear regression are in the range of 0.09 to 0.21 MOS and from 0.02 to 0.06 for nonlinear regression and the last quartile (25% of data) of linear regression between 0.21 and 0.39 MOS and 0.06 to 0.13 for non-linear regression. The boxplots explain that non-linear equation is more accurate than linear regression as it approximates the value of empirical MOS more accurately.

4.3.2 The Effect of Burst Ratio

Burst ratio has an effect on voice quality, so this effect is explained through using a training set to build the regression methods as in equations (4.6) and (4.7) where burst ratio acts as input and empirical MOS as output. The results of two equations are compared with empirical MOS values. burst ratio has many values, but this study used five discrete values ranging from 1 to 5, the experiments show that when the burst ratio is increased, the voice quality decreased as in packet loss.

$$MOS_{LQO} = -0.023 * BurstRatio + 3.511 \tag{4.6}$$

$$MOS_{LQO} = -0.069 * BurstRatio + 0.008 * BurstRatio^2 + 3.564 \tag{4.7}$$

Figure 4.12 explains the effect of burst ratio on voice quality by using linear and non-linear equations and plotting burst ratio values on x axis and MOS values on y axis. As shown in Figure 4.12 linear MOS values are lower than empirical and non-linear MOS values from 1% to 1.5% percentage of burst ratio, but non-linear MOS values decrease from 2% to 3% percentage, nevertheless empirical MOS values decrease from 3% to 4.5%. From 4.5% to 5% Empirical MOS values decreases from non-linear MOS values. It can be inferred that burst ratio percent from 3% to 4.5% may have an effect on empirical MOS voice quality. burst ratio has an inverse relation effect on voice quality and has more effect on voice quality than packet loss effect as illustrated in figure 4.12, because burst

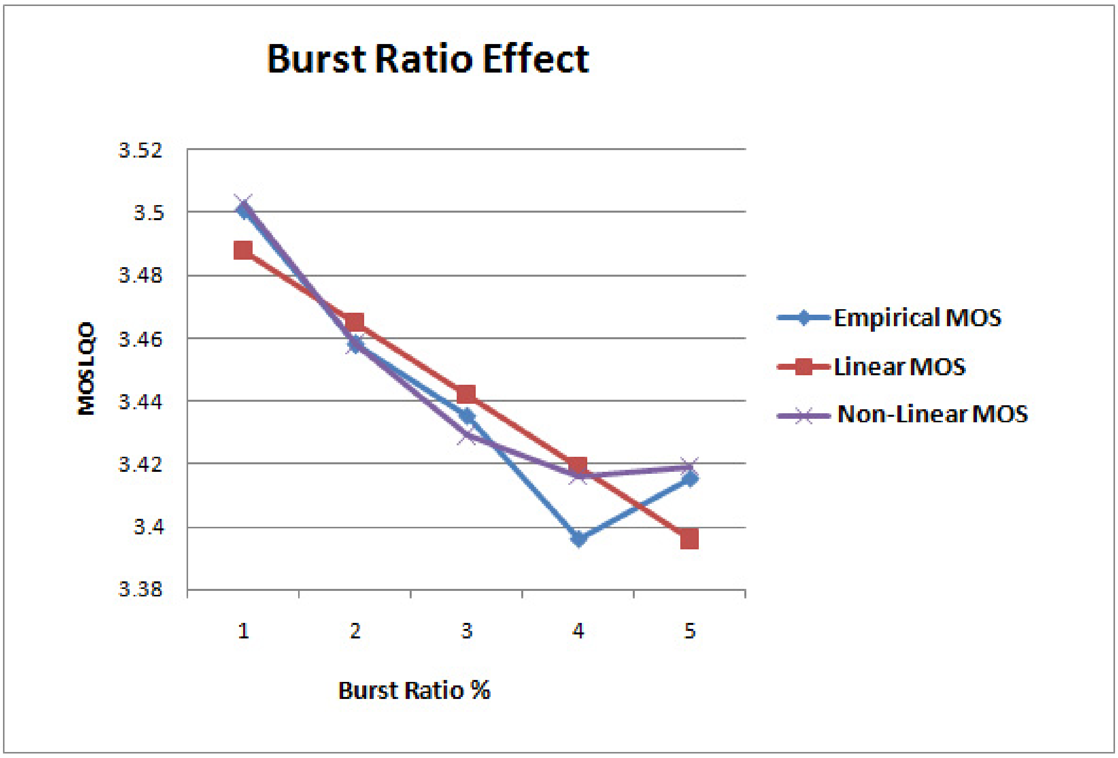

Figure 4.12: Burst Ratio Effect Using Regression Methods

ratio refers to ratio of the average length of observed bursts in a packet arrival sequences over the average length of bursts expected for a random loss in packet_based network, where the burst length refer to the number of packets in a single loss bursts.

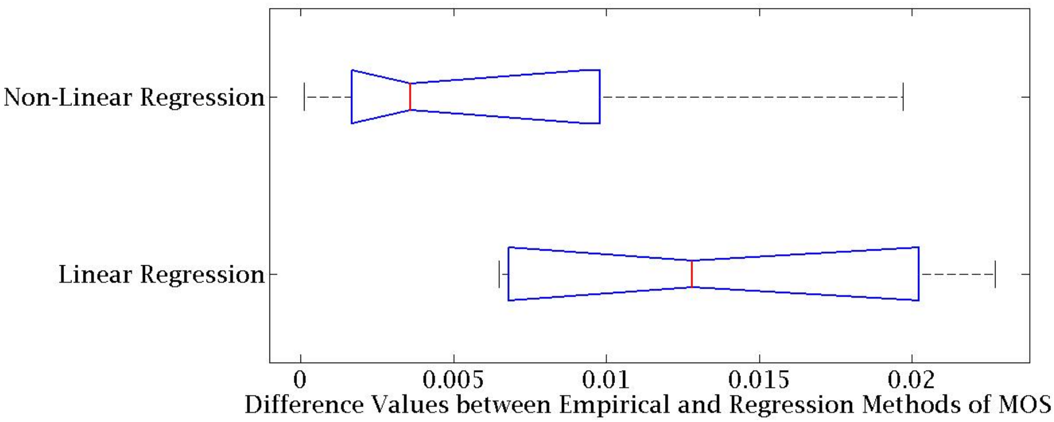

Figure 4.13: Boxplot of absolute values between MOS empirical and MOS resulted from equations of Burst Ratio

The difference values between empirical MOS and predicted MOS of regression methods is clarified through Figure 4.13. The boxplots for linear and non-linear regression clarify the absolute differences between the Empirical MOS values and MOS values resulted from the two equations where the absolute differences are distributed in the range of 0.006 to 0.035 MOS for linear regression and 0 to 0.02 for non-linear regression. The first quartile (25% of data) of linear regression lie in the range of 0.006 to 0.007 and from 0 to 0.004 for non-linear regression, the two middle interquartiles (50% of data) of linear regression are in the range of 0.007 to 0.02 MOS and from 0.004 to 0.01 for nonlinear regression and the last quartile (25% of data) of linear regression between 0.02 and 0.035 MOS and 0.01 to 0.02 for non-linear regression. The boxplots explain that non-linear equation is more accurate than linear regression as it approximates the value of empirical MOS more accurately.

4.3.3 The Effect of Language

Many languages were used in the experiment as in Table 4.1, the used language has an effect on voice quality, hence in this subsection the effect of language is studied. In the experiments the language parameter is represented as a set of number from 1 to 20 as shown in Table 4.1. From the training set, regression methods were derived to produce the linear and non-linear equations that link the MOS value with language number while other parameters are fixed . Equations (4.8) and (4.9) show the derived linear and non-linear regression equations respectively.

$$MOSLQO = 0.003 * Language + 3.547 \tag{4.8}$$

$$MOSLQO = -0.035 * Language + 0.002 * Language^2 + 3.686 \tag{4.9}$$

Figure 4.14 illustrates the relationship between empirical MOS and MOS produced from the regression methods. It can be noticed from the figure that the empirical MOS values do not have a linear or non-linear underlying relation which indicates that the language on its own can not be used to derive the quality. However, when the language is used with other parameters, it could have an effect on the quality. This is to be confirmed from the subsequent experiments. It can also be noticed that some languages have an overall quality less or more than others. Language number 7, for example, which is Finnish has less average quality than other languages. This can be referred to less tolerability (robustness) for the language to other parameters used in the experiment.

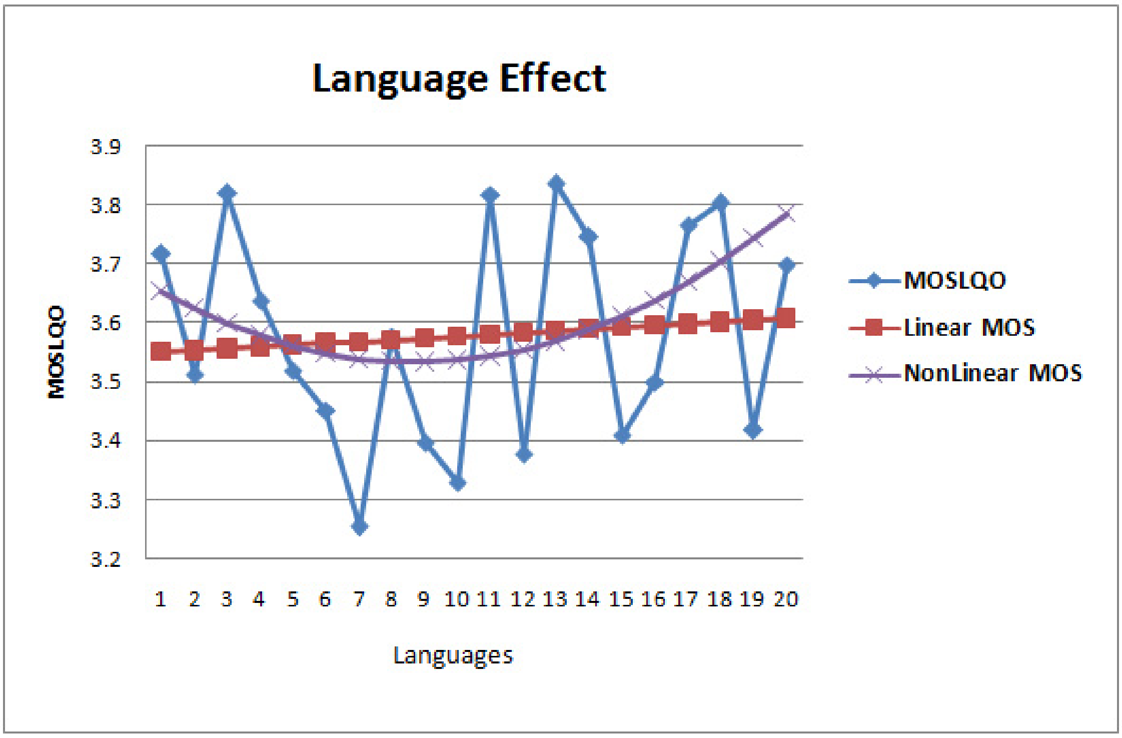

Figure 4.14: Language Effect Using Regression Methods

The difference between empirical MOS and the MOS values derived from the Regression methods is shown in Figure 4.15 which is indicates that the language effect on its own can not be used to infer the quality which is something consistent with our comments above.

The boxplots for linear and non-linear regression clarify the absolute differences between the Empirical MOS values and MOS values resulted from the two equations where the absolute differences are distributed in the range of 0 to 0.32 MOS for linear regression

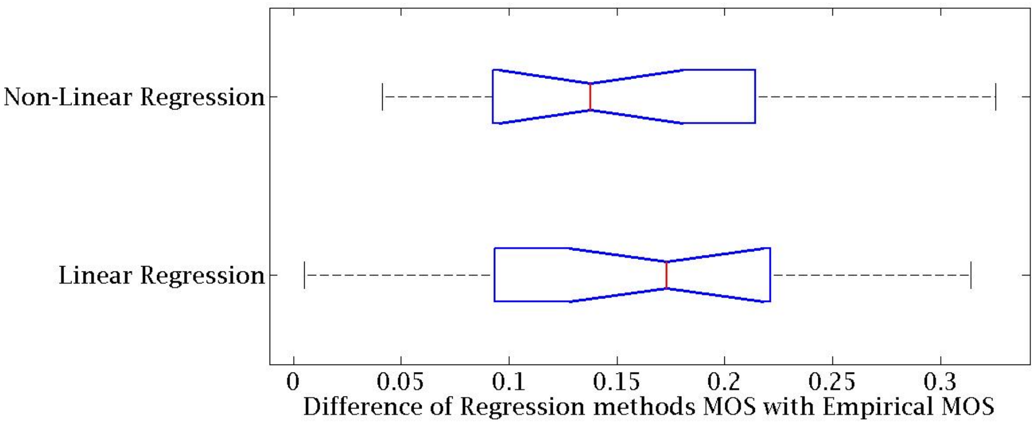

Figure 4.15: Boxplot of absolute values between MOS empirical and MOS resulted from equations of Language

and 0.04 to 0.35 for non-linear regression. The first quartile (25% of data) of linear regression lie in the range of 0 to 0.09 and from 0.04 to 0.09 for non-linear regression, the two middle interquartiles (50% of data) of linear regression are in the range of 0.09 to 0.23 MOS and from 0.09 to 0.21 for nonlinear regression and the last quartile (25% of data) of linear regression between 0.23 and 0.32 MOS and 0.21 to 0.35 for non-linear regression. The boxplots explain that non-linear equation is more accurate than linear regression as it approximates the value of empirical MOS more accurately.

4.3.4 The Effect of Gender Combined with Packet Loss

Based on our observation that the Language on its own can not be used to characterize the MOS value, we avoided to use the gender on its own as it has similar characteristics as the language, therefore, the effect of the gender is studied with the effect of packet loss. Each used language is studied with the male and female signals, the effect of the two genders on voice quality studied with statistical method after discretizing them as 1 female and 2 male as shown in equations (4.10) and (4.11) that define the relationship between the empirical MOS value and the gender value with fixed language (1), codec (2) and burst ratio (1).

$$MOSLQO = (-0.041) * PacketLoss + 0.134 * Gender + 2.840 \tag{4.10}$$

$$MOSLQO = (-5.813E+04) * Gender + 19377.410 * Gender^2 + (-8.937E-02) * PacketLoss + .001 * PacketLoss^2 + 38758.082 \tag{4.11}$$

Figure 4.16 and 4.17 depict the relationship between gender and MOS for regression methods and empirical MOS, where the x axis define the packet loss 1-50 and y axis define the different MOS values.

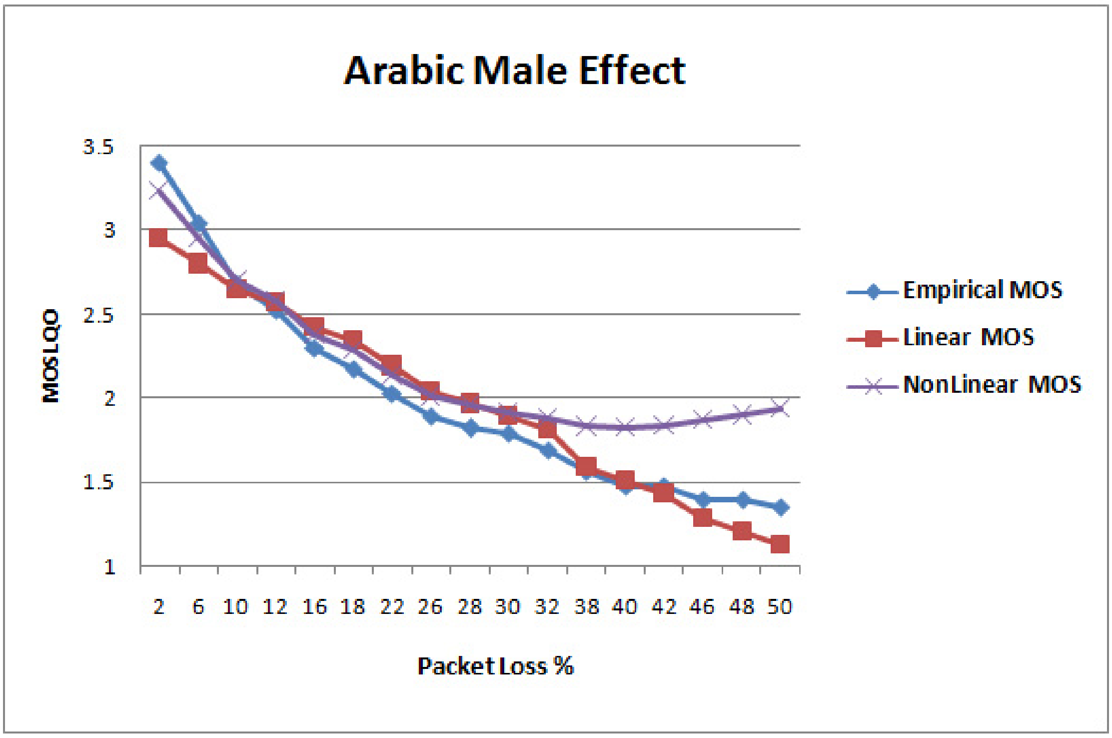

Figure 4.16: Male Gender Effect

In Figure 4.16 empirical MOS decreases in its' values from 2% to 50%. From 2% to 10% Empirical MOS values are upper than other regression methods of packet loss percentage with high quality MOS. From 10% to 32% of packet loss percentage empirical MOS values decrease lower than the two regression methods which inferred that male language speech has effect on voice quality as packet loss percentage increase and the relation that tied the empirical MOS and male language speech is squared relation as illustrated in Equation (4.10) and (4.11) because of empirical MOS line in Figure 4.16 upper than linear MOS line.

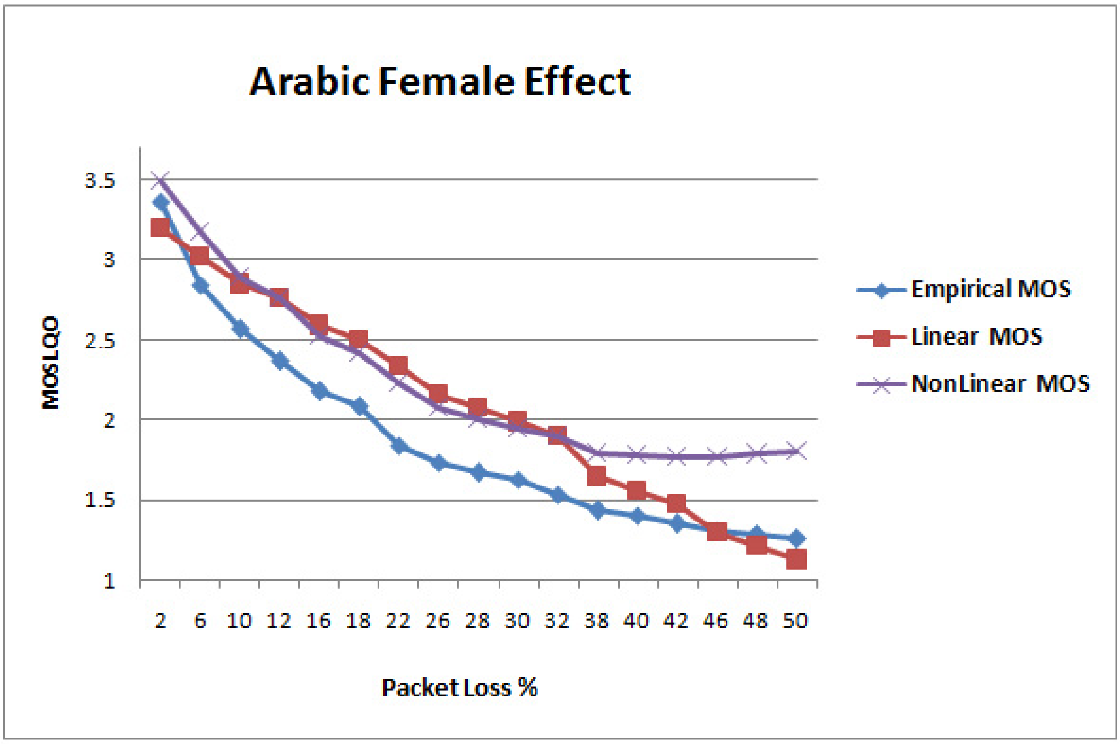

Figure 4.17: Female Gender Effect

As illustrated in male gender effect on voice quality, Figure 4.17 shows the effect of female gender on voice quality. As result of the effect of male gender on voice quality female gender has an effect on voice quality from 2% to 50% and the relation between empirical MOS and female gender are squared relation and as packet loss values increase empirical MOS values decreased.

From two previous figures, we can conclude that gender parameter has no effect on voice quality unless it is combined with other parameters such as packet loss as will illustrated in studying each parameters together.

The boxplot of differences between empirical MOS and the regression methods is shown in Figure 4.18. It can be noticed that the non-linear method has less deviation than the linear method.

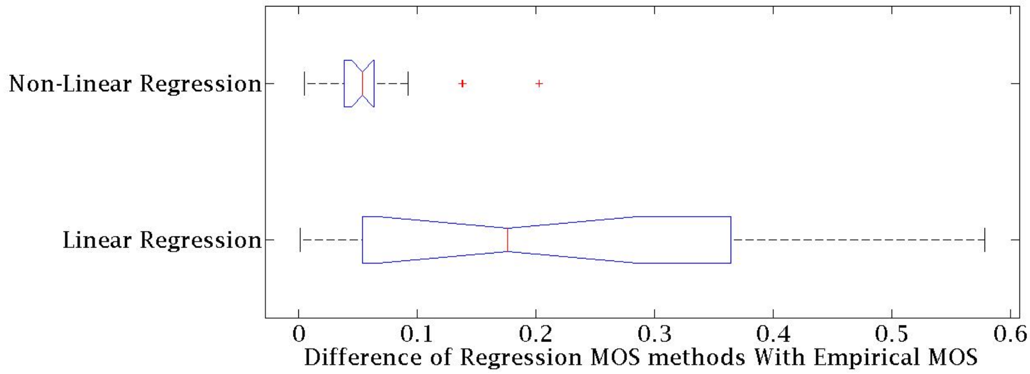

Figure 4.18: Boxplot of absolute values between MOS empirical and MOS resulted from equations of Gender

The boxplots for linear and non-linear regression clarify the absolute differences between the Empirical MOS values and MOS values resulted from the two equations where the absolute differences are distributed in the range of 0 to 0.59 MOS for linear regression and 0 to 0.09 for non-linear regression. The first quartile (25% of data) of linear regression lie in the range of 0 to 0.04 and from 0 to 0.02 for non-linear regression, the two middle interquartiles (50% of data) of linear regression are in the range of 0.04 to 0.39 MOS and from 0.02 to 0.05 for nonlinear regression and the last quartile (25% of data) of linear regression between 0.39 and 0.59 MOS and 0.05 to 0.09 for non-linear regression. The boxplots show that non-linear equation is more accurate than linear regression as it approximates the value of empirical MOS more accurately.

4.3.5 The Effect Of Codec Combined with Packet Loss

As mentioned earlier there two codec were used in the experiments (ITU-T Recommendation G.729, 1996) and (ITU-T Recommendation G.723.1, 1996). The used codec has an effect on transmitted voice quality. These two codecs are represented as two values 1 for (ITU-T Recommendation G.723.1, 1996) and 2 for (ITU-T Recommendation G.729,

1996) in the training set, they connected with packet loss through in following regression equations (4.12) and (4.13):

$$MOSLQO = (-0.420) * Codec + (-0.039) * PacketLoss + 3.611 \tag{4.12}$$

$$MOSLQO = (-9.021E-02) * packetloss + .001 * packetloss^2 + (-5.813E+04) * codec + 19377.246 * codec^2 + 38758.549 \tag{4.13}$$

Figures 4.19 and 4.20 depict the relationship between the codec and the MOS for regression methods and empirical MOS.

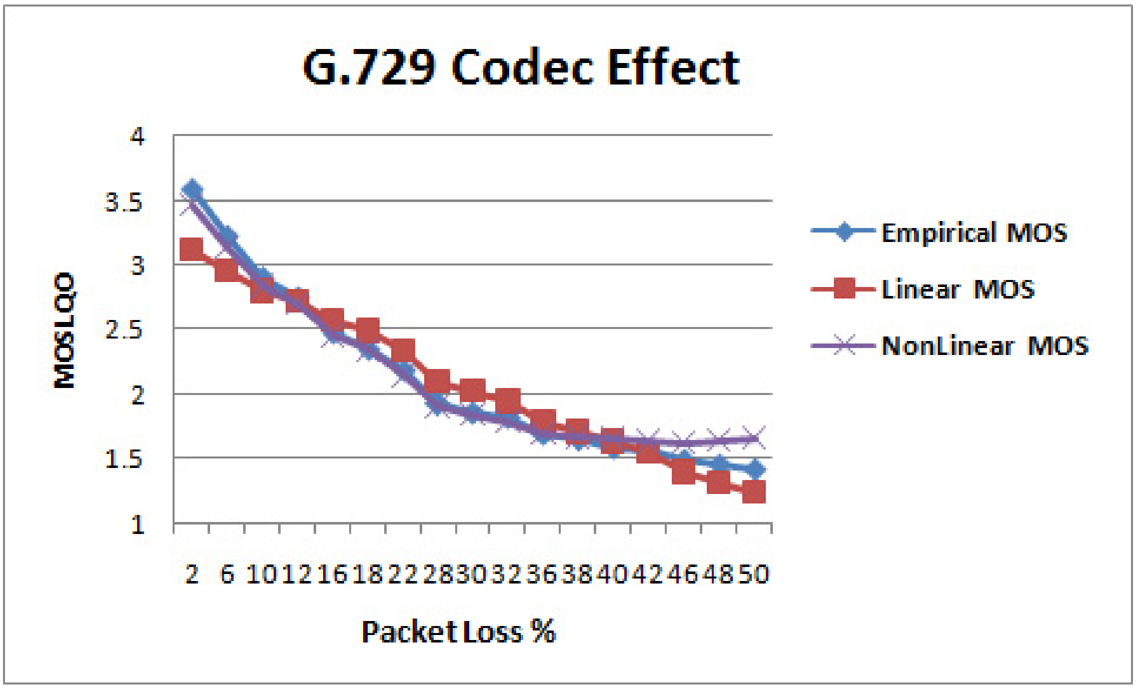

Figure 4.19: G.729 codec effect

The effect of ITU-T Recommendation G.729 (1996) is depicted in Figure 4.19 that shows that empirical MOS decreased when packet loss used with ITU-T Recommendation G.729 (1996) increased. From 40% to 50% packet loss the empirical MOS values under ITU-T Recommendation G.729 (1996) effect is decreased although it is closed with non-linear MOS from 2% to 40% packet loss percentage which is conclude that linear MOS is stable and has no effect on voice quality which leads to conclusion of empirical MOS has squared relation when combine the ITU-T Recommendation G.729 (1996) codec with packet loss effect.

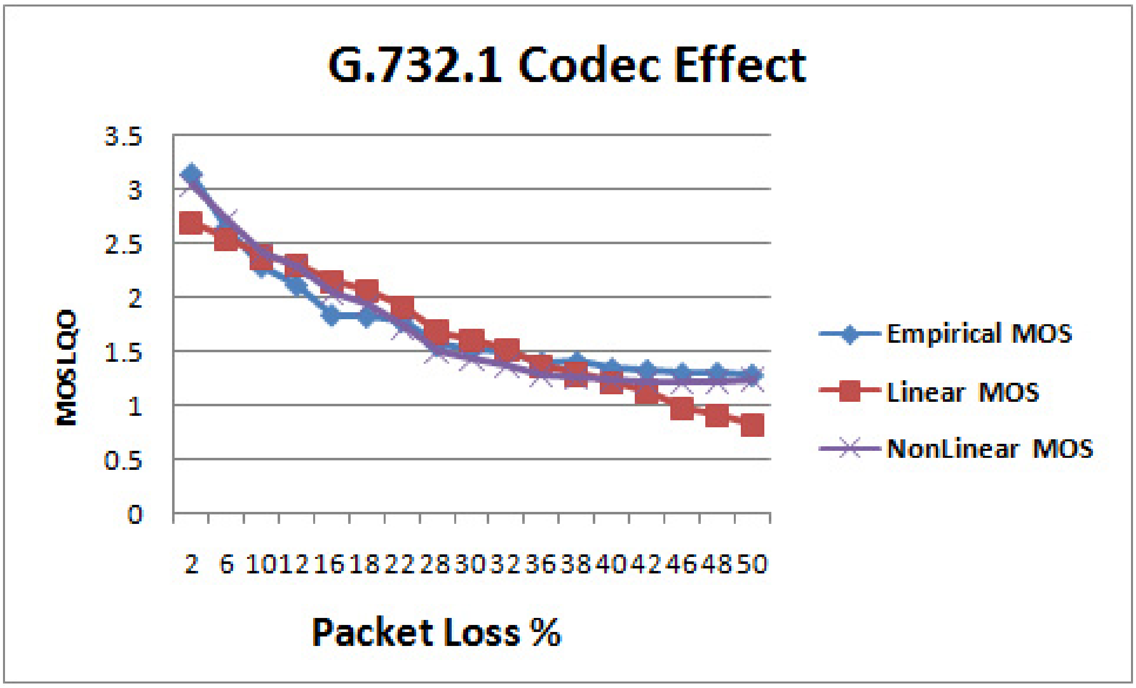

Figure 4.20: G.732.1 codec effect

In Figure 4.20 from 10% to 22% percentage packet loss, empirical MOS under the effect of ITU-T Recommendation G.723.1 (1996) codec decreases especially in 16% percentage effect, from 38% to 50% packet loss percentage the empirical MOS values increase which conclude that ITU-T Recommendation G.723.1 (1996) is efficient codec than ITU-T Recommendation G.729 (1996), when packet loss values increase.

The difference between empirical MOS and the regression methods is shown in the following Figure 4.21 which explains that Codec factor has low effect on voice quality.

The boxplots for Linear and Non-Linear regression clarify the absolute differences between the Empirical MOS values and MOS values resulted from the two equations where the absolute differences are distributed in the range of 0 to 0.48 MOS for linear regression and 0 to 0.28 for non-linear regression. The first quartile (25% of data) of

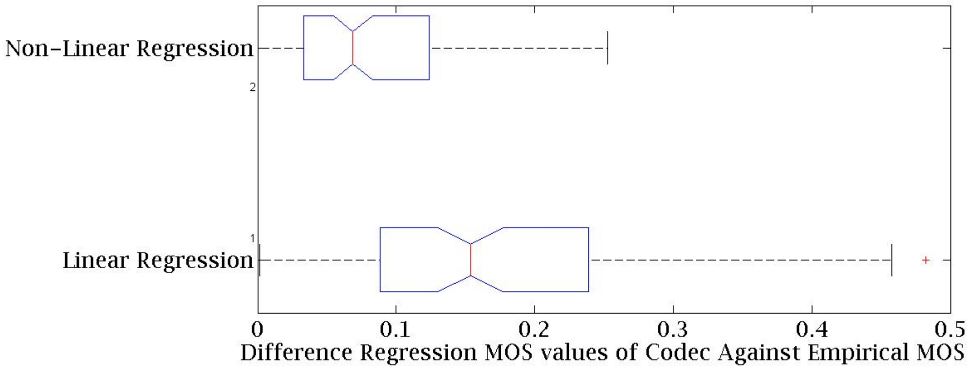

Figure 4.21: Boxplot of absolute values between MOS empirical and MOS resulted from equations of Codec

linear regression lie in the range of 0 to 0.08 and from 0 to 0.05 for non-linear regression, the two middle interquartiles (50% of data) of linear regression are in the range of 0.08 to 0.24 MOS and from 0.05 to 0.13 for nonlinear regression and the last quartile (25% of data) of linear regression between 0.24 and 0.48 MOS and 0.13 to 0.28 for non-linear regression. The boxplots explain that non-linear equation is more accurate than linear regression as it approximates the value of empirical MOS more accurately.

4.4 Study the Effect of All Parameters Together on Voice Quality

Each parameter has an effect on voice quality as discussed in the previous section, this section aimed to study the effect of all parameters together on voice quality by producing classifiers which are regression, ANN and Ant Miner to predict the values of MOS from all parameters values and clarify in details the accuracy of each classifier. As mentioned earlier 20,000 records of training set were established from the proposed system simulation, this training set will be used to clarify the effect of all parameters on voice quality, this study apply ten folds technique which means that the 20,000 records were divided into 10% of data as testing data and 90% of data as training, the next subsections explain the ten experiments for each classifier with continuous and discrete training data analysis in details:

4.4.1 Continuous Data Analysis

Three classification methods were used as classifier to predict the value of MOS which are regression method, including linear and non-linear regression and ANN. The study using the three methods with continuous data to compare these methods with each other. As Ant Miner methods can only be used with discrete data because it is used for classification not for regression, so it's class label defined as nominal (categorical) not numeric attribute, from this point of view it is excluded from this experiment.

Regression methods produce two equations for each fold, these equations relate the predicted MOS value with the five parameters values by using SPSS v.17 (n.d.). The two equations with different types, the first one was linear and the other was cubic polynomial equations. The ANN builds a network for each fold where the input layer contains the parameter values and the output layer includes the predicted MOS values with two hidden layers by using Matlab v.7.6.0.324 (n.d.). The following points explain one fold of the ten folds that is fold ten with its regression equation, ANN accuracy and the accuracy of each equations and ANN network against empirical MOS. The results for the other folds will be discussed in details in appendix B:

In fold 10, a linear and non-linear (cubic) equation and an ANN network were produced. The linear and non-linear equation for this fold are as follows:

$$MOSLinear = (-.033) * PacketLoss + (-.007) * Language + .006 * BurstRatio + (-.163) * Gender + .099 * Codec + 2.757 \quad (4.14)$$

$$
\begin{aligned}
MOSNonLinear = & (-1.217E-01)*PacketLoss + .003*PacketLoss^2 \\
& + (-2.455E-05)*PacketLoss^3 + (-1.128E-01)*Language \\
& + .011*Language^2 + (-3.102E-04)*Language^3 \\
& + (-1.394E-01)*BurstRatio + .043*BurstRatio^2 \\
& + (-3.700E-03)*BurstRatio^3 + (-4.134E+05)*Gender \\
& + (-2.369E+05)*Gender^2 + 160585.343*Gender^3 \\
& + (2.625E+06)*Codec + (-3.426E+05)*Codec^2 \\
& + (-2.282E+05)*Codec^3 + (-1.565E+06) \qquad (4.15)
\end{aligned}
$$

The statistical analysis for regression methods and ANN can be shown as in following Table 4.3:

Table 4.3: Fold 10 Statistical Analysis with Regression Methods and ANN

	Linear Regression	Non-Linear Regression	Artificial Neural Network
Average	0.2237	0.1547	0.1096
Standard Deviation	0.1730	0.1353	0.1009
Variation	0.0299	0.0183	0.0102
Max	1.0682	0.8363	0.6488
Min	0.0002	0.0003	0.0000

In order to clarify the accuracy of continuous data the following boxplot 4.22 explains the difference value of each MOS of classifiers against empirical MOS for fold 10.

As shown in boxplot 4.22 the difference values of ANN are closed to zero which explain that ANN classifiers is more accurate than other methods such as non-linear and linear regression then comes non-linear regression method to be the second accurate method to estimate the voice quality. So ANN and non-linear method are more accurate than linear regression where the absolute differences are distributed in the range of 0 to 0.65 MOS for linear regression , 0 to 0.5 for non-linear regression and 0 to 0.38 for ANN. The first quartile (25% of data) of linear regression lie in the range of 0 to 0.1, from 0 to

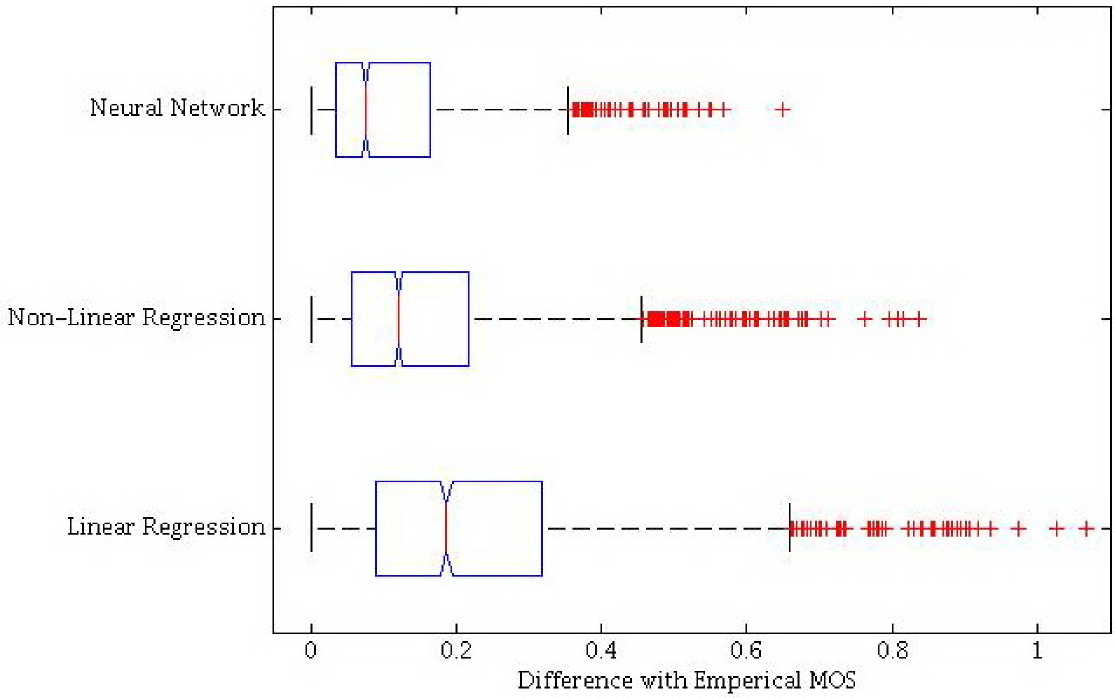

Figure 4.22: Boxplot for Fold 10

0.05 for non-linear regression and 0 to 0.03 for ANN, the two middle interquartiles (50% of data) of linear regression are in the range of 0.1 to 0.36 MOS, from 0.05 to 0.25 for nonlinear regression and 0.03 to 0.18 for ANN and the last quartile (25% of data) of linear regression between 0.36 to 0.65 MOS, 0.25 to 0.5 for non-linear regression and 0.18 to 0.38 for ANN.

Each fold has two regression linear and non-linear equations and ANN that estimate the empirical MOS with predicted MOS, from these ten folds, Table 4.4 clarifies the average difference of the ten statistical analysis to summarize the results exist in the tables available in Appendix B.

Table 4.5 shows the Standard Deviation difference of the ten folds. Table 4.6 shows the variation difference of the ten folds. Table 4.7 shows the Maximum difference of the ten folds. Table 4.8 shows the Minimum difference of the ten folds.

In Table 4.9, the correlation between each method with the empirical MOS values is listed for each fold. Then average over the 10 folds is also computed. From the Table, it

Table 4.4: Average Difference of Each Ten Folds

Average Difference Of Each ten Folds			
	Linear Regression	Non-Linear Regression	ANN
Fold1	0.2202	0.1483	0.0942
Fold2	0.2378	0.1554	0.1012
Fold3	0.2279	0.1519	0.0943
Fold4	0.2235	0.1505	0.0938
Fold5	0.2257	0.1531	0.1073
Fold6	0.2235	0.1493	0.1008
Fold7	0.2214	0.1516	0.0936
Fold8	0.2284	0.1561	0.0999
Fold9	0.2240	0.1509	0.0944
Fold10	0.2237	0.1547	0.1096

Table 4.5: Standard Deviation Difference of Each Ten Folds

Standard Deviation Difference Of Each ten Folds			
	Linear Regression	Non-Linear Regression	ANN
Fold1	0.1675	0.1326	0.0900
Fold2	0.1855	0.1372	0.0947
Fold3	0.1764	0.1329	0.0902
Fold4	0.1733	0.1321	0.0947
Fold5	0.1754	0.1366	0.1016
Fold6	0.1719	0.1293	0.0918
Fold7	0.1686	0.1316	0.0924
Fold8	0.1714	0.1380	0.0977
Fold9	0.1755	0.1350	0.0952
Fold10	0.1730	0.1353	0.1009

can be noticed that ANN method is most correlated method in comparison with empirical values with correlation of 0.970, non-linear regression and linear regression accuracy comes next with correlation of 0.933 and 0.863 respectively.

Table 4.6: Variation Difference of Each Ten Folds

Variation Difference Of Each ten Folds			
	Linear Regression	Non-Linear Regression	ANN
Fold1	0.0281	0.0176	0.0081
Fold2	0.0344	0.0188	0.0090
Fold3	0.0311	0.0177	0.0081
Fold4	0.0300	0.0175	0.0090
Fold5	0.0308	0.0187	0.0103
Fold6	0.0295	0.0167	0.0084
Fold7	0.0284	0.0173	0.0085
Fold8	0.0294	0.0190	0.0095
Fold9	0.0308	0.0182	0.0091
Fold10	0.0299	0.0183	0.0102

Table 4.7: Maximum Difference of Each Ten Folds

Maximum Difference Of Each ten Folds			
	Linear Regression	Non-Linear Regression	ANN
Fold1	1.1363	0.8414	0.6829
Fold2	1.1691	0.7738	0.5944
Fold3	1.1894	0.8023	0.6265
Fold4	1.0783	0.8823	0.5982
Fold5	1.0932	0.8527	0.7865
Fold6	1.1523	0.8361	0.6722
Fold7	1.0422	0.8574	0.6668
Fold8	1.0587	0.8510	0.6362
Fold9	1.1073	0.8486	0.6816
Fold10	1.0682	0.8363	0.6488

Table 4.8: Minimum Difference of Each Ten Folds

Minimum Difference Of Each ten Folds			
	Linear Regression	Non-Linear Regression	ANN
Fold1	0.0000	0.0000	0.0000
Fold2	0.0000	0.0001	0.0000
Fold3	0.0001	0.0000	0.0000
Fold4	0.0001	0.0001	0.0000
Fold5	0.0000	0.0001	0.0002
Fold6	0.0001	0.0000	0.0001
Fold7	0.0001	0.0001	0.0001
Fold8	0.0000	0.0001	0.0000
Fold9	0.0000	0.0003	0.0001
Fold10	0.0002	0.0003	0.0000

Table 4.9: The Pearson Correlation of each Classifier for Continuous Data Analysis

Pearson Correlation			
Method	Linear	Non-Linear	ANN
Fold1	0.866	0.933	0.972
Fold2	0.858	0.935	0.972
Fold3	0.859	0.933	0.973
Fold4	0.862	0.934	0.971
Fold5	0.863	0.932	0.965
Fold6	0.860	0.934	0.969
Fold7	0.869	0.934	0.972
Fold8	0.859	0.927	0.968
Fold9	0.871	0.937	0.973
Fold10	0.863	0.930	0.964
Average	0.863	0.933	0.970

4.4.2 Discrete Data Analysis

The empirical MOS values are continuous data as in each of ten folds in previous subsection that used classifiers with continuous data, in this study each training data used with two forms continuous and discrete analysis, continuous analysis was discussed in preceding subsection, so in this subsection the empirical MOS in training data was discretized by dividing the range of MOS values into intervals or scales where each value of empirical MOS took a discrete value or scale form these intervals as shown in Table 4.10.

Table 4.10: Empirical MOS Scaling

Interval	Scale
$MOS \leq 1$	Bad
$1 < MOS \leq 1.25$	Poor1
$1.25 < MOS \leq 1.5$	Poor2
$1.5 < MOS \leq 1.75$	Poor3
$1.75 < MOS \leq 2$	Poor4
$2 < MOS \leq 2.25$	Fair1
$2.5 < MOS \leq 2.5$	Fair2
$2.5 < MOS \leq 2.75$	Fair3
$2.75 < MOS \leq 3$	Fair4
$3 < MOS \leq 3.25$	Good1
$3.25 < MOS \leq 3.5$	Good2
$3.5 < MOS \leq 3.75$	Good3
$3.75 < MOS \leq 4$	Good4

The discrete training data means that the empirical MOS values were discretized, then ten fold procedure applied to the discrete training data, but the difference from continuous training data that the Ant Miner classifiers were used in discrete training data. The Ant miner classifiers were used in this type of training data which includes Ant miner, $_{C}Ant-Miner$ and $_{C}Ant-Miner$ type 2 with 60 ants value for the number of ants that anticipate in discovering the required rules. Because the continuous class label in continuous training data which is a MOS value do not fit with Ant Miner, the Ant miner classifiers used discrete class label.

The following points explain one fold of the ten folds that is fold ten with its Ant-Miner variation rules (Ant-Miner, ${}_{C}Ant-Miner$ and ${}_{C}Ant-Miner$ type 2). The results for the other folds will be discussed in details in appendix C:

In fold 10, Ant Miner variation rules were produced. The Ant-Miner, ${}_{C}Ant-Miner$ and ${}_{C}Ant-Miner$ type 2 rules are as follows:

1. Ant Miner

   IF Gender = 2 THEN MOS= Poor2

   IF Packet_Loss = 1 THEN MOS= Good2

   IF Packet_Loss = 2 THEN MOS= Good1

   IF Codec = 2 THEN MOS= Poor2

   IF Packet_Loss = 3 THEN MOS= Fair4

   IF Packet_Loss = 4 THEN MOS= Fair4

   IF Language = 3 AND Packet_Loss = 5 THEN MOS= Good2

   IF Packet_Loss = 5 THEN MOS= Fair4

   IF Packet_Loss = 6 AND Language = 11 THEN MOS= Good1

   IF Packet_Loss = 6 THEN MOS= Fair3

   IF Packet_Loss = 7 THEN MOS= Fair3

   IF Burst_Ratio = 2 THEN MOS= Poor2

   IF Burst_Ratio = 5 AND Packet_Loss = 9 THEN MOS= Fair4

   IF Burst_Ratio = 3 THEN MOS= Poor2

   IF Burst_Ratio = 4 THEN MOS= Poor2

   IF Packet_Loss = 8 THEN MOS= Fair3

   IF Burst_Ratio = 1 THEN MOS= Poor2

   IF Packet_Loss = 11 THEN MOS= Fair2

   IF Packet_Loss = 10 THEN MOS= Fair2

   IF Packet_Loss = 15 THEN MOS= Fair1

   IF Packet_Loss = 18 THEN MOS= Fair1

IF Packet_Loss = 13 THEN MOS= Fair1

IF Packet_Loss = 20 THEN MOS= Fair1

IF Packet_Loss = 17 THEN MOS= Fair1

IF Packet_Loss = 14 THEN MOS= Fair1

IF Packet_Loss = 12 THEN MOS= Fair1

IF Packet_Loss = 23 THEN MOS= Poor4

IF Packet_Loss = 25 THEN MOS= Poor4

IF Packet_Loss = 21 THEN MOS= Poor4

IF Language = 11 THEN MOS= Poor4

IF Packet_Loss = 19 THEN MOS= Poor4

IF Packet_Loss = 22 THEN MOS= Poor4

IF Packet_Loss = 26 THEN MOS= Poor4

IF Packet_Loss = 16 THEN MOS= Poor4

IF Packet_Loss = 33 THEN MOS= Poor3

IF Packet_Loss = 29 THEN MOS= Poor3

IF Packet_Loss = 30 THEN MOS= Poor3

IF Packet_Loss = 28 THEN MOS= Poor3

IF Language = 1 THEN MOS= Poor3

IF Packet_Loss = 24 THEN MOS= Poor3

IF Language = 3 THEN MOS= Poor3

IF Packet_Loss = 32 THEN MOS= Poor3

IF Packet_Loss = 35 THEN MOS= Poor3

IF Packet_Loss = 31 THEN MOS= Poor3

IF Packet_Loss = 27 THEN MOS= Poor3

IF Packet_Loss = 37 THEN MOS= Poor3

IF Packet_Loss = 34 THEN MOS= Poor3

IF Packet_Loss = 46 THEN MOS= Poor2

IF Language = 18 THEN MOS= Poor3

IF Packet_Loss = 43 THEN MOS= Poor2

IF Packet_Loss = 44 THEN MOS= Poor2

IF Language = 17 THEN MOS= Poor3

IF Packet_Loss = 40 THEN MOS= Poor2

IF Packet_Loss = 42 THEN MOS= Poor2

IF Packet_Loss = 47 THEN MOS= Poor2

IF Packet_Loss = 41 THEN MOS= Poor2

IF Packet_Loss = 50 THEN MOS= Poor2

IF Language = 13 THEN MOS= Poor3

IF Packet_Loss = 39 THEN MOS= Poor2

IF Packet_Loss = 48 THEN MOS= Poor2

IF Packet_Loss = 45 THEN MOS= Poor2

IF Packet_Loss = 38 THEN MOS= Poor2

IF Packet_Loss = 49 THEN MOS= Poor2

IF Burst_Ratio = 5 THEN MOS= Poor2

2. $_{C}AntMiner$

IF Packet_Loss $\geq$ 26.5 THEN MOS= Poor2

IF Burst_Ratio $\geq$ 1.5 AND Packet_Loss $\geq$ 14.5 THEN MOS= Poor3

IF Codec = 1 AND Burst_Ratio $<$ 1.5 AND Packet_Loss $\geq$ 11.5 THEN MOS= Poor2

IF Packet_Loss $<$ 5.5 THEN MOS= Good1

IF Gender = 2 AND Packet_Loss $\geq$ 8.5 THEN MOS= Poor4

IF Gender = 1 AND Packet_Loss $<$ 10.5 THEN MOS= Fair3

IF Gender = 1 AND Burst_Ratio $\geq$ 1.5 THEN MOS= Fair1

IF Codec = 2 AND Packet_Loss $<$ 17.5 THEN MOS= Fair2

IF Language = 4 AND Burst_Ratio $\geq$ 1.5 THEN MOS= Fair4

IF Packet_Loss $\geq$ 7.5 THEN MOS= Poor4

IF Burst_Ratio $<$ 1.5 THEN MOS= Poor3

IF Language = 1 THEN MOS= Fair3

IF Language = 9 THEN MOS= Fair3

IF Language = 14 AND Burst_Ratio $\geq$ 2.5 THEN MOS= Fair3

IF Language = 2 THEN MOS= Fair3

IF Packet_Loss $<$ 6.5 THEN MOS= Fair1

IF Burst_Ratio $<$ 4.5 THEN MOS= Fair1

IF Gender = 2 THEN MOS= Poor4

3. $_{C}AntMiner$ type2

IF Burst_Ratio $\geq$ 1.5 AND Packet_Loss $\geq$ 21.5 THEN MOS= Poor2

IF Codec = 1 AND Packet_Loss $\geq$ 21.5 THEN MOS= Poor1

IF Packet_Loss $\geq$ 10.5 THEN MOS= Poor3

IF Burst_Ratio $\geq$ 1.5 AND Packet_Loss $\geq$ 4.5 THEN MOS= Fair2

IF Codec = 1 AND Packet_Loss $\geq$ 4.5 AND Gender = 2 THEN MOS= Poor3

IF Gender = 2 AND Codec = 1 THEN MOS= Fair1

IF Gender = 1 AND Burst_Ratio $\geq$ 1.5 THEN MOS= Good1

IF Gender = 2 THEN MOS= Fair4

IF Packet_Loss $<$ 5.5 AND Codec = 2 THEN MOS= Good2

IF Codec = 1 AND Packet_Loss $\geq$ 5.5 THEN MOS= Fair2

IF Gender = 1 AND Codec = 2 THEN MOS= Fair3

IF Packet_Loss $\geq$ 3.5 THEN MOS= Fair3

IF Packet_Loss $<$ 1.5 THEN MOS= Good1

IF Packet_Loss $\geq$ 2.5 THEN MOS= Good1

IF Codec = 1 THEN MOS= Fair4

In order to compute the accuracy of each of the six classifiers, comparisons are made between the classifiers' MOS values and the empirical MOS, from these comparisons, the accuracy of each ten folds were computed as shown in Table 4.11.

Table 4.11: The Accuracy of Each Classifier

Accuracy						
	Linear	Non-Linear	ANN	Ant-Miner	CAnt-Miner	CAnt-Miner2
Fold1	33.48%	49.38%	63.07%	33.83%	45.13%	40.03%
Fold2	30.23%	49.28%	62.72%	33.98%	45.48%	41.83%
Fold3	34.08%	46.83%	65.37%	32.78%	44.18%	42.13%
Fold4	32.78%	47.73%	64.87%	32.58%	43.63%	38.53%
Fold5	33.58%	49.28%	58.52%	33.63%	45.13%	42.28%
Fold6	33.03%	46.63%	59.77%	32.53%	45.38%	41.73%
Fold7	32.38%	49.23%	63.87%	32.33%	42.33%	41.18%
Fold8	31.83%	48.18%	61.87%	33.33%	43.08%	43.23%
Fold9	32.78%	47.43%	63.97%	33.38%	44.83%	41.98%
Fold10	35.53%	46.98%	59.77%	33.08%	43.28%	41.03%
Average	32.97%	48.10%	62.38%	33.15%	44.25%	41.40%

From Table 4.11 the ANN seems the most accurate classifier with 62.38%, the Ant-miner classifier is not accurate in comparison with the most accurate classifier, but the Ant-miner classifiers has advantages which include readability where the user can read the resulted rules and usability which means that the user can use these rules through editing them.

In Table 4.12, the correlation between each method with the empirical MOS values is listed for each fold. Then average over the 10 folds is also computed. From the Table, it can be noticed that ANN method is most correlated method in comparison with empirical values with correlation of 0.955, non-linear regression, $_{C}Ant-Miner$, $_{C}Ant-Miner$ type 2, linear regression and Ant Miner accuracy comes next with correlation of 0.917, 0.875, 0.866, 0.848 and 0.470 respectively.

Table 4.12: The Pearson Correlation of each Classifier for Discrete Data Analysis

Pearson Correlation						
Method	Linear	Non-Linear	ANN	Ant-Miner	CAnt-Miner	CAnt-Miner2
Fold1	0.848	0.918	0.956	0.437	0.869	0.824
Fold2	0.842	0.920	0.958	0.483	0.872	0.883
Fold3	0.846	0.916	0.960	0.489	0.879	0.879
Fold4	0.844	0.916	0.955	0.457	0.872	0.836
Fold5	0.850	0.917	0.949	0.438	0.878	0.873
Fold6	0.846	0.916	0.952	0.501	0.875	0.882
Fold7	0.853	0.918	0.957	0.510	0.875	0.865
Fold8	0.845	0.910	0.952	0.413	0.872	0.877
Fold9	0.855	0.920	0.958	0.474	0.873	0.878
Fold10	0.855	0.915	0.949	0.497	0.880	0.865
Average	0.848	0.917	0.955	0.470	0.875	0.866

Chapter 5

Conclusions And Future Work

5.1 Conclusions

Merging the data and voice traffic media within VoIP application has many advantages, nevertheless several challenges may arise which decreases the quality of VoIP traffic because of the nature of IP network for real time transmission of data rather than voice traffic. This challenging motivates researchers to propose new techniques to assess the quality in VoIP networks for legal, commercial and technical reasons; measuring the quality also helps VoIP service providers to evaluate their own and their competitors' service using a standard scale and to achieve customers' satisfaction.

Several methods have been proposed to estimate the quality: subjective and objective methods. As speech quality is a subjective as it is determined by the listener's perception, the most reliable approach for assessing speech quality is through subjective tests. Although subjective tests is a reliable technique but it is expensive, time-consuming and needs strict lab conditions. On another hand, Objective speech quality assessment methods base their measures on objective metrics of physical parameters and properties of the speech signal by building an algorithmic model to assess the voice quality.

In order to study the effect of some parameters on speech quality, 5 parameters such as packet loss, burst ration, language, gender and codec was selected and a training set of 20000 record were built, such that selected parameters with its' values present the attributes for the training set and voice quality which preforms with MOS_{LQO} as class label for the training set. The effect of 5 parameters were studied by initial experiments where the effects of each parameter on voice quality studied and presented, and by studying the effects of each parameters on voice quality together through two data analysis continuous and discrete; continuous data analysis combines linear and non-linear as regression methods and ANN as classifiers to predict their correlation against empirical voice quality prediction. As a result ANN method is most correlated method in comparison with empirical values with correlation of 0.970, non-linear regression and linear regression accuracy comes next with correlation of 0.933 and 0.863 respectively. As Ant Miner methods can only be used with discrete data because it is used for classification not for regression, so it's class label defined as nominal (categorical) not numeric attribute, from this point a discrete data analysis discretized the MOS_{LQO} value in training data by dividing the range of MOS_{LQO} values into intervals or scales where each value of MOS_{LQO} took a discrete value or scale form. In discrete data analysis six classifiers linear and non-linear regression methods, ANN and a proposed technique Ant Miner with its' variations, including: original Ant Miner, an extension to Ant-Miner called *c*Ant-Miner and an extension to *c*Ant-Miner called *c*Ant-Miner type 2 were used and compared with each other to assess their accuracy in predicting the voice quality. The ANN was the most accurate classifier with 62.38%, the Ant-miner classifier is not accurate in comparison with the most accurate classifier, but the Ant-miner classifiers has advantages which include readability where the user can read the resulted rules and usability which means that the user can use these rules through editing them. The Pearson correlation of each six classifiers ANN method is most correlated method in comparison with empirical values with correlation of 0.955, non-linear regression, $_{C}Ant-Miner$, $_{C}Ant-Miner$ type 2, linear regression and Ant Miner accuracy comes next with correlation of 0.917, 0.875, 0.866, 0.848 and 0.470 respectively.

5.2 Future Work

This thesis studied several parameters that have an effect on transmitted voice over IP network such as packet loss, language, gender, burst ratio and codec, also a set of classifiers were studied to determine their accuracy on estimating the voice quality such as linear and non-linear regression, ANN and Ant Colony Optimization technique called Ant Miner with its' variations, including: original Ant Miner, an extension to Ant-Miner called *c*Ant-Miner and an extension to *c*Ant-Miner called *c*Ant-Miner type 2. Future work will be discussed in this section.

Several improvement and direction are possible as future work:

1. Other parameters than the studied five parameters can be studied to determine their effect on voice quality independently and together. Possible parameters that classified as signal and parametric will studied to determine their effects on voice quality such as delay, jitter, echo and loudness.

2. Three Ant miner methods were studied original Ant Miner, *c*Ant-Miner and *c*Ant-Miner type 2. Ant Miner variant of ACO which is one of swarm Intelligence subfields that classified within Artificial Intelligence (AI) discipline. Swarm Intelligence has several methods such as bird flocking, animal herding, bacterial growth, bee colony optimization and fish schooling, these methods will be tested to estimate the voice quality and compare them with the proposed methods.

3. Combining two methods to compute the voice quality such as using Genetic Algorithm (GA) to select the optimal weights values for the training of ANNs and using ACO with ANN to find the best paths for the estimation.

4. Select set of features from a training set by using ACO which captures the optimal features that will be used by the methods to estimate the voice quality.

5. Study the effect of different network devices on voice quality where the types of devices and their values will be an input for the proposed methods.

References

AL-Akhras, M. (2007), Qualitlty of Media Traffic over Lossy Internet Protocol Networks: Measurement and Improvment, PhD thesis, **De Montfort University**.

AL-Akhras, M., Zedan, H., John, R. and ALMomani, I. (2009), 'Non-intrusive speech quality prediction in VoIP networks using a Neural network approach', ***Neurocomputing*** (72), 2595–2608.

Aleksander, I. and Morton, H. (1990), *An introduction to neural computing*, 2 edn, Van Nostrand Reinhold Co., New York, NY, USA.

Campbell, D., Jones, E. and Glavin, M. (2009), 'Audio quality assessment techniques-A review, and recent developments', ***signal processing*** **89**(8), 1489 – 1500.

Collins, D. (2003), *Carrier Grade Voice over IP*, 2nd edn, McGraw-Hill Companies.

Da Silva, A., Varela, M., de Souza e Silva, E., Rosa, L. and G.Rubino, G. (2008), 'Quality assessment of interaction voice applications', ***Computer Networks*** **52**(6), 1179–1192.

Deneubourg and Goss (1990), 'Collective patterns and decision making', ***Ethology, Ecology and Evolution*** **1**, 295–311.

Ding, L., Lin, Z., Radwan, A., El-Hennawey, M. S. and Goubran, R. A. (2007), 'Non-intrusive single-ended speech quality assessment in VoIP', ***Speech Communication*** **49**(6), 477–489.

Dorigo, M. and Gambardella, L. M. (1997), 'Ant Colony System: A Cooperative Learning Approach to the Traveling Salesman Problem', ***Evolutionary Computation, IEEE Transactions on*** **1**(1), 53–66.

Dorigo, M., Maniezzo, V. and Colorni, A. (1991), Positive Feedback as a Search Strategy, Technical report.

Dorigo, M., Maniezzo, V. and Colorni, A. (1996), 'The Ant System: Optimization by a colony of cooperating agents', ***IEEE Transactions on Systems, Man, and Cybernetics-Part B*** **26**(1), 29–41.

Fausett, L. (1994), *Fundamentals of Neural Networks: Architectures, Algorithms and Applications*, Prentice-Hall, New Jersey.

Grassé, P. P. (1959), 'La reconstruction du nid et les coordinations interindividuelles chez Bellicositermes netelensis et Cubitermes sp. La théorie de la stigmergie: Essai d'interpretation du comportement des termites constructeurs', ***Insectes Sociaux*** **6**, 41–83.

Grothmann, R. (2002), Multi-agent market modeling based on neural networks, Technical report, **Faculty of Economics, University of Bremen**.

Haupt, R. L. and Haupt, S. E. (2004), *Practical Genetic Algorithms with CD-ROM*, **Wiley-Interscience**.

Haykin, S. (1994), *Neural Networks: A Comprehensive Foundation*, 2 edn, **Macmillan**, New York.

Itakura, F. (1975), 'Minimum prediction residual principle applied to speech recognition', ***Acoustics, Speech and Signal Processing, IEEE Transactions on*** **23**(1), 67 – 72.

Itakura, F. and Saito, S. (1978), 'Analysis synthesis telephony based on the maximum likelihood method', ***Acoust. Speech Signal Process*** pp. C17–C20.

ITU Recommendation P.800.1 (1994), 'Terms and definitions related to quality of service and network performance including dependability', ***International Telecommunication Union-Telecommunication Standardization Sector (ITU-T),Geneva*** .

ITU-T Recommendation G.107 (2000), 'The E-model, a computational model for use in transmission planning', ***International Telecommunication Union-Telecommunication Standardization Sector (ITU-T)***,*Geneva* .

ITU-T Recommendation G.723.1 (1996), 'Dual rate speech coder for multimedia communications transmitting at 5.3 and 6.3 kbit/s', ***International Telecommunication Union-Telecommunication Standardization Sector (ITU-T)*** .

ITU-T Recommendation G.729 (1996), 'Coding of speech at 8 kbit/s using conjugate-structure algebraic-code-excited linear prediction (CS-ACELP)', ***International Telecommunication Union-Telecommunication Standardization Sector (ITU-T)*** .

ITU-T Recommendation P.50 (1999), 'Objective measuring apparatus', ***International Telecommunication Union-Telecommunication Standardization Sector (ITU-T)***,*Geneva* .

ITU-T Recommendation P.563 (2004), 'Single-ended method for objective speech quality assessment in narrow-band telephony applications', ***International Telecommunication Union-Telecommunication Standardization Sector (ITU-T)***,*Geneva* .

ITU-T Recommendation P.800 (1996), 'Methods for subjective determination of transmission quality', ***International Telecommunication Union-Telecommunication Standardization Sector (ITU-T)***,*Geneva* .

ITU-T Recommendation P.861 (1996), 'Objective quality measurement of telephone-band (300 - 3400 Hz) speech codecs', ***International Telecommunication Union-Telecommunication Standardization Sector (ITU-T)***,*Geneva* .

ITU-T Recommendation P.862 (2001), 'Perceptual evaluation of speech quality (PESQ):An objective method for end-to-end speech quality assessment of narrow-band telephone networks and speech codecs', ***International Telecommunication Union-Telecommunication Standardization Sector (ITU-T)***,*Geneva* .

J. W. McGowan (2005), 'Burst Ratio: A Measure of Bursty Loss on Packet-Based Networks', ***United States Patent*** **B2**(6,931,017).

Karapantazis, S. and Pavlidou, F.-N. (2009), 'Voip: A comprehensive survey on a promising technology', ***Computer Networks*** **53**(12), 2050 – 2090.

Karjalainen, M. (1985), 'A new auditory model for the evaluation of sound quality of audio systems', ***Acoustics, Speech, and Signal Process*** pp. 608–611.

Kekre, H., C, S. and H, S. (1977), 'A two-state Markov model of speech in conversation and its application to computer communication systems', ***Computers and Electrical Engineering*** **4**(2), 133–141.

Kitawaki, N., Nagabuchi, H. and Itoh, K. (1988), 'Objective quality evaluation for low-bit-rate speech coding systems', ***IEEE Journal on Selected Areas in Communications*** **6**(2), 242–248.

Mahdi and Picoviciv (2009), 'Advances in voice quality measurement in modern telecommunications', ***Digital Signal Processing*** **19**, 79–103.

Matlab v.7.6.0.324 (n.d.).

URL: *www.mathwork.com*

Nygren, K. (2004), Stock Prediction - A Neural Network Approach , Master's thesis, **Royal Institute of Technology, KTH**.

Otero, E., Freitas, A. and Johnson, G. (2008), 'cAnt-Miner: An Ant Colony Classification Algorithm to Cope with Continuous Attributes', ***Springer-Verlag Berlin Heidelberg*** pp. 48–59.

P.862.1, I.-T. R. (2003), 'Mapping function for transforming P.862 raw result scores to MOS-LQO', ***International Telecommunication Union-Telecommunication Standardization Sector (ITU-T),Geneva*** .

P.862.2, I.-T. R. (2005), 'Wideband extension to recommendation P.862 for the assessment of wideband telephone networks and speech codecs', ***International Telecommunication Union-Telecommunication Standardization Sector (ITU-T),Geneva*** .

Parpinelli, R. S., Lopes, H. S. and Freitas, A. A. (2002), 'Data Mining With an Ant Colony Optimization Algorithm', ***IEEE Transaction on Evolutionary Compuring*** **6**(4), 321–332.

Quackenbush, S., Barnawell, T. and Clements, M. (1988), *Objective Measures of Speech Quality*, **Prentice Hall**, Englewood Cliffs, NJ.

Raja, A., Azad, R. M. A., Flanagan, C., Picovici, D. and Ryan, C. (2006), Non-Intrusive Quality Evaluation of VoIP Using Genetic Programming, *in* '**BIONETICS '06: Proceedings of the 1st international conference on Bio inspired models of network, information and computing systems**', ACM, New York, NY, USA, p. 31.

Raja, A., Azad, R. M. A., Flanagan, C. and Ryan, C. (2007), *Genetic Programming*, Vol. 4445, chapter Real-Time, Non-intrusive Evaluation of VoIP, pp. 217–228.

Raja, A., Azad, R. M. A., Flanagan, C. and Ryan, C. (2008), VoIP speech quality estimation in a mixed context with genetic programming, *in* '**GECCO '08: Proceedings of the 10th annual conference on Genetic and evolutionary computation**', ACM, New York, NY, USA, pp. 1627–1634.

Raja, A. and Flanagan, C. (2008), *Genetic Programming*, chapter Real-Time, Non-intrusive Speech Quality Estimation: A Signal-Based Model, pp. 37–48.

Rango, F., Tropea, M., Fazio, P. and Marano, S. (2006), 'Overview on VoIP: Subjective and Objective Measurement Methods', ***International Journal of Computer Science and Network Security*** **6**(1), 140–153.

Rix, A. W., Beerends, J. G., Hollier, M. P. and Hekstra, A. P. (2001), Perceptual Evaluation of Speech Quality (PESQ) A new method for speech quality assessment of telephone networks and codecs, *in* '**ICASSP '01: Proceedings of the Acoustics, Speech, and Signal Processing, 200. on IEEE International Conference**', IEEE Computer Society, Washington, DC, USA, pp. 749–752.

Rix, W. and Hollier, P. (2000), 'The perceptual analysis measurement system for robust end-to-end speech quality assessment', ***Acoustics, Speech, and Signal Process*** pp. 1515–1518.

Schroeder, M. R., Atal, B. S. and Hal l, J. L. (1979), 'Optimizing digital speech coders by exploiting masking properties of the human ear', ***Acoustical Society of America Journal*** **66**(6), 1647–1652.

SPSS v.17 (n.d.).

URL: *www.SPSS.com*

Stützle, T. and Hoos, H. (2000), 'MAX-MIN Ant System', ***Future Generation Computing System*** **16**(9), 889–914.

Sun,L. (2004), Speech Quality Prediction For Voice over Internet Protocol Networks, Technical report, Doctor of Philosophy: University of Plymouth.

Sun,L. and Ifeachor,E. C. (2002), 'Perceived Speech Quality Prediction for Voice Over IP-based Networks', ***In IEEE International Conference on communications*** **4**, 2573–2577.

Voran, S. D. and States., U. (1998), *Objective Estimation of Perceived Speech Quality Using Measuring Normalizing Blocks*, **U.S. Dept. of Commerce, National Telecommunications and Information Administration** .

Wang, S., Sekey, A. and Gersho, A. (1992), 'An objective measure for predicting subjective quality of speech coders', ***Selected Areas in Communications, IEEE Journal on*** **10**(5), 819–829.

Witten, H. and Frank, E. (2005), *Data Mining: Practical Machine Learning Tools and Techniques*, Morgan Kaufmann Series in Data Management Systems, second edn, Morgan Kaufmann.

Yang, W. (1999), Enhanced Modified Bark Spectral Distortion (EMBSD):An Objective Speech Quality Measure Based On Audible Distortion And Cognition Model, PhD thesis, Philadelphia, PA, USA. Chair-Yantorno, Robert.

Yang, W., Dixon, M. and Yantorno, R. (1997), A modified bark spectral distortion measure which uses noise masking threshold, *in* '**Speech Coding For Telecommunications Proceeding, 1997, 1997 IEEE Workshop on**', pp. 55 –56.

Appendix A

The Proposed Technique Pseudocode

The built training data which used to produced classifiers comes with five processes as discussed in chapter 4 to produce 20,000 records. These steps are explained in the following program 3. These processes inferred from the VoIP system, so it is important to clarify each process as program calling as in the next program that written in MATLAB Matlab v.7.6.0.324 (n.d.):

A.1 Initialization Variables

In this code there are many variables used for building the training data, first previous variables that stored in MATLAB are cleared then MATLAB window is cleaned by clc command. Some parameters are then declared to help in storing the results such as Starttime=datestr(now) that compute the start time of program running, MOSoriginal=[] which is an array of 30 MOS values, PESQoriginal=[] like MOSoriginal which is an array of 30 PESQ values, nIterations=30 that is a value for 30 iteration obtained from the pilot study (AL-Akhras, 2007), Result=zeros(20000,7) this variable initialize an array with zeros values for 20,000 Records and 7 columns where the 20,000 records contain the parameters values that effect the voice quality that are packet loss, language, burst ratio, gender and codec values and the MOS value computed from converting the PESQ MOS value into

Program 3 The Proposed Method Pseudocode, Copyright ©Rami Al-Khawaldeh-2010

```
function AntColonyOptimizationForVoIPQualityAssessments
% Initialization Variables
clear    %Clear the workspace Buffer
clc  %Clear windows command of Matlab
% Declaration some of arrays and Variables
Starttime=datestr(now) %Compute the current time of system as start value
MOSoriginal=[]; %buffer of 30 MOS  value
PESQoriginal=[]; %Buffer of 30 PESQ value
nIterations=30;         %Value obtained from the Pilot study
Result=zeros(20000,7);  %Array of 20,000 Record and 7 column
i=1;  %i which is an index that control the decide the current record.
% Starting the process of building training data
for Codec=1:2  %Codec 1: G.723.1   2: G.729
for Gender=1:2 %Gender 1: Female 2: Male
for Burst_Ratio=1:5 %Burst_Ratio [1,2,3,4,5]
for Language=1:20 %Language There are 20 language from ITU-T P.50
%Reference Voice: Language + Gender
% Coder : Coder + Language + Gender
     L=num2str(Language); %Convert Number to string
     G=num2str(Gender);   %Convert Number to string
     C=num2str(Codec);    %Convert Number to string
     S=[L G '.16p'];  %Reference Voice
     B=[C L G '.BIT'];% Coder
     original=read16bitPCM(S);  %Reading the Original Language
     bitstream=read16bitPCM(B); %Reading the coding value of Original Language
     for Packet_Loss=1:50       %Packet Loss [1,2,...,50]
  poriginal=p(Packet_Loss,Burst_Ratio);
  qoriginal=q(Packet_Loss,Burst_Ratio);
  for iteration=1:nIterations % 30 iterations
degradedcopy=insertloss(bitstream,poriginal,qoriginal,Packet_Loss);
if Codec==1
   [PESQval1,MOSval1]=computeMOSofG7231(degradedcopy,S);
else
   [PESQval1,MOSval1]=computeMOSofG729(original,degradedcopy);
end
 PESQoriginal=[PESQoriginal PESQval1];
 MOSoriginal=[MOSoriginal MOSval1];
 end % nIterations
 if Codec==1
delete 'degraded.16p';
 end
%Compute the mean and Standard deviation for 30 PESQ and MOS values
PESQMean=mean(PESQoriginal);
PESQStd=std(PESQoriginal);
MOSMean=mean(MOSoriginal);
MOSStd=std(MOSoriginal);
% Saving the result of each iteration
 Result(i,1:7)=[Packet_Loss,Language,Burst_Ratio,Gender,Codec,PESQMean,MOSMean];
save('Result'); %Save the current value of result matrix when it takes new value
MOSoriginal=[];
PESQoriginal=[];
i=i+1;
```

MOS Listening Quality Estimation (MOS_{LQE}) value, seven columns contains the attribute labels such as packetloss, language, burstratio, gender, codec and MOSLQE and the i variable which is an index that control the decide the current record.

A.2 Starting the process of building training data

There are 5 For loop statements that determine the values for each parameters 2 for gender, 2 for codec, 5 for burst ratio measured in percent, 20 for language where each language has a number to refer to it and 50 values measured in percent for packet loss. Firstly the voice signal is coded and read from calling function called read16bitPCM as follows:

original=read16bitPCM(S); %Reading the Original Language

bitstream=read16bitPCM(B); %Reading the coding value of Original Language

The read16bitPCM is shown in the following program 4:

Program 4 The read16bitPCM Pseudocode calling function

```
function speechstream=read16bitPCM(sfilename)

%  read16bitstream reads 16 bit pcm binary file and returns the resultant
%  vector
%  The input is a filename containing the 16 bit pcm data
%  The output is a vector containing the data

%open the input filename
f_speech = fopen(sfilename,'r');
%read the whole speech file with int16 (2-bytes per element) precision
speechstream=fread(f_speech,inf,'int16');
speechstream=int16(speechstream);
fclose(f_speech);
```

In order to compute p and q values to insert loss in read streams the following calling programs were used:

poriginal=p(Packet_Loss,Burst_Ratio);

qoriginal=q(Packet_Loss,Burst_Ratio);

The two values take packet_loss and Burst_Ratio as input and poriginal and qoriginal as output the p and q function are as follow programs 5,6:

Program 5 The p Pseudocode calling function

```
function pout=p(Ppl,BurstR)

% P calculates the value of p which is the transition probability between
% "found" and "loss" state whereas q is the transition probability of moving
% between "loss" and "found" state in a 2-state markov model used to characteriz
% packet loss as described in the ITU-T recommendation G.107 (03/2005) Sec 3.5.
%
%
%The equations used to derive p in the recommendation are:
%
%   BurstR = 1/(p+q) = (Ppl/100)/p = (1-(Ppl/100))/q
%
%  From the above equations we derived the formula for p and q as following
%       p=Ppl/(100.BurstR),  q=(100-Ppl)/(100.BurstR)
%
%  The input are Packet-loss probability and Burst Ratio
%
%  The output is p, the transition probability between "found" and "loss"
%  states

pout=Ppl/(100*BurstR);
```

After computing the two values of p and q, it is the time to insert loss for coded bitstream by calling the insertloss function as follows:

degradedcopy=insertloss(bitstream, poriginal, qoriginal, Packet_Loss);

The reference parameters are bitstream, poriginal, qoriginal and packet_loss where degradedcopy is the output of the function. The psedocode of insertloss is as follow program 7: After insert loss in bitstream, this process repeated 30 times as in pilot study to compute accurate MOS when loss is inserted randomly, through this process the study decode the resulting degradedcopy by using one of decoder and compute the average of MOS as a final results as following calling functions:

$[PESQval1, MOSval1]$=computeMOSofG7231(degradedcopy,S);

$[PESQval1, MOSval1]$=computeMOSofG729(original,degradedcopy);

These two calling functions decode the degradedcopy of signal then compute PESQ MOS

Program 6 The q Pseudocode calling function

```
function qout=q(Ppl,BurstR)

% P calculates the value of p which is the transition probability between
% "found" and "loss" state whereas q is the transition probability of moving
% between "loss" and "found" state in a 2-state markov model used to characterize
% packet loss as described in the ITU-T recommendation G.107 (03/2005) Sec 3.5.
%
%
%The equations used to derive p in the recommendation are:
%
%   BurstR = 1/(p+q) = (Ppl/100)/p = (1-(Ppl/100))/q
%
%  From the above equations we derived the formula for p and q as following
%       p=Ppl/(100.BurstR),  q=(100-Ppl)/(100.BurstR)
%
%  The input are Packet-loss probability and Burst Ratio
%
%  The output is p, the transition probability between "found" and "loss"
%  states

qout=(1-Ppl/100)/BurstR;
```

values and convert the value into MOS_{LQE} where the psedocode of computeMOSofG7231 and computeMOSofG729 are as follow programs 8,9:

A.3 Saving the result of each iteration

Finally the value of MOS for each combination of the five parameters is averaged over 30 iterations and stored as one record in training data set by saving it in a result array and initialize each array and variables for next iterations as follows:

Result(i,1:7)=[Packet_Loss,Language,Burst_Ratio,Gender,Codec,PESQMean,MOSMean];

save('Result'); %Save the current value of result matrix when it takes new value

MOSoriginal=[]; PESQoriginal=[]; i=i+1;

Program 7 The InsertLoss Pseudocode calling function

```
function loss_stream=insertloss(stream,p,q,losspercent)

%Insert packet loss into the stream according to 2-state markov model where
%p represents the transition probability between "found" and "loss" and q
%represents the transition probability between "loss" and "found"
%
%"found" will be represented by while "loss" will be represented by 0

nFrames=floor(size(stream,1)/82+0.5);%# of frames (82 values/frame)
loss_stream=stream;
choose=rand();
state=1;

if choose<=losspercent/100
     state=0;
     loss_stream(3:82)=0;
end

for i=2:nFrames
   choose=rand();
   if state==1%found
       if choose<p %move to loss state
           state=0;%loss
           loss_stream((i-1)*82+3:i*82)=0;
       end
   elseif state==0%currently in loss state
       if choose<q %move to found state
           state=1;%found
       else%stay in loss state
           loss_stream((i-1)*82+3:i*82)=0;
       end
   end
end
```

Program 8 The computeMOSofG7231 Pseudocode calling function

```
function [PESQ_Score,MOS]=computeMOSofG7231(degraded,S)
%Compute the MOS value according to the PESQ
%ITU-T recommendation P.862 (02/2001).
%ITU-T recommendation P.862.1 (11/2003).
%ITU-T recommendation P.862 ammendment 2 (11/2005).
g723DECOD(degraded);
[PESQ_Score,MOS]=PESQMain(S);
```

Program 9 The computeMOSofG729 Pseudocode calling function

```
function [PESQ_Score,MOS]=computeMOSofG729(original,degraded)
%Compute the MOS value according to the PESQ
%ITU-T recommendation P.862 (02/2001).
%ITU-T recommendation P.862.1 (11/2003).
%ITU-T recommendation P.862 ammendment 2 (11/2005
degradedsignal=int16(decodertocombine(degraded));
[PESQ_Score,MOS]=pesqtocombine('+8000',original,degradedsignal);
```

Appendix B

Ten Folds Experiments

1. First Fold

   In this fold a linear and non-linear (cubic) equation and ANN network were produced. The linear and non-linear equation for the first fold are as follows:

$$\begin{aligned} MOSLinear = {} & (-.033) * PacketLoss + (-.007) * Language + .005 * BurstRatio + \\ & (-.163) * Gender + .100 * Codec + 2.757 \end{aligned} \tag{B.1}$$

$$\begin{aligned} MOSNonLinear = {} & (-1.217E-01) * PacketLoss + .003 * PacketLoss^2 \\ & + (-2.448E-05) * PacketLoss^3 + (-1.119E-01) * Language \\ & + .011 * Langauge^2 + (-3.061E-04) * Language^3 \\ & + (-1.482E-01) * BurstRatio + .046 * BurstRatio^2 \\ & + (-4.040E-03) * BurstRatio^3 + (2.061E+06) * Gender \\ & + (-9.826E+05) * Gender^2 + 126685.763 * Gender^3 \\ & + (-4.085E+05) * Codec + (-4.509E+05) * Codec^2 \\ & + 251591.467 * Codec^3 + (-5.974E+05) \end{aligned} \tag{B.2}$$

The statistical analysis for regression methods and ANN can be shown in following Table B.1, this statistical analysis is the Average, Standard Deviation, Variation, Max, and Min difference of each MOS classifiers values from empirical MOS values:

Table B.1: Fold 1 Statistical Analysis with Regression Methods and ANN

	Linear Regression	Non-Linear Regression	Artificial Neural Network
Average	0.2202	0.1483	0.0942
Standard Deviation	0.1675	0.1326	0.0900
Variation	0.0281	0.0176	0.0081
Max	1.1363	0.8414	0.6829
Min	0.0000	0.0000	0.0000

In order to clarify the accuracy of continuous data, the following boxplot B.1 explains the difference value of each MOS of the derived classifiers against empirical MOS.

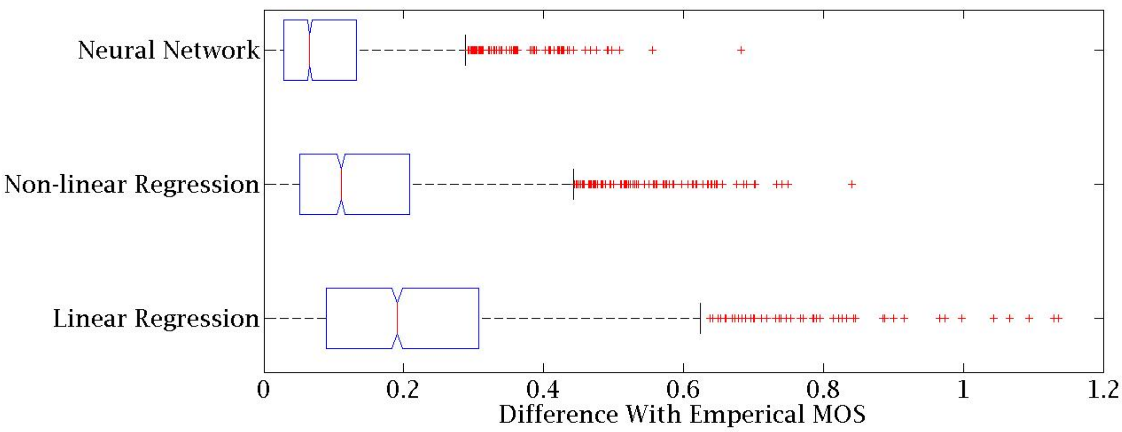

Figure B.1: Boxplot for Fold 1

2. Second Fold

In this fold a linear and non-linear (cubic) equation and ANN network were pro-

duced. The linear and non-linear equation for the first fold are as follows:

$$MOSLinear = (-.033) * PacketLoss + (-.006) * Language + .005 * BurstRatio + (-.162) * Gender + .098 * Codec + 2.753 \tag{B.3}$$

$$\begin{aligned} MOSNonLinear = {} & (-1.199E-01) * PacketLoss + .003 * PacketLoss^2 \\ & + (-2.370E-05) * PacketLoss^3 + (-1.118E-01) * Language \\ & + .011 * Languge^2 + (-3.038E-04) * Language^3 \\ & + (-1.397E-01) * BusrtRatio + .043 * BurstRatio^2 \\ & + (-3.641E-03) * BurstRatio^3 + (-1.226E+06) * Gender \\ & + (-6.306E+05) * Gender^2 + 445376.397 * Gender^3 \\ & + (-2.616E+05) * Codec + (-6.200E+05) * Codec^2 \\ & + 303084.029 * Codec^3 + (1.990E+06) \end{aligned} \tag{B.4}$$

The statistical analysis for regression methods and ANN can be shown in following Table B.2.

Table B.2: Fold 2 Statistical Analysis with Regression Methods and ANN

	Linear Regression	Non-Linear Regression	Artificial Neural Network
Average	0.2378	0.1554	0.1012
Standard Deviation	0.1855	0.1372	0.0947
Variation	0.0344	0.0188	0.0090
Max	1.1691	0.7738	0.5944
Min	0.0000	0.0001	0.0000

In order to clarify the accuracy of continuous data, the following boxplot B.2 explains the difference value of each MOS of the derived classifiers against empirical MOS.

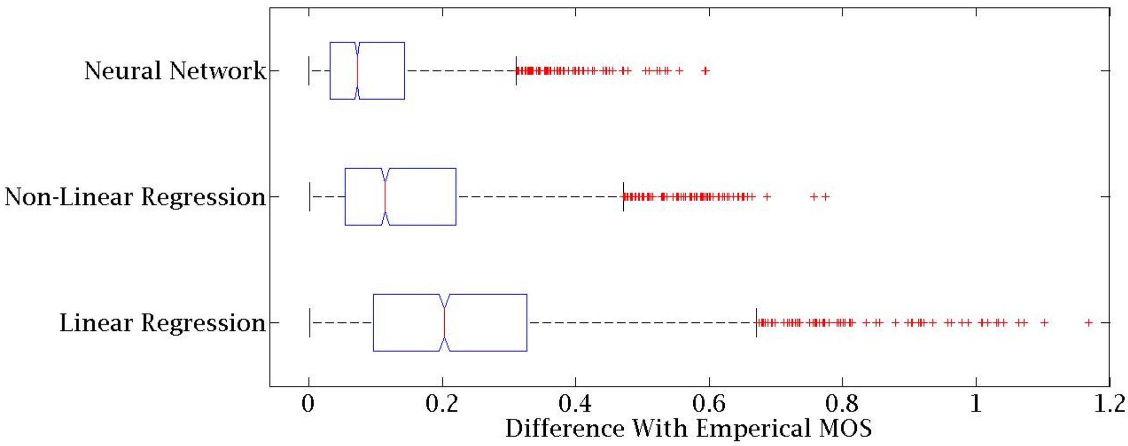

Figure B.2: Boxplot for Fold 2

3. Third Fold

In this fold a linear and non-linear (cubic) equation and ANN network were produced. The linear and non-linear equation for the first fold are as follows:

$$MOSLinear = (-.033)*PacketLoss + (-.007)*Language + .006*BurstRatio + (-.163)*Gender + .100*Codec + 2.756 \quad (B.5)$$

$$\begin{aligned} MOSNonLinear = {} & (-1.213E-01)*PacketLoss + .003*PacketLoss^2 \\ & + (-2.437E-05)*PacketLoss^3 + (-1.100E-01)*Language \\ & + .010*Langugae^2 + (-2.986E-04)*Language^3 \\ & + (-1.320E-01)*BurstRatio + .040*burstRatio^2 \\ & + (-3.427E-03)*BurstRatio^3 + (-1.780E+06)*Gender \\ & + (-5.973E+05)*Gender^2 + 510298.576*Gender^3 \\ & + (1.077E+06)*Codec + (-2.966E+05)*Codec^2 \\ & + (-2.667E+04)*Codec^3 + (1.114E+06) \quad (B.6) \end{aligned}$$

The statistical analysis for regression methods and ANN can be shown in following

Table B.3.

Table B.3: Fold 3 Statistical Analysis with Regression Methods and ANN

	Linear Regression	Non-Linear Regression	Artificial Neural Network
Average	0.2279	0.1519	0.0943
Standard Deviation	0.1764	0.1329	0.0902
Variation	0.0311	0.0177	0.0081
Max	1.1894	0.8023	0.6265
Min	0.0001	0.0000	0.0000

In order to clarify the accuracy of continuous data, the following boxplot B.3 explains the difference value of each MOS of the derived classifiers against empirical MOS.

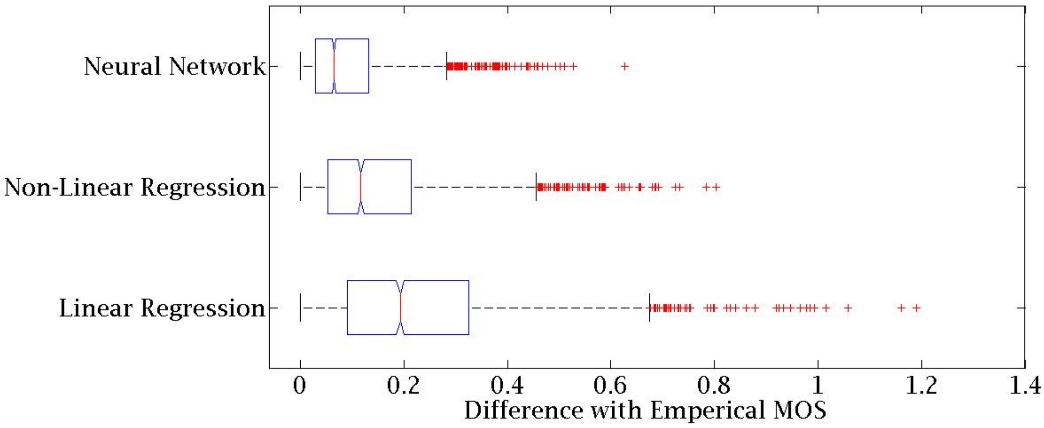

Figure B.3: Boxplot for Fold 3

4. Fourth Fold

In this fold a linear and non-linear (cubic) equation and ANN network were produced. The linear and non-linear equation for the first fold are as follows:

$$MOSLinear = (-.033) * PacketLoss + (-.006) * Language + .006 * BurstRatio + (-.165) * Gender + .100 * Codec + 2.753 \tag{B.7}$$

$$
\begin{aligned}
MOSNonLinear = {} & (-1.214E-01)*PacketLoss + .003*PacketLoss^2 \\
& + (-2.449E-05)*PacketLoss^3 + (-1.115E-01)*Language \\
& + .011*Language^2 + (-3.054E-04)*Language^3 \\
& + (-1.313E-01)*BurstRatio + .040*BurstRatio^2 \\
& + (-3.380E-03)*BurstRatio^3 + (-1.839E+06)*Gender \\
& + 496969.589*Gender^2 + 49731.044*Gender^3 \\
& + (2.118E+06)*Codec + (-1.356E+05)*Codec^2 \\
& + (-2.445E+05)*Codec^3 + (-4.457E+05) \qquad \text{(B.8)}
\end{aligned}
$$

The statistical analysis for regression methods and ANN can be shown in following Table B.4.

Table B.4: Fold 4 Statistical Analysis with Regression Methods and ANN

	Linear Regression	Non-Linear Regression	Artificial Neural Network
Average	0.2235	0.1505	0.0938
Standard Deviation	0.1733	0.1321	0.0947
Variation	0.0300	0.0175	0.0090
Max	1.0783	0.8823	0.5982
Min	0.0001	0.0001	0.0000

In order to clarify the accuracy of continuous data, the following boxplot B.4 explains the difference value of each MOS of the derived classifiers against empirical MOS.

5. Fifth Fold

In this fold a linear and non-linear (cubic) equation and ANN network were pro-

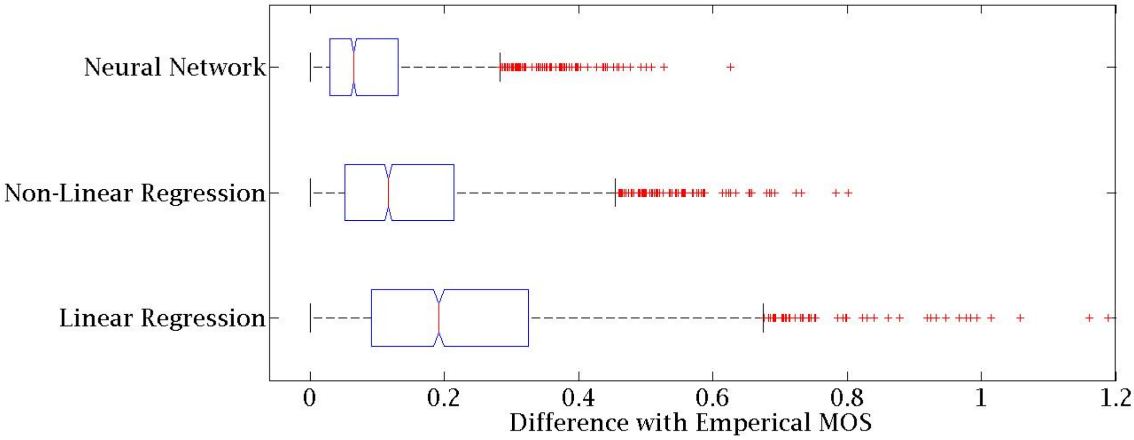

Figure B.4: Boxplot for Fold 4

duced. The linear and non-linear equation for the first fold are as follows:

$$MOSLinear = (-.033) * PacketLoss + (-.006) * Language + .006 * BurstRatio + (-.164) * Gender + .099 * Codec + 2.748 \quad \text{(B.9)}$$

$$\begin{aligned} MOSNonLinear = {} & (-1.220E-01) * PacketLoss + .003 * PacketLoss^2 \\ & + (-2.483E-05) * PacketLoss^3 + (-1.122E-01) * Language \\ & + .011 * Language^2 + (-3.075E-04) * Language^3 \\ & + (-1.407E-01) * BurstRatio + .044 * BurstRatio^2 \\ & + (-3.888E-03) * BurstRatio^3 + (-6.781E+05) * Gender \\ & + (-6.353E+05) * Gender^2 + (369141.619) * Gender^3 \\ & + (2.012E+06) * Codec + (-7.590E+05) * Codec^2 \\ & + 37910.085 * Codec^3 + (-3.464E+05) \quad \text{(B.10)} \end{aligned}$$

The statistical analysis for regression methods and ANN can be shown in following Table B.5.

In order to clarify the accuracy of continuous data, the following boxplot B.5 ex-

Table B.5: Fold 5 Statistical Analysis with Regression Methods and ANN

	Linear Regression	Non-Linear Regression	Artificial Neural Network
Average	0.2257	0.1531	0.1073
Standard Deviation	0.1754	0.1366	0.1016
Variation	0.0308	0.0187	0.0103
Max	1.0932	0.8527	0.7865
Min	0.0000	0.0001	0.0002

plains the difference value of each MOS of the derived classifiers against empirical MOS.

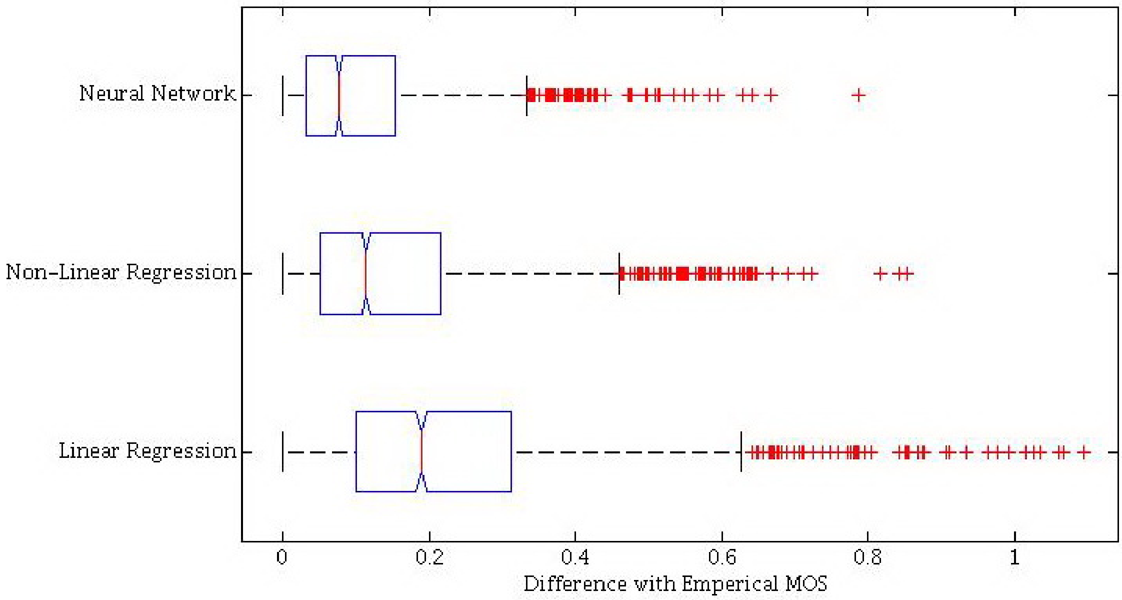

Figure B.5: Boxplot for Fold 5

6. Sixth Fold

In this fold a linear and non-linear (cubic) equation and ANN network were produced. The linear and non-linear equation for the first fold are as follows:

$$MOSLinear = (-.033) * PacketLoss + (-.006) * Language + .006 * BurstRatio + (-.163) * Gender + .100 * Codec + 2.754 \tag{B.11}$$

$$\begin{aligned} MOSNonLinear = (-1.222E-01)*PacketLoss + .003*PacketLoss^2 \\ + (-2.487E-05)*PacketLoss^3 + (-1.115E-01)*Language \\ + .011*Language^2 + (-3.062E-04)*Language^3 \\ + (-1.270E-01)*BurstRatio + .039*BurstRatio^2 \\ + (-3.250E-03)*BurstRatio^3 + (-3.893E+05)*Gender \\ + (-3.919E+05)*Gender^2 + (223552.431)*Gender^3 \\ + (1.900E+06)*Codec + (-1.099E+06)*Codec^2 \\ + 199566.807*Codec^3 + (-4.431E+05) \end{aligned} \tag{B.12}$$

The statistical analysis for regression methods and ANN can be shown in following Table B.6.

Table B.6: Fold 6 Statistical Analysis with Regression Methods and ANN

	Linear Regression	Non-Linear Regression	Artificial Neural Network
Average	0.2235	0.1493	0.1008
Standard Deviation	0.1719	0.1293	0.0918
Variation	0.0295	0.0167	0.0084
Max	1.1523	0.8361	0.6722
Min	0.0001	0.0000	0.0001

In order to clarify the accuracy of continuous data, the following boxplot B.6 explains the difference value of each MOS of the derived classifiers against empirical MOS.

7. Seventh Fold

In this fold a linear and non-linear (cubic) equation and ANN network were pro-

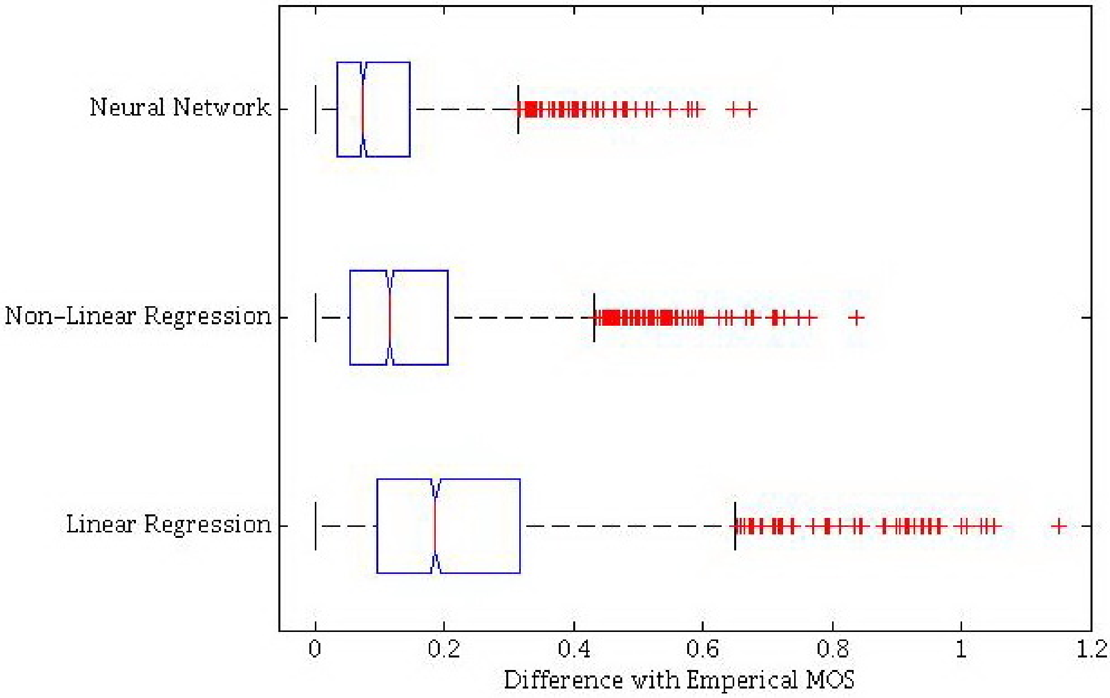

Figure B.6: Boxplot for Fold 6

duced. The linear and non-linear equation for the first fold are as follows:

$$MOSLinear = (-.033)*PacketLoss + (-.006)*Language + .006*BurstRatio + (-.165)*Gender + .102*Codec + 2.752 \tag{B.13}$$

$$
\begin{aligned}
MOSNonLinear = & (-1.222E-01) * PacketLoss + .003 * PacketLoss^2 \\
& + (-2.482E-05) * PacketLoss^3 + (-1.113E-01) * Language \\
& + .011 * Language^2 + (-3.065E-04) * Language^3 \\
& + (-1.355E-01) * BurstRatio + .041 * BurstRatio^2 \\
& + (-3.542E-03) * BurstRatio^3 + (-2.330E+05) * Gender \\
& + (-3.538E+05) * Gender^2 + (184913.346) * Gender^3 \\
& + (2.302E+06) * Codec + (-9.207E+05) * Codec^2 \\
& + 65816.116 * Codec^3 + (-1.045E+06) \qquad \text{(B.14)}
\end{aligned}
$$

The statistical analysis for regression methods and ANN can be shown in following Table B.7.

Table B.7: Fold 7 Statistical Analysis with Regression Methods and ANN

	Linear Regression	Non-Linear Regression	Artificial Neural Network
Average	0.2214	0.1516	0.0936
Standard Deviation	0.1686	0.1316	0.0924
Variation	0.0284	0.0173	0.0085
Max	1.0422	0.8574	0.6668
Min	0.0001	0.0001	0.0001

In order to clarify the accuracy of continuous data, the following boxplot B.7 explains the difference value of each MOS of the derived classifiers against empirical MOS.

8. Eighth Fold

In this fold a linear and non-linear (cubic) equation and ANN network were pro-

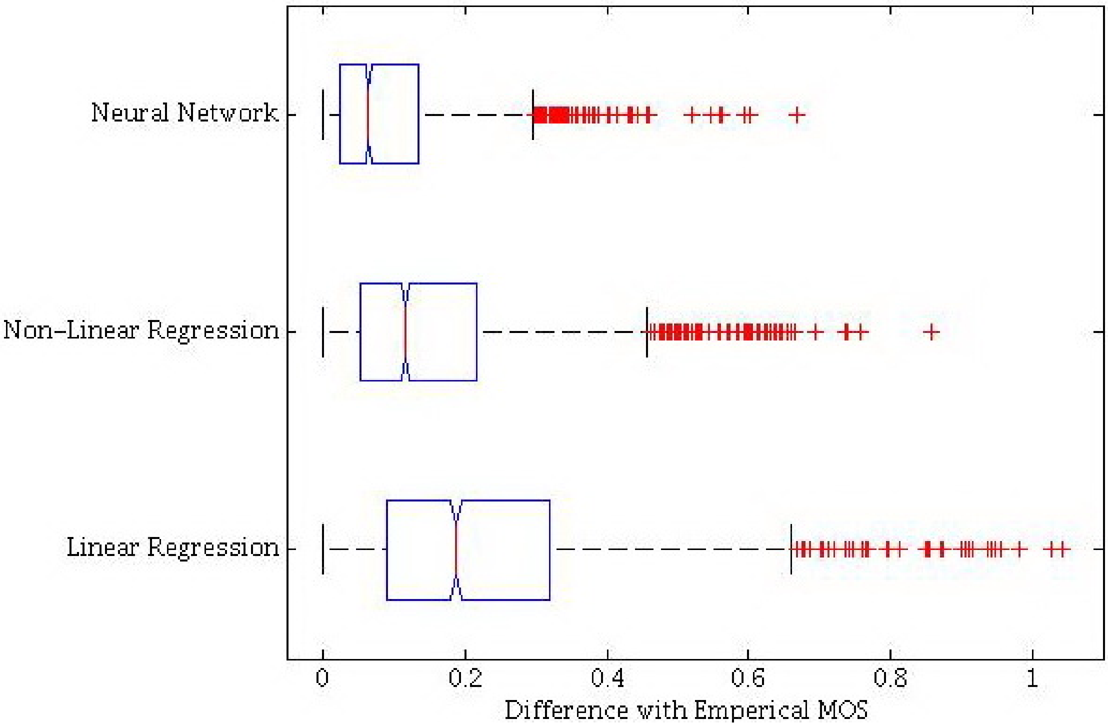

Figure B.7: Boxplot for Fold 7

duced. The linear and non-linear equation for the first fold are as follows:

$$MOSLinear = (-.033) * PacketLoss + (-.006) * Language + .006 * BurstRatio + (-.167) * Gender + .101 * Codec + 2.756 \tag{B.15}$$

$$\begin{aligned} MOSNonLinear = {} & (-1.220E-01)*PacketLoss + .003*PacketLoss^2 \\ & + (-2.470E-05)*PacketLoss^3 + (-1.118E-01)*Language \\ & + .011*Language^2 + (-3.046E-04)*Language^3 \\ & + (-1.380E-01)*BurstRatio + .043*BurstRatio^2 \\ & + (-3.694E-03)*BurstRatio^3 + 924113.479*Gender \\ & + 65597.710*Gender^2 + (-1.601E+05)*Gender^3 \\ & + (2.071E+06)*Codec + (-8.168E+05)*Codec^2 \\ & + 54224.171*Codec^3 + (-2.138E+06) \end{aligned} \tag{B.16}$$

The statistical analysis for regression methods and ANN can be shown in following Table B.8.

Table B.8: Fold 8 Statistical Analysis with Regression Methods and ANN

	Linear Regression	Non-Linear Regression	Artificial Neural Network
Average	0.2284	0.1561	0.0999
Standard Deviation	0.1714	0.1380	0.0977
Variation	0.0294	0.0190	0.0095
Max	1.0587	0.8510	0.6362
Min	0.0000	0.0001	0.0000

In order to clarify the accuracy of continuous data, the following boxplot B.8 explains the difference value of each MOS of the derived classifiers against empirical MOS.

9. Ninth Fold

In this fold a linear and non-linear (cubic) equation and ANN network were pro-

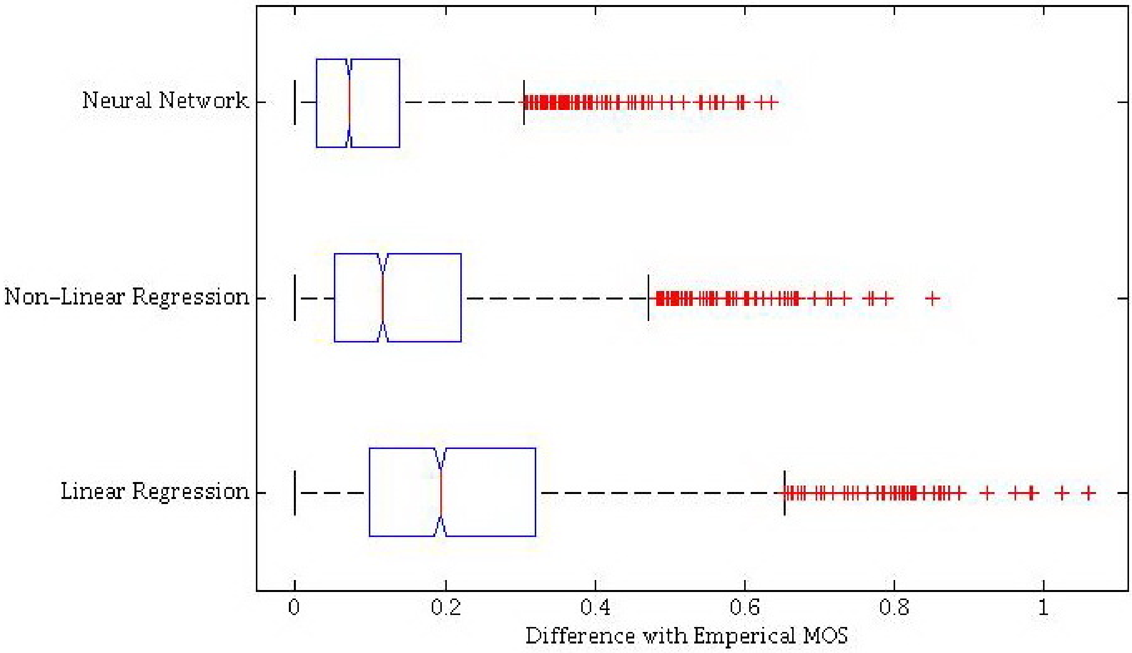

Figure B.8: Boxplot for Fold 8

duced. The linear and non-linear equation for the first fold are as follows:

$$MOSLinear = (-.033)*PacketLoss + (-.007)*Language + .004*BurstRatio + (-.163)*Gender + .099*Codec + 2.755 \tag{B.17}$$

$$\begin{aligned} MOSNonLinear = {} & (-1.221E-01)*PacketLoss + .003*PacketLoss^2 \\ & + (-2.484E-05)*PacketLoss^3 + (-1.117E-01)*Language \\ & + .011*Language^2 + (-3.078E-04)*Language^3 \\ & + (-1.380E-01)*BurstRatio + .042*BurstRatio^2 \\ & + (-3.570E-03)*BurstRatio^3 + (-5.727E+05)*Gender \\ & + 66915.141*Gender^2 + 53131.259*Gender^3 \\ & + (1.549E+06)*Codec + (-1.393E+06)*Codec^2 \\ & + 375829.046*Codec^3 + (-7.907E+04) \end{aligned} \tag{B.18}$$

The statistical analysis for regression methods and ANN can be shown in following Table B.9.

Table B.9: Fold 9 Statistical Analysis with Regression Methods and ANN

	Linear Regression	Non-Linear Regression	Artificial Neural Network
Average	0.2240	0.1509	0.0944
Standard Deviation	0.1755	0.1350	0.0952
Variation	0.0308	0.0182	0.0091
Max	1.1073	0.8486	0.6816
Min	0.0000	0.0003	0.0001

In order to clarify the accuracy of continuous data, the following boxplot B.9 explains the difference value of each MOS of the derived classifiers against empirical MOS.

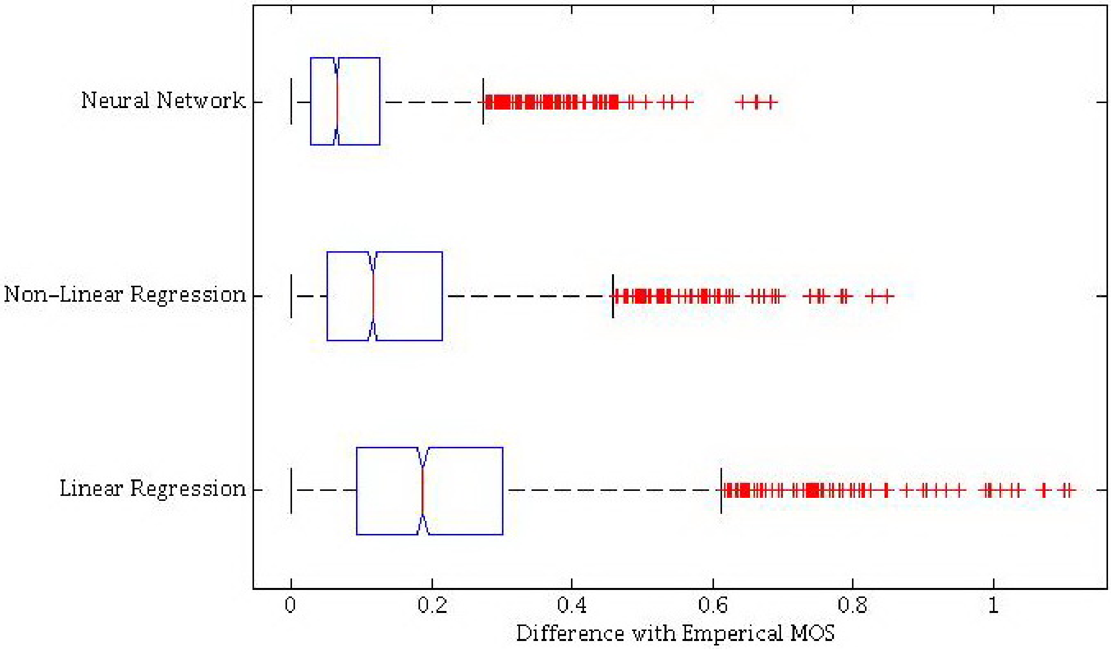

Figure B.9: Boxplot for Fold 9

10. Tenth Fold

In this fold a linear and non-linear (cubic) equation and ANN network were pro-

duced. The linear and non-linear equation for the first fold are as follows:

$$MOSLinear = (-.033)*PacketLoss + (-.007)*Language + .006*BurstRatio + (-.163)*Gender + .099*Codec + 2.757 \quad (B.19)$$

$$\begin{aligned} MOSNonLinear = {} & (-1.217E-01)*PacketLoss + .003*PacketLoss^2 \\ & + (-2.455E-05)*PacketLoss^3 + (-1.128E-01)*Language \\ & + .011*Language^2 + (-3.102E-04)*Language^3 \\ & + (-1.394E-01)*BurstRatio + .043*BurstRatio^2 \\ & + (-3.700E-03)*BurstRatio^3 + (-4.134E+05)*Gender \\ & + (-2.369E+05)*Gender^2 + 160585.343*Gender^3 \\ & + (2.625E+06)*Codec + (-3.426E+05)*Codec^2 \\ & + (-2.282E+05)*Codec^3 + (-1.565E+06) \quad (B.20) \end{aligned}$$

The statistical analysis for regression methods and ANN can be shown in following Table B.10.

Table B.10: Fold 10 Statistical Analysis with Regression Methods and ANN

	Linear Regression	Non-Linear Regression	Artificial Neural Network
Average	0.2237	0.1547	0.1096
Standard Deviation	0.1730	0.1353	0.1009
Variation	0.0299	0.0183	0.0102
Max	1.0682	0.8363	0.6488
Min	0.0002	0.0003	0.0000

In order to clarify the accuracy of continuous data, the following boxplot B.10 explains the difference value of each MOS of the derived classifiers against empirical MOS.

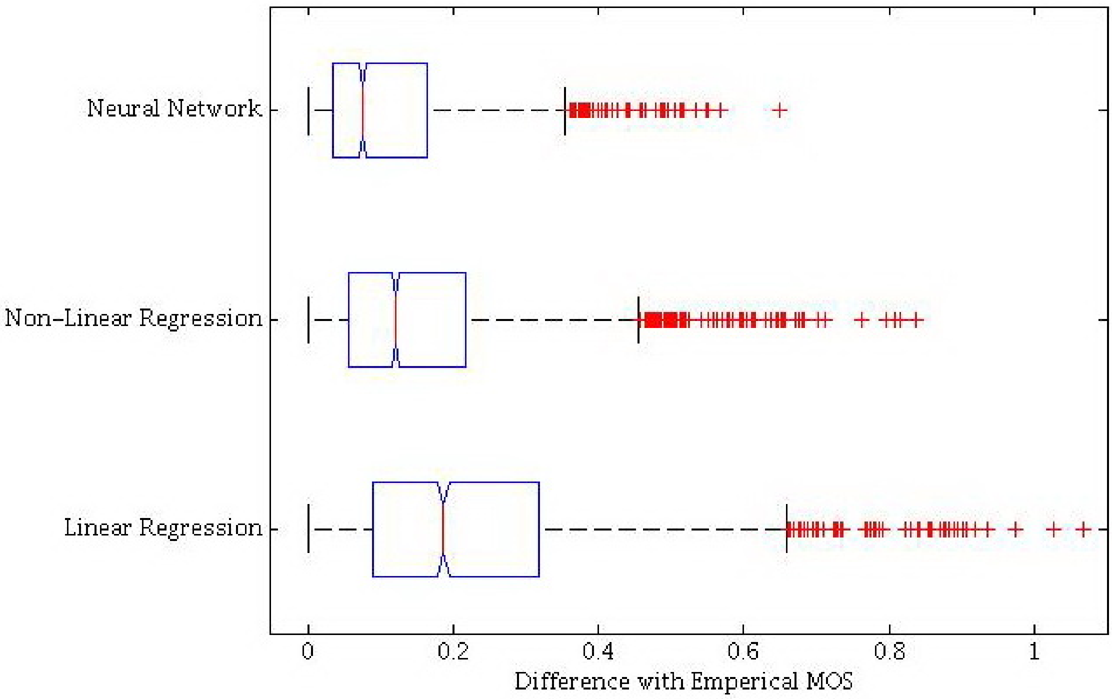

Figure B.10: Boxplot for Fold 10

Appendix C

Ten Folds Ant Miner Rules

In this Appendix, The rules produced By the Ant-Miner variations are listed for reference.

1. First Fold

   In this fold Ant-Miner variation rules were produced. The Ant-Miner, CAnt-Miner and CAnt-Miner rules are as follows:

   (a) Ant Miner

   IF Gender = 2 THEN MOS= Poor2

   IF Packet_Loss = 1 AND Language = 13 THEN MOS= Good4

   IF Packet_Loss = 1 THEN MOS= Good2

   IF Codec = 2 THEN MOS= Poor2

   IF Packet_Loss = 2 THEN MOS= Good1

   IF Packet_Loss = 4 THEN MOS= Fair4

   IF Packet_Loss = 3 THEN MOS= Fair4

   IF Burst_Ratio = 3 THEN MOS= Poor2

   IF Burst_Ratio = 2 THEN MOS= Poor2

   IF Packet_Loss = 5 THEN MOS= Fair4

   IF Packet_Loss = 6 THEN MOS= Fair4

   IF Burst_Ratio = 1 THEN MOS= Poor2

   IF Burst_Ratio = 4 THEN MOS= Poor2

IF Packet_Loss = 9 THEN MOS= Fair4

IF Packet_Loss = 8 THEN MOS= Fair3

IF Packet_Loss = 7 THEN MOS= Fair3

IF Packet_Loss = 11 THEN MOS= Fair2

IF Packet_Loss = 10 THEN MOS= Fair2

IF Packet_Loss = 12 THEN MOS= Fair2

IF Packet_Loss = 15 THEN MOS= Fair1

IF Packet_Loss = 17 THEN MOS= Fair1

IF Packet_Loss = 18 THEN MOS= Fair1

IF Packet_Loss = 14 THEN MOS= Fair1

IF Packet_Loss = 13 THEN MOS= Fair1

IF Packet_Loss = 25 THEN MOS= Poor4

IF Packet_Loss = 19 THEN MOS= Poor4

IF Packet_Loss = 21 THEN MOS= Poor4

IF Packet_Loss = 27 THEN MOS= Poor4

IF Packet_Loss = 22 THEN MOS= Poor4

IF Packet_Loss = 23 THEN MOS= Poor4

IF Packet_Loss = 26 THEN MOS= Poor4

IF Packet_Loss = 20 THEN MOS= Poor4

IF Language = 11 THEN MOS= Poor4

IF Packet_Loss = 16 THEN MOS= Poor4

IF Packet_Loss = 31 THEN MOS= Poor3

IF Packet_Loss = 32 THEN MOS= Poor3

IF Packet_Loss = 29 THEN MOS= Poor3

IF Packet_Loss = 33 THEN MOS= Poor3

IF Packet_Loss = 28 THEN MOS= Poor3

IF Packet_Loss = 30 THEN MOS= Poor3

IF Language = 3 THEN MOS= Poor3

IF Packet_Loss = 35 THEN MOS= Poor3

IF Packet_Loss = 24 THEN MOS= Poor3

IF Language = 1 THEN MOS= Poor3

IF Packet_Loss = 34 THEN MOS= Poor3

IF Language = 18 THEN MOS= Poor3

IF Language = 17 THEN MOS= Poor3

IF Packet_Loss = 44 THEN MOS= Poor2

IF Language = 7 THEN MOS= Poor2

IF Language = 8 THEN MOS= Poor2

IF Language = 15 THEN MOS= Poor2

IF Packet_Loss = 37 THEN MOS= Poor3

IF Packet_Loss = 46 THEN MOS= Poor2

IF Language = 5 THEN MOS= Poor2

IF Language = 6 THEN MOS= Poor2

IF Language = 20 THEN MOS= Poor2

IF Packet_Loss = 36 THEN MOS= Poor3

IF Language = 12 THEN MOS= Poor2

IF Packet_Loss = 38 THEN MOS= Poor3

IF Language = 13 THEN MOS= Poor2

IF Language = 9 THEN MOS= Poor2

IF Language = 4 THEN MOS= Poor2

IF Language = 14 THEN MOS= Poor2

IF Language = 16 THEN MOS= Poor2

IF Language = 10 THEN MOS= Poor2

IF Language = 19 THEN MOS= Poor2

IF Codec = 1 THEN MOS= Poor2

(b) CAnt Miner

IF Packet_Loss $\geq$ 25.5 THEN MOS= Poor2

IF Packet_Loss <8.5 THEN MOS= Fair4

IF Packet_Loss $\geq$ 16.5 AND Burst_Ratio $\geq$ 1.5 THEN MOS= Poor3

IF Packet_Loss $\geq$ 13.5 AND Codec = 1 AND Burst_Ratio <1.5 THEN MOS= Poor2

IF Gender = 2 AND Codec = 1 THEN MOS= Poor3

IF Codec = 2 AND Packet_Loss $\geq$ 12.5 THEN MOS= Poor4

IF Burst_Ratio $\geq$ 1.5 AND Packet_Loss $\geq$ 10.5 THEN MOS= Fair1

IF Codec = 1 AND Burst_Ratio <1.5 THEN MOS= Poor4

IF Burst_Ratio $\geq$ 1.5 AND Codec = 2 THEN MOS= Fair1

IF Language = 7 THEN MOS= Poor4

IF Packet_Loss $\geq$ 9.5 AND Gender = 1 AND Burst_Ratio <1.5 THEN MOS= Fair4

IF Packet_Loss <9.5 THEN MOS= Fair3

IF Codec = 2 THEN MOS= Fair2

IF Burst_Ratio $\geq$ 2.5 THEN MOS= Fair2

IF Gender = 1 THEN MOS= Fair1

(c) CAnt Miner type2

IF Burst_Ratio $\geq$ 1.5 AND Packet_Loss $\geq$ 21.5 THEN MOS= Poor2

IF Packet_Loss $\geq$ 12.5 AND Codec = 1 THEN MOS= Poor1

IF Codec = 2 AND Packet_Loss $\geq$ 11.5 THEN MOS= Poor4

IF Gender = 2 AND Packet_Loss $\geq$ 5.5 AND Codec = 1 AND Burst_Ratio <2.5 THEN MOS= Poor3

IF Packet_Loss $\geq$ 6.5 THEN MOS= Fair2

IF Codec = 1 AND Gender = 2 THEN MOS= Fair1

IF Packet_Loss <2.5 THEN MOS= Good2

IF Packet_Loss <4.5 AND Burst_Ratio $\geq$ 1.5 THEN MOS= Fair4

IF Burst_Ratio $\geq$ 1.5 AND Gender = 2 THEN MOS= Fair2

IF Burst_Ratio $\geq$ 1.5 THEN MOS= Fair3

IF Codec = 2 AND Gender = 1 THEN MOS= Good1

IF Packet_Loss <4.5 THEN MOS= Fair4

IF Gender = 1 THEN MOS= Fair2

IF Packet_Loss <5.5 THEN MOS= Fair4

IF Codec = 2 THEN MOS= Fair3

2. Second Fold

   In this fold Ant-Miner variation rules were produced. The Ant-Miner, CAnt-Miner and CAnt-Miner rules are as follows:

   (a) Ant Miner

   IF Gender = 2 THEN MOS= Poor2

   IF Packet_Loss = 1 AND Language = 13 THEN MOS= Good4

   IF Packet_Loss = 1 THEN MOS= Good2

   IF Codec = 2 THEN MOS= Poor2

   IF Packet_Loss = 2 THEN MOS= Good1

   IF Packet_Loss = 4 THEN MOS= Fair4

   IF Language = 18 AND Packet_Loss = 3 THEN MOS= Good3

   IF Burst_Ratio = 5 AND Packet_Loss = 3 THEN MOS= Good2

   IF Packet_Loss = 5 THEN MOS= Fair4

   IF Packet_Loss = 3 THEN MOS= Fair4

   IF Packet_Loss = 6 THEN MOS= Fair3

   IF Language = 11 AND Packet_Loss = 7 THEN MOS= Good1

   IF Packet_Loss = 7 THEN MOS= Fair3

   IF Burst_Ratio = 2 THEN MOS= Poor2

   IF Burst_Ratio = 3 THEN MOS= Poor2

IF Burst_Ratio = 1 THEN MOS= Poor2

IF Burst_Ratio = 4 THEN MOS= Poor2

IF Packet_Loss = 8 THEN MOS= Fair3

IF Packet_Loss = 9 THEN MOS= Fair3

IF Packet_Loss = 11 THEN MOS= Fair2

IF Packet_Loss = 14 THEN MOS= Fair2

IF Packet_Loss = 10 THEN MOS= Fair2

IF Packet_Loss = 16 THEN MOS= Fair2

IF Packet_Loss = 12 THEN MOS= Fair2

IF Packet_Loss = 15 THEN MOS= Fair1

IF Packet_Loss = 13 THEN MOS= Fair1

IF Packet_Loss = 18 THEN MOS= Fair1

IF Packet_Loss = 17 THEN MOS= Fair1

IF Packet_Loss = 21 THEN MOS= Poor4

IF Language = 4 THEN MOS= Poor4

IF Packet_Loss = 19 THEN MOS= Poor4

IF Packet_Loss = 22 THEN MOS= Poor4

IF Packet_Loss = 23 THEN MOS= Poor4

IF Language = 11 THEN MOS= Poor4

IF Packet_Loss = 20 THEN MOS= Poor4

IF Packet_Loss = 26 THEN MOS= Poor4

IF Language = 1 THEN MOS= Poor3

IF Packet_Loss = 25 THEN MOS= Poor4

IF Packet_Loss = 31 THEN MOS= Poor3

IF Packet_Loss = 29 THEN MOS= Poor3

IF Packet_Loss = 24 THEN MOS= Poor3

IF Packet_Loss = 33 THEN MOS= Poor3

IF Packet_Loss = 28 THEN MOS= Poor3

IF Language = 3 THEN MOS= Poor3

IF Packet_Loss = 30 THEN MOS= Poor3

IF Packet_Loss = 32 THEN MOS= Poor3

IF Packet_Loss = 35 THEN MOS= Poor3

IF Packet_Loss = 27 THEN MOS= Poor3

IF Packet_Loss = 34 THEN MOS= Poor3

IF Language = 18 THEN MOS= Poor3

IF Language = 15 THEN MOS= Poor2

IF Language = 16 THEN MOS= Poor2

IF Language = 5 THEN MOS= Poor2

IF Language = 8 THEN MOS= Poor2

IF Language = 6 THEN MOS= Poor2

IF Packet_Loss = 37 THEN MOS= Poor3

IF Language = 20 THEN MOS= Poor2

IF Language = 19 THEN MOS= Poor2

IF Packet_Loss = 38 THEN MOS= Poor3

IF Language = 9 THEN MOS= Poor2

IF Packet_Loss = 36 THEN MOS= Poor3

IF Language = 14 THEN MOS= Poor2

IF Language = 7 THEN MOS= Poor2

IF Language = 10 THEN MOS= Poor2

IF Language = 13 THEN MOS= Poor2

IF Language = 12 THEN MOS= Poor2

IF Language = 17 THEN MOS= Poor2

IF Language = 2 THEN MOS= Poor2

(b) CAnt Miner

IF Packet_Loss $\geq$ 25.5 THEN MOS= Poor2

IF Packet_Loss <8.5 THEN MOS= Fair4

IF Packet_Loss ≥ 15.5 AND Codec = 1 AND Gender = 2 THEN MOS= Poor2

IF Burst_Ratio ≥ 1.5 AND Packet_Loss ≥ 16.5 THEN MOS= Poor3

IF Codec = 1 AND Packet_Loss ≥ 16.5 THEN MOS= Poor2

IF Codec = 1 AND Gender = 2 THEN MOS= Poor3

IF Codec = 2 AND Packet_Loss ≥ 12.5 THEN MOS= Poor4

IF Burst_Ratio ≥ 1.5 AND Codec = 2 THEN MOS= Fair1

IF Codec = 1 AND Burst_Ratio <1.5 AND Packet_Loss ≥ 11.5 THEN MOS= Poor3

IF Packet_Loss ≥ 12.5 AND Burst_Ratio <3.5 THEN MOS= Poor4

IF Packet_Loss ≥ 10.5 AND Codec = 1 THEN MOS= Fair1

IF Language = 7 AND Burst_Ratio <3.5 THEN MOS= Poor4

IF Packet_Loss <9.5 THEN MOS= Fair3

IF Codec = 2 THEN MOS= Fair2

IF Language = 3 THEN MOS= Fair4

IF Burst_Ratio <1.5 THEN MOS= Fair1

IF Burst_Ratio ≥ 2.5 THEN MOS= Fair2

IF Codec = 1 THEN MOS= Fair1

(c) CAnt Miner type2

IF Packet_Loss ≥ 20.5 THEN MOS= Poor2

IF Packet_Loss ≥ 8.5 THEN MOS= Poor4

IF Gender = 2 AND Codec = 1 THEN MOS= Fair1

IF Packet_Loss ≥ 4.5 THEN MOS= Fair3

IF Gender = 1 AND Packet_Loss <2.5 THEN MOS= Good2

IF Gender = 1 AND Burst_Ratio ≥ 1.5 THEN MOS= Fair4

IF Burst_Ratio ≥ 1.5 AND Packet_Loss ≥ 2.5 THEN MOS= Fair3

IF Codec = 1 THEN MOS= Fair3

IF Burst_Ratio $\geq$ 1.5 THEN MOS= Fair4

IF Language = 7 THEN MOS= Fair2

IF Gender = 2 THEN MOS= Fair4

IF Codec = 2 AND Packet_Loss <3.5 THEN MOS= Good1

IF Codec = 2 THEN MOS= Good1

3. Third Fold

   In this fold Ant-Miner variation rules were produced. The Ant-Miner, CAnt-Miner and CAnt-Miner rules are as follows:

   (a) Ant Miner

   IF Gender = 2 THEN MOS= Poor2

   IF Packet_Loss = 1 THEN MOS= Good2

   IF Codec = 2 THEN MOS= Poor2

   IF Packet_Loss = 2 THEN MOS= Good1

   IF Packet_Loss = 4 THEN MOS= Fair4

   IF Packet_Loss = 5 THEN MOS= Fair4

   IF Packet_Loss = 3 THEN MOS= Fair4

   IF Packet_Loss = 6 THEN MOS= Fair3

   IF Packet_Loss = 7 AND Language = 11 THEN MOS= Good1

   IF Packet_Loss = 7 THEN MOS= Fair3

   IF Burst_Ratio = 2 THEN MOS= Poor2

   IF Burst_Ratio = 3 THEN MOS= Poor2

   IF Burst_Ratio = 1 THEN MOS= Poor2

   IF Burst_Ratio = 4 THEN MOS= Poor2

   IF Packet_Loss = 9 THEN MOS= Fair4

   IF Packet_Loss = 8 THEN MOS= Fair3

   IF Packet_Loss = 11 THEN MOS= Fair2

IF Packet_Loss = 10 THEN MOS= Fair2

IF Packet_Loss = 12 THEN MOS= Fair2

IF Packet_Loss = 15 THEN MOS= Fair1

IF Packet_Loss = 18 THEN MOS= Fair1

IF Packet_Loss = 13 THEN MOS= Fair1

IF Packet_Loss = 14 THEN MOS= Fair1

IF Packet_Loss = 17 THEN MOS= Fair1

IF Packet_Loss = 19 THEN MOS= Poor4

IF Packet_Loss = 25 THEN MOS= Poor4

IF Language = 11 THEN MOS= Poor4

IF Packet_Loss = 22 THEN MOS= Poor4

IF Packet_Loss = 20 THEN MOS= Poor4

IF Packet_Loss = 23 THEN MOS= Poor4

IF Packet_Loss = 26 THEN MOS= Poor4

IF Packet_Loss = 16 THEN MOS= Poor4

IF Packet_Loss = 21 THEN MOS= Poor4

IF Packet_Loss = 29 THEN MOS= Poor3

IF Language = 1 THEN MOS= Poor3

IF Packet_Loss = 31 THEN MOS= Poor3

IF Packet_Loss = 30 THEN MOS= Poor3

IF Packet_Loss = 32 THEN MOS= Poor3

IF Packet_Loss = 33 THEN MOS= Poor3

IF Packet_Loss = 28 THEN MOS= Poor3

IF Language = 3 THEN MOS= Poor3

IF Packet_Loss = 24 THEN MOS= Poor3

IF Packet_Loss = 34 THEN MOS= Poor3

IF Packet_Loss = 35 THEN MOS= Poor3

IF Packet_Loss = 27 THEN MOS= Poor3

IF Packet_Loss = 37 THEN MOS= Poor3

IF Language = 18 THEN MOS= Poor3

IF Packet_Loss = 46 THEN MOS= Poor2

IF Packet_Loss = 41 THEN MOS= Poor2

IF Language = 17 THEN MOS= Poor3

IF Packet_Loss = 44 THEN MOS= Poor2

IF Packet_Loss = 42 THEN MOS= Poor2

IF Packet_Loss = 47 THEN MOS= Poor2

IF Packet_Loss = 43 THEN MOS= Poor2

IF Packet_Loss = 48 THEN MOS= Poor2

IF Packet_Loss = 40 THEN MOS= Poor2

IF Language = 19 THEN MOS= Poor1

IF Language = 13 THEN MOS= Poor3

IF Packet_Loss = 50 THEN MOS= Poor2

IF Packet_Loss = 39 THEN MOS= Poor2

IF Packet_Loss = 45 THEN MOS= Poor2

IF Packet_Loss = 49 THEN MOS= Poor2

IF Packet_Loss = 38 THEN MOS= Poor2

IF Packet_Loss = 36 THEN MOS= Poor2

(b) CAnt Miner

IF Packet_Loss $\geq$ 25.5 THEN MOS= Poor2

IF Packet_Loss <7.5 THEN MOS= Fair4

IF Packet_Loss $\geq$ 16.5 AND Burst_Ratio $\geq$ 1.5 THEN MOS= Poor3

IF Packet_Loss $\geq$ 12.5 AND Codec = 1 AND Burst_Ratio <1.5 THEN MOS= Poor2

IF Gender = 2 AND Packet_Loss $\geq$ 9.5 AND Codec = 1 THEN MOS= Poor3

IF Packet_Loss $\geq$ 12.5 THEN MOS= Poor4

IF Burst_Ratio $\geq$ 1.5 AND Packet_Loss $\geq$ 9.5 THEN MOS= Fair1

IF Codec = 1 AND Gender = 2 THEN MOS= Poor4

IF Burst_Ratio $\geq$ 1.5 AND Gender = 2 THEN MOS= Fair1

IF Codec = 1 AND Packet_Loss $\geq$ 9.5 THEN MOS= Poor4

IF Burst_Ratio $\geq$ 1.5 THEN MOS= Fair2

IF Packet_Loss $\geq$ 9.5 THEN MOS= Fair2

IF Codec = 1 THEN MOS= Fair2

IF Packet_Loss $\geq$ 8.5 THEN MOS= Fair3

IF Gender = 1 THEN MOS= Fair3

IF Gender = 2 THEN MOS= Fair3

(c) CAnt Miner type2

IF Packet_Loss $\geq$ 21.5 THEN MOS= Poor2

IF Gender = 2 AND Packet_Loss $\geq$ 8.5 THEN MOS= Poor3

IF Packet_Loss $\geq$ 11.5 THEN MOS= Poor4

IF Gender = 2 AND Packet_Loss $\geq$ 4.5 AND Codec = 1 THEN MOS= Poor4

IF Packet_Loss $\geq$ 5.5 THEN MOS= Fair2

IF Language = 7 AND Gender = 2 AND Codec = 1 THEN MOS= Poor4

IF Language = 13 AND Codec = 2 AND Packet_Loss $<$1.5 AND Gender = 1 THEN MOS= Good4

IF Gender = 1 AND Packet_Loss $<$2.5 THEN MOS= Good2

IF Gender = 2 AND Codec = 1 THEN MOS= Fair1

IF Language = 7 AND Codec = 2 AND Burst_Ratio $\geq$ 1.5 AND Packet_Loss $\geq$ 4.5 THEN MOS= Fair1

IF Gender = 1 AND Burst_Ratio $\geq$ 1.5 THEN MOS= Fair4

IF Packet_Loss $<$2.5 THEN MOS= Good1

IF Codec = 2 AND Burst_Ratio $\geq$ 1.5 THEN MOS= Fair3

IF Codec = 1 THEN MOS= Fair3

IF Gender = 2 THEN MOS= Fair4

IF Packet_Loss $\geq$ 4.5 THEN MOS= Good2

IF Packet_Loss <3.5 THEN MOS= Good1

IF Gender = 1 THEN MOS= Good1

4. Fourth Fold

   In this fold Ant-Miner variation rules were produced. The Ant-Miner, CAnt-Miner and CAnt-Miner rules are as follows:

   (a) Ant Miner

   IF Gender = 2 THEN MOS= Poor2

   IF Packet_Loss = 1 THEN MOS= Good2

   IF Codec = 1 THEN MOS= Poor2

   IF Packet_Loss = 3 THEN MOS= Good2

   IF Packet_Loss = 2 THEN MOS= Good1

   IF Packet_Loss = 5 THEN MOS= Good1

   IF Packet_Loss = 4 THEN MOS= Fair4

   IF Burst_Ratio = 1 THEN MOS= Poor3

   IF Packet_Loss = 6 THEN MOS= Fair3

   IF Burst_Ratio = 2 THEN MOS= Poor2

   IF Burst_Ratio = 3 THEN MOS= Poor2

   IF Burst_Ratio = 5 THEN MOS= Poor2

   IF Packet_Loss = 7 THEN MOS= Fair2

   IF Packet_Loss = 8 THEN MOS= Fair2

   IF Packet_Loss = 10 THEN MOS= Fair1

   IF Packet_Loss = 14 THEN MOS= Fair1

   IF Packet_Loss = 11 THEN MOS= Fair1

   IF Packet_Loss = 12 THEN MOS= Fair1

IF Packet_Loss = 9 THEN MOS= Fair1

IF Packet_Loss = 16 THEN MOS= Poor4

IF Packet_Loss = 18 THEN MOS= Poor4

IF Packet_Loss = 13 THEN MOS= Poor4

IF Packet_Loss = 15 THEN MOS= Poor4

IF Packet_Loss = 17 THEN MOS= Poor4

IF Language = 2 THEN MOS= Poor1

IF Language = 7 THEN MOS= Poor1

IF Language = 10 THEN MOS= Poor1

IF Packet_Loss = 20 THEN MOS= Poor4

IF Packet_Loss = 19 THEN MOS= Poor4

IF Language = 9 THEN MOS= Poor1

IF Language = 12 THEN MOS= Poor1

IF Language = 19 THEN MOS= Poor1

IF Packet_Loss = 50 THEN MOS= Poor1

IF Packet_Loss = 48 THEN MOS= Poor1

IF Packet_Loss = 47 THEN MOS= Poor1

IF Packet_Loss = 49 THEN MOS= Poor1

IF Packet_Loss = 46 THEN MOS= Poor1

IF Packet_Loss = 25 THEN MOS= Poor3

IF Packet_Loss = 26 THEN MOS= Poor3

IF Packet_Loss = 28 THEN MOS= Poor3

IF Packet_Loss = 24 THEN MOS= Poor3

IF Packet_Loss = 21 THEN MOS= Poor3

IF Packet_Loss = 27 THEN MOS= Poor3

IF Packet_Loss = 22 THEN MOS= Poor3

IF Language = 3 THEN MOS= Poor3

IF Packet_Loss = 23 THEN MOS= Poor3

IF Language = 1 THEN MOS= Poor3

IF Language = 4 THEN MOS= Poor2

IF Packet_Loss = 45 THEN MOS= Poor1

IF Packet_Loss = 31 THEN MOS= Poor3

IF Language = 20 THEN MOS= Poor2

IF Packet_Loss = 29 THEN MOS= Poor3

IF Language = 5 THEN MOS= Poor2

IF Language = 8 THEN MOS= Poor2

IF Language = 17 THEN MOS= Poor2

IF Language = 18 THEN MOS= Poor2

IF Packet_Loss = 30 THEN MOS= Poor3

IF Packet_Loss = 43 THEN MOS= Poor1

IF Language = 14 THEN MOS= Poor2

IF Language = 15 THEN MOS= Poor2

IF Language = 6 THEN MOS= Poor2

IF Language = 13 THEN MOS= Poor2

IF Language = 11 THEN MOS= Poor2

IF Language = 16 THEN MOS= Poor2

(b) CAnt Miner

IF Packet_Loss $\geq$ 23.5 THEN MOS= Poor2

IF Packet_Loss <7.5 THEN MOS= Fair4

IF Packet_Loss $\geq$ 14.5 AND Codec = 1 AND Burst_Ratio <2.5 THEN MOS= Poor2

IF Packet_Loss $\geq$ 15.5 AND Burst_Ratio $\geq$ 1.5 THEN MOS= Poor3

IF Gender = 2 AND Codec = 1 AND Burst_Ratio <4.5 THEN MOS= Poor3

IF Packet_Loss $\geq$ 11.5 AND Codec = 2 THEN MOS= Poor4

IF Burst_Ratio $\geq$ 1.5 AND Gender = 1 THEN MOS= Fair2

IF Codec = 1 THEN MOS= Poor4

IF Burst_Ratio <1.5 THEN MOS= Fair3

IF Packet_Loss <9.5 THEN MOS= Fair2

IF Language = 1 THEN MOS= Fair3

IF Language = 4 THEN MOS= Fair3

IF Language = 14 THEN MOS= Fair2

IF Language = 2 THEN MOS= Fair2

IF Language = 11 THEN MOS= Fair2

IF Burst_Ratio $\geq$ 2.5 AND Packet_Loss <10.5 THEN MOS= Fair1

IF Burst_Ratio <2.5 THEN MOS= Fair1

IF Burst_Ratio $\geq$ 3.5 THEN MOS= Fair1

IF Gender = 2 THEN MOS= Fair1

(c) CAnt Miner type2

IF Packet_Loss $\geq$ 20.5 THEN MOS= Poor2

IF Packet_Loss $\geq$ 8.5 THEN MOS= Poor4

IF Gender = 2 AND Burst_Ratio $\geq$ 1.5 AND Codec = 1 AND Packet_Loss $\geq$ 4.5 THEN MOS= Poor4

IF Language = 13 AND Packet_Loss <1.5 AND Gender = 1 AND Codec = 2 THEN MOS= Good4

IF Gender = 2 AND Codec = 1 AND Packet_Loss $\geq$ 4.5 THEN MOS= Poor4

IF Codec = 1 AND Gender = 2 THEN MOS= Fair1

IF Packet_Loss $\geq$ 4.5 AND Codec = 2 THEN MOS= Fair2

IF Codec = 2 AND Language = 11 AND Packet_Loss <1.5 AND Gender = 1 THEN MOS= Good4

IF Gender = 1 AND Burst_Ratio $\geq$ 1.5 AND Packet_Loss <3.5 THEN MOS= Good2

IF Packet_Loss $\geq$ 5.5 AND Burst_Ratio <1.5 THEN MOS= Fair1

IF Language = 7 AND Packet_Loss $\geq$ 6.5 THEN MOS= Fair1

IF Gender = 1 AND Packet_Loss $\geq$ 4.5 THEN MOS= Fair3

IF Burst_Ratio $\geq$ 1.5 AND Gender = 2 AND Packet_Loss $\geq$ 2.5 THEN MOS= Fair3

IF Burst_Ratio $\geq$ 1.5 THEN MOS= Fair4

IF Codec = 1 AND Packet_Loss $\geq$ 2.5 THEN MOS= Fair3

IF Gender = 2 THEN MOS= Fair4

IF Packet_Loss $\geq$ 2.5 THEN MOS= Good1

IF Codec = 1 THEN MOS= Good1

IF Packet_Loss <1.5 THEN MOS= Good2

IF Codec = 2 THEN MOS= Good3

5. Fifth Fold

In this fold Ant-Miner variation rules were produced. The Ant-Miner, CAnt-Miner and CAnt-Miner rules are as follows:

(a) Ant Miner

IF Gender = 2 THEN MOS= Poor2

IF Packet_Loss = 1 AND Language = 13 THEN MOS= Good4

IF Packet_Loss = 1 THEN MOS= Good2

IF Codec = 2 THEN MOS= Poor2

IF Packet_Loss = 2 THEN MOS= Good1

IF Packet_Loss = 4 THEN MOS= Fair4

IF Packet_Loss = 3 THEN MOS= Fair4

IF Burst_Ratio = 2 THEN MOS= Poor2

IF Packet_Loss = 5 THEN MOS= Fair4

IF Burst_Ratio = 3 THEN MOS= Poor2

IF Packet_Loss = 6 THEN MOS= Fair4

IF Burst_Ratio = 4 THEN MOS= Poor2

IF Packet_Loss = 9 AND Burst_Ratio = 5 THEN MOS= Fair4

IF Packet_Loss = 7 THEN MOS= Fair3

IF Burst_Ratio = 1 THEN MOS= Poor2

IF Packet_Loss = 8 THEN MOS= Fair3

IF Packet_Loss = 11 THEN MOS= Fair2

IF Packet_Loss = 14 THEN MOS= Fair2

IF Packet_Loss = 10 THEN MOS= Fair2

IF Packet_Loss = 16 THEN MOS= Fair2

IF Packet_Loss = 15 THEN MOS= Fair1

IF Packet_Loss = 18 THEN MOS= Fair1

IF Packet_Loss = 13 THEN MOS= Fair1

IF Packet_Loss = 17 THEN MOS= Fair1

IF Packet_Loss = 12 THEN MOS= Fair1

IF Language = 4 THEN MOS= Poor4

IF Packet_Loss = 19 THEN MOS= Poor4

IF Packet_Loss = 21 THEN MOS= Poor4

IF Packet_Loss = 23 THEN MOS= Poor4

IF Packet_Loss = 25 THEN MOS= Poor4

IF Packet_Loss = 20 THEN MOS= Poor4

IF Packet_Loss = 26 THEN MOS= Poor4

IF Packet_Loss = 22 THEN MOS= Poor4

IF Language = 6 THEN MOS= Poor2

IF Packet_Loss = 27 THEN MOS= Poor4

IF Language = 7 THEN MOS= Poor2

IF Language = 3 THEN MOS= Poor3

IF Language = 1 THEN MOS= Poor3

IF Language = 18 THEN MOS= Poor3

IF Packet_Loss = 30 THEN MOS= Poor3

IF Packet_Loss = 31 THEN MOS= Poor3

IF Packet_Loss = 33 THEN MOS= Poor3

IF Packet_Loss = 32 THEN MOS= Poor3

IF Packet_Loss = 34 THEN MOS= Poor3

IF Packet_Loss = 29 THEN MOS= Poor3

IF Packet_Loss = 28 THEN MOS= Poor3

IF Packet_Loss = 35 THEN MOS= Poor3

IF Packet_Loss = 24 THEN MOS= Poor3

IF Language = 11 THEN MOS= Poor3

IF Language = 16 THEN MOS= Poor2

IF Packet_Loss = 37 THEN MOS= Poor3

IF Language = 15 THEN MOS= Poor2

IF Packet_Loss = 38 THEN MOS= Poor3

IF Language = 20 THEN MOS= Poor2

IF Language = 8 THEN MOS= Poor2

IF Language = 13 THEN MOS= Poor2

IF Language = 9 THEN MOS= Poor2

IF Language = 12 THEN MOS= Poor2

IF Language = 14 THEN MOS= Poor2

IF Language = 5 THEN MOS= Poor2

IF Language = 2 THEN MOS= Poor2

IF Language = 10 THEN MOS= Poor2

IF Language = 19 THEN MOS= Poor2

IF Gender = 1 THEN MOS= Poor2

(b) CAnt Miner

IF Packet_Loss $\geq$ 26.5 THEN MOS= Poor2

IF Burst_Ratio $\geq$ 1.5 AND Packet_Loss $\geq$ 14.5 THEN MOS= Poor3

IF Codec = 1 AND Packet_Loss $\geq$ 9.5 AND Burst_Ratio <1.5 THEN MOS= Poor2

IF Packet_Loss <5.5 THEN MOS= Good1

IF Packet_Loss $\geq$ 11.5 THEN MOS= Poor4

IF Gender = 2 AND Codec = 1 THEN MOS= Poor4

IF Packet_Loss $\geq$ 8.5 AND Burst_Ratio $\geq$ 1.5 THEN MOS= Fair1

IF Burst_Ratio $\geq$ 1.5 AND Packet_Loss $\geq$ 6.5 THEN MOS= Fair2

IF Gender = 2 THEN MOS= Fair2

IF Burst_Ratio $\geq$ 1.5 THEN MOS= Fair3

IF Codec = 1 THEN MOS= Fair2

IF Packet_Loss <8.5 THEN MOS= Fair4

IF Packet_Loss <9.5 THEN MOS= Fair3

IF Packet_Loss $\geq$ 10.5 THEN MOS= Fair2

IF Gender = 1 THEN MOS= Fair2

(c) CAnt Miner type2

IF Packet_Loss $\geq$ 21.5 THEN MOS= Poor2

IF Gender = 2 AND Packet_Loss $\geq$ 9.5 THEN MOS= Poor3

IF Burst_Ratio $\geq$ 1.5 AND Packet_Loss $\geq$ 11.5 THEN MOS= Poor4

IF Gender = 1 AND Packet_Loss <6.5 THEN MOS= Good1

IF Burst_Ratio $\geq$ 1.5 AND Packet_Loss $\geq$ 4.5 THEN MOS= Fair2

IF Codec = 1 AND Packet_Loss $\geq$ 4.5 THEN MOS= Poor3

IF Codec = 1 THEN MOS= Fair1

IF Gender = 1 THEN MOS= Fair2

IF Gender = 2 AND Packet_Loss <2.5 THEN MOS= Good1

IF Burst_Ratio $\geq$ 1.5 THEN MOS= Fair3

IF Packet_Loss $\geq$ 4.5 AND Language = 7 THEN MOS= Fair1

IF Language = 12 THEN MOS= Fair1

IF Packet_Loss $\geq$ 5.5 THEN MOS= Fair3

IF Packet_Loss $\geq$ 4.5 THEN MOS= Fair4

IF Codec = 2 AND Packet_Loss $\geq$ 3.5 THEN MOS= Fair4

IF Codec = 2 THEN MOS= Fair4

6. Sixth Fold

   In this fold Ant-Miner variation rules were produced. The Ant-Miner, CAnt-Miner and CAnt-Miner rules are as follows:

   (a) Ant Miner

   IF Gender = 2 THEN MOS= Poor2

   IF Packet_Loss = 1 THEN MOS= Good2

   IF Codec = 1 THEN MOS= Poor2

   IF Packet_Loss = 3 THEN MOS= Good2

   IF Packet_Loss = 2 THEN MOS= Good1

   IF Packet_Loss = 5 THEN MOS= Good1

   IF Packet_Loss = 4 THEN MOS= Fair4

   IF Packet_Loss = 7 THEN MOS= Fair4

   IF Burst_Ratio = 1 THEN MOS= Poor3

   IF Packet_Loss = 6 THEN MOS= Fair3

   IF Packet_Loss = 8 AND Burst_Ratio = 4 THEN MOS= Fair4

   IF Burst_Ratio = 3 THEN MOS= Poor2

   IF Burst_Ratio = 4 THEN MOS= Poor2

   IF Burst_Ratio = 2 THEN MOS= Poor2

   IF Packet_Loss = 9 THEN MOS= Fair3

   IF Packet_Loss = 8 THEN MOS= Fair2

   IF Packet_Loss = 10 THEN MOS= Fair1

IF Packet_Loss = 16 THEN MOS= Fair1

IF Packet_Loss = 11 THEN MOS= Fair1

IF Packet_Loss = 50 THEN MOS= Poor1

IF Language = 10 THEN MOS= Poor1

IF Packet_Loss = 14 THEN MOS= Fair1

IF Language = 7 THEN MOS= Poor1

IF Language = 2 THEN MOS= Poor1

IF Packet_Loss = 15 THEN MOS= Fair1

IF Packet_Loss = 17 THEN MOS= Poor4

IF Packet_Loss = 19 THEN MOS= Poor4

IF Packet_Loss = 18 THEN MOS= Poor4

IF Packet_Loss = 49 THEN MOS= Poor1

IF Packet_Loss = 47 THEN MOS= Poor1

IF Packet_Loss = 48 THEN MOS= Poor1

IF Packet_Loss = 13 THEN MOS= Poor4

IF Packet_Loss = 12 THEN MOS= Poor4

IF Packet_Loss = 26 THEN MOS= Poor3

IF Packet_Loss = 23 THEN MOS= Poor3

IF Packet_Loss = 27 THEN MOS= Poor3

IF Packet_Loss = 24 THEN MOS= Poor3

IF Packet_Loss = 28 THEN MOS= Poor3

IF Language = 12 THEN MOS= Poor1

IF Packet_Loss = 25 THEN MOS= Poor3

IF Language = 9 THEN MOS= Poor1

IF Packet_Loss = 21 THEN MOS= Poor4

IF Packet_Loss = 20 THEN MOS= Poor4

IF Packet_Loss = 29 THEN MOS= Poor3

IF Packet_Loss = 22 THEN MOS= Poor3

IF Packet_Loss = 30 THEN MOS= Poor3

IF Language = 4 THEN MOS= Poor2

IF Language = 17 THEN MOS= Poor2

IF Language = 8 THEN MOS= Poor2

IF Language = 18 THEN MOS= Poor2

IF Packet_Loss = 45 THEN MOS= Poor1

IF Language = 5 THEN MOS= Poor2

IF Language = 20 THEN MOS= Poor2

IF Packet_Loss = 43 THEN MOS= Poor1

IF Packet_Loss = 34 THEN MOS= Poor3

IF Packet_Loss = 32 THEN MOS= Poor3

IF Packet_Loss = 31 THEN MOS= Poor3

IF Language = 1 THEN MOS= Poor2

IF Language = 14 THEN MOS= Poor2

IF Packet_Loss = 42 THEN MOS= Poor1

IF Packet_Loss = 46 THEN MOS= Poor1

IF Language = 15 THEN MOS= Poor2

IF Packet_Loss = 33 THEN MOS= Poor3

IF Language = 11 THEN MOS= Poor2

IF Packet_Loss = 44 THEN MOS= Poor1

IF Language = 6 THEN MOS= Poor2

IF Language = 13 THEN MOS= Poor2

IF Language = 19 THEN MOS= Poor2

IF Language = 3 THEN MOS= Poor2

(b) CAnt Miner

IF Packet_Loss $\geq$ 25.5 THEN MOS= Poor2

IF Packet_Loss <8.5 THEN MOS= Fair4

IF Gender = 2 AND Codec = 1 AND Packet_Loss $\geq$ 15.5 THEN MOS= Poor2

IF Packet_Loss $\geq$ 16.5 AND Burst_Ratio $\geq$ 1.5 THEN MOS= Poor3

IF Codec = 1 AND Packet_Loss $\geq$ 17.5 THEN MOS= Poor2

IF Gender = 2 AND Codec = 1 THEN MOS= Poor3

IF Packet_Loss $\geq$ 13.5 THEN MOS= Poor4

IF Codec = 2 AND Burst_Ratio $\geq$ 1.5 THEN MOS= Fair1

IF Codec = 1 AND Burst_Ratio <1.5 THEN MOS= Poor4

IF Packet_Loss $\geq$ 10.5 AND Codec = 1 THEN MOS= Fair1

IF Language = 7 AND Burst_Ratio <3.5 THEN MOS= Poor4

IF Codec = 1 THEN MOS= Fair3

IF Gender = 2 THEN MOS= Fair2

IF Packet_Loss $\geq$ 10.5 THEN MOS= Fair2

IF Packet_Loss <9.5 THEN MOS= Good1

IF Gender = 1 THEN MOS= Fair3

(c) CAnt Miner type2

IF Packet_Loss $\geq$ 20.5 THEN MOS= Poor2

IF Gender = 2 AND Packet_Loss $\geq$ 8.5 THEN MOS= Poor3

IF Packet_Loss $\geq$ 11.5 THEN MOS= Poor4

IF Burst_Ratio $\geq$ 1.5 AND Packet_Loss $\geq$ 4.5 THEN MOS= Fair2

IF Gender = 2 AND Codec = 1 THEN MOS= Fair1

IF Gender = 1 AND Packet_Loss <4.5 AND Burst_Ratio $\geq$ 2.5 THEN MOS= Good2

IF Packet_Loss <3.5 THEN MOS= Good1

IF Codec = 1 AND Packet_Loss $\geq$ 6.5 THEN MOS= Fair1

IF Gender = 2 THEN MOS= Fair3

IF Codec = 2 AND Packet_Loss <6.5 THEN MOS= Good1

IF Packet_Loss $\geq$ 6.5 THEN MOS= Fair3

IF Burst_Ratio $\geq$ 1.5 THEN MOS= Good1

IF Packet_Loss <5.5 THEN MOS= Fair3

IF Codec = 1 THEN MOS= Fair2

7. Seventh Fold

   In this fold Ant-Miner variation rules were produced. The Ant-Miner, CAnt-Miner and CAnt-Miner rules are as follows:

   (a) Ant Miner

   IF Gender = 2 THEN MOS= Poor2

   IF Packet_Loss = 1 THEN MOS= Good2

   IF Codec = 1 THEN MOS= Poor2

   IF Packet_Loss = 3 THEN MOS= Good2

   IF Packet_Loss = 2 THEN MOS= Good2

   IF Packet_Loss = 5 THEN MOS= Good1

   IF Packet_Loss = 4 THEN MOS= Fair4

   IF Packet_Loss = 6 THEN MOS= Fair4

   IF Packet_Loss = 7 THEN MOS= Fair4

   IF Packet_Loss = 9 THEN MOS= Fair3

   IF Burst_Ratio = 1 THEN MOS= Poor3

   IF Packet_Loss = 8 THEN MOS= Fair2

   IF Burst_Ratio = 5 THEN MOS= Poor2

   IF Burst_Ratio = 2 THEN MOS= Poor2

   IF Burst_Ratio = 3 THEN MOS= Poor2

   IF Packet_Loss = 12 THEN MOS= Fair1

   IF Packet_Loss = 11 THEN MOS= Fair1

   IF Packet_Loss = 10 THEN MOS= Fair1

   IF Packet_Loss = 17 THEN MOS= Fair1

IF Packet_Loss = 16 THEN MOS= Poor4

IF Packet_Loss = 15 THEN MOS= Poor4

IF Packet_Loss = 14 THEN MOS= Poor4

IF Packet_Loss = 20 THEN MOS= Poor4

IF Packet_Loss = 19 THEN MOS= Poor4

IF Packet_Loss = 13 THEN MOS= Poor4

IF Packet_Loss = 50 THEN MOS= Poor1

IF Packet_Loss = 49 THEN MOS= Poor1

IF Packet_Loss = 47 THEN MOS= Poor1

IF Packet_Loss = 45 THEN MOS= Poor1

IF Packet_Loss = 48 THEN MOS= Poor1

IF Packet_Loss = 46 THEN MOS= Poor1

IF Packet_Loss = 44 THEN MOS= Poor1

IF Packet_Loss = 21 THEN MOS= Poor3

IF Language = 13 THEN MOS= Poor3

IF Packet_Loss = 43 THEN MOS= Poor1

IF Language = 1 THEN MOS= Poor3

IF Language = 3 THEN MOS= Poor3

IF Packet_Loss = 42 THEN MOS= Poor1

IF Packet_Loss = 23 THEN MOS= Poor3

IF Packet_Loss = 18 THEN MOS= Poor3

IF Packet_Loss = 24 THEN MOS= Poor3

IF Language = 11 THEN MOS= Poor3

IF Packet_Loss = 41 THEN MOS= Poor1

IF Packet_Loss = 22 THEN MOS= Poor3

IF Packet_Loss = 25 THEN MOS= Poor3

IF Language = 15 THEN MOS= Poor2

IF Packet_Loss = 39 THEN MOS= Poor1

IF Packet_Loss = 34 THEN MOS= Poor2

IF Packet_Loss = 32 THEN MOS= Poor2

IF Language = 17 THEN MOS= Poor3

IF Packet_Loss = 35 THEN MOS= Poor2

IF Packet_Loss = 33 THEN MOS= Poor2

IF Language = 14 THEN MOS= Poor3

IF Packet_Loss = 40 THEN MOS= Poor1

IF Packet_Loss = 30 THEN MOS= Poor2

IF Language = 20 THEN MOS= Poor3

IF Packet_Loss = 31 THEN MOS= Poor2

IF Language = 4 THEN MOS= Poor3

IF Packet_Loss = 28 THEN MOS= Poor2

IF Packet_Loss = 29 THEN MOS= Poor2

IF Language = 2 THEN MOS= Poor1

IF Packet_Loss = 36 THEN MOS= Poor2

IF Packet_Loss = 26 THEN MOS= Poor2

IF Packet_Loss = 27 THEN MOS= Poor2

IF Packet_Loss = 37 THEN MOS= Poor2

IF Gender = 1 THEN MOS= Poor2

(b) CAnt Miner

IF Packet_Loss $\geq$ 25.5 THEN MOS= Poor2

IF Packet_Loss <8.5 THEN MOS= Fair4

IF Packet_Loss $\geq$ 15.5 AND Codec = 1 AND Gender = 2 THEN MOS= Poor2

IF Packet_Loss $\geq$ 16.5 AND Burst_Ratio $\geq$ 1.5 THEN MOS= Poor3

IF Codec = 1 AND Packet_Loss $\geq$ 16.5 THEN MOS= Poor2

IF Codec = 1 AND Gender = 2 THEN MOS= Poor3

IF Codec = 2 AND Packet_Loss $\geq$ 12.5 THEN MOS= Poor4

IF Gender = 1 AND Burst_Ratio <1.5 AND Packet_Loss ≥ 11.5 THEN MOS= Poor3

IF Packet_Loss <10.5 AND Gender = 1 THEN MOS= Fair3

IF Burst_Ratio ≥ 1.5 AND Codec = 2 THEN MOS= Fair1

IF Packet_Loss ≥ 13.5 AND Burst_Ratio <4.5 THEN MOS= Poor4

IF Packet_Loss ≥ 11.5 AND Codec = 1 THEN MOS= Fair1

IF Codec = 1 THEN MOS= Fair1

IF Language = 7 THEN MOS= Poor4

IF Gender = 2 AND Packet_Loss ≥ 9.5 THEN MOS= Fair2

IF Gender = 1 THEN MOS= Fair2

IF Gender = 2 THEN MOS= Fair3

(c) CAnt Miner type2

IF Burst_Ratio ≥ 1.5 AND Packet_Loss ≥ 21.5 THEN MOS= Poor2

IF Burst_Ratio <1.5 AND Packet_Loss ≥ 19.5 AND Codec = 1 THEN MOS= Poor1

IF Packet_Loss ≥ 11.5 THEN MOS= Poor3

IF Packet_Loss ≥ 5.5 THEN MOS= Fair2

IF Codec = 2 AND Language = 13 AND Gender = 1 AND Packet_Loss <1.5 THEN MOS= Good4

IF Gender = 1 AND Packet_Loss <2.5 THEN MOS= Good2

IF Gender = 2 AND Codec = 1 THEN MOS= Fair1

IF Gender = 1 THEN MOS= Fair4

IF Packet_Loss <2.5 THEN MOS= Good1

IF Packet_Loss ≥ 4.5 AND Burst_Ratio ≥ 2.5 THEN MOS= Fair2

IF Language = 20 THEN MOS= Fair2

IF Burst_Ratio ≥ 1.5 AND Packet_Loss ≥ 3.5 THEN MOS= Fair3

IF Language = 7 THEN MOS= Fair2

IF Language = 1 AND Packet_Loss <4.5 THEN MOS= Good2

IF Language = 4 THEN MOS= Good2

IF Language = 11 AND Packet_Loss <4.5 THEN MOS= Good1

IF Language = 2 THEN MOS= Good1

IF Language = 14 THEN MOS= Good1

IF Burst_Ratio ≥ 1.5 THEN MOS= Fair4

IF Packet_Loss <3.5 THEN MOS= Fair4

IF Packet_Loss <4.5 THEN MOS= Fair4

IF Codec = 2 THEN MOS= Fair4

8. Eighth Fold

In this fold Ant-Miner variation rules were produced. The Ant-Miner, CAnt-Miner and CAnt-Miner rules are as follows:

(a) Ant Miner

IF Gender = 2 THEN MOS= Poor2

IF Codec = 2 AND Language = 13 AND Packet_Loss = 1 THEN MOS= Good4

IF Packet_Loss = 1 THEN MOS= Good2

IF Codec = 2 THEN MOS= Poor2

IF Packet_Loss = 2 THEN MOS= Good1

IF Packet_Loss = 3 AND Language = 18 THEN MOS= Good3

IF Packet_Loss = 3 THEN MOS= Fair4

IF Packet_Loss = 4 THEN MOS= Fair4

IF Burst_Ratio = 3 THEN MOS= Poor2

IF Packet_Loss = 5 THEN MOS= Fair4

IF Burst_Ratio = 2 THEN MOS= Poor2

IF Burst_Ratio = 1 THEN MOS= Poor2

IF Burst_Ratio = 4 THEN MOS= Poor2

IF Packet_Loss = 8 THEN MOS= Fair3

IF Packet_Loss = 6 THEN MOS= Fair3

IF Packet_Loss = 7 THEN MOS= Fair3

IF Packet_Loss = 9 THEN MOS= Fair3

IF Packet_Loss = 11 THEN MOS= Fair2

IF Packet_Loss = 10 THEN MOS= Fair2

IF Packet_Loss = 16 THEN MOS= Fair2

IF Packet_Loss = 12 THEN MOS= Fair2

IF Packet_Loss = 15 THEN MOS= Fair1

IF Packet_Loss = 13 THEN MOS= Fair1

IF Packet_Loss = 18 THEN MOS= Fair1

IF Packet_Loss = 14 THEN MOS= Fair1

IF Packet_Loss = 23 THEN MOS= Poor4

IF Packet_Loss = 19 THEN MOS= Poor4

IF Packet_Loss = 25 THEN MOS= Poor4

IF Packet_Loss = 27 THEN MOS= Poor4

IF Packet_Loss = 26 THEN MOS= Poor4

IF Packet_Loss = 21 THEN MOS= Poor4

IF Packet_Loss = 20 THEN MOS= Poor4

IF Packet_Loss = 22 THEN MOS= Poor4

IF Packet_Loss = 17 THEN MOS= Poor4

IF Packet_Loss = 31 THEN MOS= Poor3

IF Packet_Loss = 29 THEN MOS= Poor3

IF Language = 1 THEN MOS= Poor3

IF Packet_Loss = 32 THEN MOS= Poor3

IF Packet_Loss = 33 THEN MOS= Poor3

IF Packet_Loss = 30 THEN MOS= Poor3

IF Packet_Loss = 28 THEN MOS= Poor3

IF Language = 3 THEN MOS= Poor3

IF Packet_Loss = 35 THEN MOS= Poor3

IF Packet_Loss = 24 THEN MOS= Poor3

IF Packet_Loss = 34 THEN MOS= Poor3

IF Language = 11 THEN MOS= Poor3

IF Language = 17 THEN MOS= Poor3

IF Language = 7 THEN MOS= Poor2

IF Language = 15 THEN MOS= Poor2

IF Language = 8 THEN MOS= Poor2

IF Language = 16 THEN MOS= Poor2

IF Language = 20 THEN MOS= Poor2

IF Packet_Loss = 36 THEN MOS= Poor3

IF Packet_Loss = 38 THEN MOS= Poor3

IF Language = 14 THEN MOS= Poor2

IF Language = 5 THEN MOS= Poor2

IF Language = 6 THEN MOS= Poor2

IF Language = 12 THEN MOS= Poor2

IF Packet_Loss = 37 THEN MOS= Poor3

IF Language = 13 THEN MOS= Poor2

IF Language = 9 THEN MOS= Poor2

IF Language = 4 THEN MOS= Poor2

IF Language = 10 THEN MOS= Poor2

IF Language = 19 THEN MOS= Poor2

IF Language = 2 THEN MOS= Poor2

IF Burst_Ratio = 5 THEN MOS= Poor2

(b) CAnt Miner

IF Packet_Loss $\geq$ 23.5 THEN MOS= Poor2

IF Packet_Loss <6.5 THEN MOS= Fair4

IF Burst_Ratio $\geq$ 1.5 AND Packet_Loss $\geq$ 15.5 THEN MOS= Poor3

IF Packet_Loss $\geq$ 11.5 AND Codec = 1 AND Burst_Ratio <1.5 THEN MOS= Poor2

IF Codec = 1 AND Gender = 2 AND Packet_Loss $\geq$ 9.5 THEN MOS= Poor3

IF Packet_Loss $\geq$ 11.5 AND Codec = 2 THEN MOS= Poor4

IF Burst_Ratio $\geq$ 1.5 AND Packet_Loss $\geq$ 9.5 THEN MOS= Fair1

IF Gender = 2 AND Codec = 1 THEN MOS= Poor4

IF Language = 7 THEN MOS= Poor4

IF Burst_Ratio $\geq$ 1.5 AND Codec = 2 THEN MOS= Fair2

IF Codec = 1 AND Burst_Ratio <2.5 THEN MOS= Fair1

IF Packet_Loss $\geq$ 8.5 THEN MOS= Fair2

IF Packet_Loss $\geq$ 7.5 AND Codec = 1 THEN MOS= Fair2

IF Gender = 1 AND Language = 3 THEN MOS= Good2

IF Language = 20 THEN MOS= Fair1

IF Gender = 1 AND Packet_Loss $\geq$ 7.5 THEN MOS= Good1

IF Language = 12 THEN MOS= Fair1

IF Codec = 1 THEN MOS= Fair3

IF Gender = 1 THEN MOS= Fair4

IF Packet_Loss <7.5 THEN MOS= Fair3

IF Gender = 2 THEN MOS= Fair3

(c) CAnt Miner type2

IF Packet_Loss $\geq$ 21.5 AND Burst_Ratio $\geq$ 1.5 THEN MOS= Poor2

IF Burst_Ratio <1.5 AND Packet_Loss $\geq$ 19.5 AND Codec = 1 THEN MOS= Poor1

IF Packet_Loss $\geq$ 10.5 AND Gender = 2 THEN MOS= Poor3

IF Packet_Loss $\geq$ 8.5 THEN MOS= Poor4

IF Gender = 2 AND Codec = 1 AND Packet_Loss $\geq$ 4.5 THEN MOS= Poor4

IF Language = 13 AND Packet_Loss <1.5 AND Gender = 1 AND Codec = 2 THEN MOS= Good4

IF Language = 7 AND Gender = 2 THEN MOS= Poor4

IF Codec = 2 AND Packet_Loss $\geq$ 4.5 AND Burst_Ratio $\geq$ 1.5 THEN MOS= Fair2

IF Packet_Loss $\geq$ 2.5 AND Codec = 1 AND Gender = 2 THEN MOS= Fair1

IF Packet_Loss $\geq$ 3.5 THEN MOS= Fair3

IF Gender = 1 AND Burst_Ratio $\geq$ 1.5 THEN MOS= Good2

IF Packet_Loss $\geq$ 1.5 AND Gender = 1 AND Codec = 2 THEN MOS= Good3

IF Gender = 2 AND Codec = 2 THEN MOS= Fair4

IF Gender = 1 AND Codec = 2 THEN MOS= Good2

IF Burst_Ratio $\geq$ 1.5 THEN MOS= Fair3

IF Gender = 2 THEN MOS= Fair2

IF Packet_Loss <1.5 THEN MOS= Good1

IF Packet_Loss $\geq$ 2.5 THEN MOS= Good1

IF Codec = 1 THEN MOS= Fair4

9. Ninth Fold

   In this fold Ant-Miner variation rules were produced. The Ant-Miner, CAnt-Miner and CAnt-Miner rules are as follows:

   (a) Ant Miner

   IF Gender = 2 THEN MOS= Poor2

   IF Packet_Loss = 1 AND Codec = 2 THEN MOS= Good3

   IF Codec = 2 THEN MOS= Poor2

   IF Packet_Loss = 1 THEN MOS= Good2

IF Packet_Loss = 2 THEN MOS= Good1

IF Packet_Loss = 4 THEN MOS= Fair4

IF Packet_Loss = 3 THEN MOS= Fair4

IF Packet_Loss = 5 THEN MOS= Fair4

IF Packet_Loss = 6 THEN MOS= Fair3

IF Packet_Loss = 7 THEN MOS= Fair3

IF Burst_Ratio = 3 THEN MOS= Poor2

IF Burst_Ratio = 2 THEN MOS= Poor2

IF Burst_Ratio = 1 THEN MOS= Poor2

IF Burst_Ratio = 4 THEN MOS= Poor2

IF Packet_Loss = 8 THEN MOS= Fair3

IF Packet_Loss = 9 THEN MOS= Fair3

IF Packet_Loss = 11 THEN MOS= Fair2

IF Packet_Loss = 10 THEN MOS= Fair2

IF Packet_Loss = 14 THEN MOS= Fair2

IF Packet_Loss = 16 THEN MOS= Fair2

IF Packet_Loss = 12 THEN MOS= Fair2

IF Packet_Loss = 18 THEN MOS= Fair1

IF Packet_Loss = 15 THEN MOS= Fair1

IF Packet_Loss = 17 THEN MOS= Fair1

IF Packet_Loss = 13 THEN MOS= Fair1

IF Packet_Loss = 23 THEN MOS= Poor4

IF Packet_Loss = 19 THEN MOS= Poor4

IF Packet_Loss = 21 THEN MOS= Poor4

IF Packet_Loss = 26 THEN MOS= Poor4

IF Packet_Loss = 20 THEN MOS= Poor4

IF Packet_Loss = 25 THEN MOS= Poor4

IF Packet_Loss = 22 THEN MOS= Poor4

IF Packet_Loss = 27 THEN MOS= Poor4

IF Packet_Loss = 29 THEN MOS= Poor3

IF Packet_Loss = 31 THEN MOS= Poor3

IF Language = 1 THEN MOS= Poor3

IF Packet_Loss = 30 THEN MOS= Poor3

IF Packet_Loss = 35 THEN MOS= Poor3

IF Language = 3 THEN MOS= Poor3

IF Packet_Loss = 24 THEN MOS= Poor3

IF Packet_Loss = 28 THEN MOS= Poor3

IF Packet_Loss = 33 THEN MOS= Poor3

IF Packet_Loss = 32 THEN MOS= Poor3

IF Packet_Loss = 34 THEN MOS= Poor3

IF Language = 18 THEN MOS= Poor3

IF Language = 11 THEN MOS= Poor3

IF Packet_Loss = 46 THEN MOS= Poor2

IF Language = 17 THEN MOS= Poor3

IF Packet_Loss = 44 THEN MOS= Poor2

IF Packet_Loss = 42 THEN MOS= Poor2

IF Packet_Loss = 48 THEN MOS= Poor2

IF Packet_Loss = 40 THEN MOS= Poor2

IF Packet_Loss = 43 THEN MOS= Poor2

IF Packet_Loss = 39 THEN MOS= Poor2

IF Packet_Loss = 45 THEN MOS= Poor2

IF Packet_Loss = 50 THEN MOS= Poor2

IF Packet_Loss = 47 THEN MOS= Poor2

IF Packet_Loss = 38 THEN MOS= Poor2

IF Packet_Loss = 41 THEN MOS= Poor2

IF Packet_Loss = 49 THEN MOS= Poor2

IF Packet_Loss = 37 THEN MOS= Poor2

IF Codec = 1 THEN MOS= Poor2

(b) CAnt Miner

IF Packet_Loss $\geq$ 24.5 THEN MOS= Poor2

IF Packet_Loss <8.5 THEN MOS= Fair4

IF Burst_Ratio $\geq$ 1.5 AND Packet_Loss $\geq$ 16.5 THEN MOS= Poor3

IF Packet_Loss $\geq$ 12.5 AND Codec = 1 AND Burst_Ratio <1.5 THEN MOS= Poor2

IF Codec = 1 AND Gender = 2 THEN MOS= Poor3

IF Codec = 2 AND Packet_Loss $\geq$ 12.5 THEN MOS= Poor4

IF Codec = 2 AND Burst_Ratio $\geq$ 1.5 THEN MOS= Fair1

IF Codec = 1 AND Packet_Loss $\geq$ 11.5 AND Burst_Ratio <3.5 THEN MOS= Poor4

IF Packet_Loss $\geq$ 10.5 AND Codec = 1 THEN MOS= Fair1

IF Language = 7 THEN MOS= Poor4

IF Codec = 2 AND Gender = 1 THEN MOS= Fair4

IF Burst_Ratio $\geq$ 2.5 THEN MOS= Fair3

IF Language = 4 THEN MOS= Fair4

IF Gender = 1 AND Burst_Ratio <1.5 THEN MOS= Poor4

IF Language = 10 THEN MOS= Poor4

IF Gender = 1 THEN MOS= Fair1

IF Packet_Loss $\geq$ 10.5 THEN MOS= Fair2

IF Packet_Loss $\geq$ 9.5 THEN MOS= Fair2

IF Gender = 2 THEN MOS= Fair3

(c) CAnt Miner type2

IF Burst_Ratio $\geq$ 1.5 AND Packet_Loss $\geq$ 21.5 THEN MOS= Poor2

IF Codec = 1 AND Packet_Loss $\geq$ 21.5 THEN MOS= Poor1

IF Packet_Loss $\geq$ 10.5 THEN MOS= Poor3

IF Packet_Loss $\geq$ 5.5 AND Gender = 2 AND Codec = 1 THEN MOS= Poor4

IF Gender = 1 AND Packet_Loss <4.5 THEN MOS= Good1

IF Packet_Loss $\geq$ 6.5 THEN MOS= Fair2

IF Language = 7 AND Codec = 1 THEN MOS= Poor4

IF Language = 14 AND Packet_Loss <1.5 AND Codec = 2 THEN MOS= Good3

IF Codec = 2 AND Packet_Loss <2.5 AND Language = 1 THEN MOS= Good3

IF Burst_Ratio $\geq$ 1.5 AND Gender = 2 AND Packet_Loss $\geq$ 3.5 THEN MOS= Fair2

IF Language = 2 AND Gender = 2 AND Packet_Loss <1.5 AND Codec = 2 THEN MOS= Good3

IF Packet_Loss $\geq$ 3.5 THEN MOS= Fair3

IF Gender = 2 AND Codec = 2 THEN MOS= Fair4

IF Language = 4 THEN MOS= Good1

IF Packet_Loss $\geq$ 2.5 THEN MOS= Fair1

IF Packet_Loss $\geq$ 1.5 THEN MOS= Fair3

IF Burst_Ratio $\geq$ 3.5 THEN MOS= Fair2

IF Burst_Ratio <2.5 THEN MOS= Fair2

IF Gender = 2 THEN MOS= Fair3

10. Tenth Fold

    In this fold Ant-Miner variation rules were produced. The Ant-Miner, CAnt-Miner and CAnt-Miner rules are as follows:

    (a) Ant Miner

IF Gender = 2 THEN MOS= Poor2

IF Packet_Loss = 1 THEN MOS= Good2

IF Packet_Loss = 2 THEN MOS= Good1

IF Codec = 2 THEN MOS= Poor2

IF Packet_Loss = 3 THEN MOS= Fair4

IF Packet_Loss = 4 THEN MOS= Fair4

IF Language = 3 AND Packet_Loss = 5 THEN MOS= Good2

IF Packet_Loss = 5 THEN MOS= Fair4

IF Packet_Loss = 6 AND Language = 11 THEN MOS= Good1

IF Packet_Loss = 6 THEN MOS= Fair3

IF Packet_Loss = 7 THEN MOS= Fair3

IF Burst_Ratio = 2 THEN MOS= Poor2

IF Burst_Ratio = 5 AND Packet_Loss = 9 THEN MOS= Fair4

IF Burst_Ratio = 3 THEN MOS= Poor2

IF Burst_Ratio = 4 THEN MOS= Poor2

IF Packet_Loss = 8 THEN MOS= Fair3

IF Burst_Ratio = 1 THEN MOS= Poor2

IF Packet_Loss = 11 THEN MOS= Fair2

IF Packet_Loss = 10 THEN MOS= Fair2

IF Packet_Loss = 15 THEN MOS= Fair1

IF Packet_Loss = 18 THEN MOS= Fair1

IF Packet_Loss = 13 THEN MOS= Fair1

IF Packet_Loss = 20 THEN MOS= Fair1

IF Packet_Loss = 17 THEN MOS= Fair1

IF Packet_Loss = 14 THEN MOS= Fair1

IF Packet_Loss = 12 THEN MOS= Fair1

IF Packet_Loss = 23 THEN MOS= Poor4

IF Packet_Loss = 25 THEN MOS= Poor4

IF Packet_Loss = 21 THEN MOS= Poor4

IF Language = 11 THEN MOS= Poor4

IF Packet_Loss = 19 THEN MOS= Poor4

IF Packet_Loss = 22 THEN MOS= Poor4

IF Packet_Loss = 26 THEN MOS= Poor4

IF Packet_Loss = 16 THEN MOS= Poor4

IF Packet_Loss = 33 THEN MOS= Poor3

IF Packet_Loss = 29 THEN MOS= Poor3

IF Packet_Loss = 30 THEN MOS= Poor3

IF Packet_Loss = 28 THEN MOS= Poor3

IF Language = 1 THEN MOS= Poor3

IF Packet_Loss = 24 THEN MOS= Poor3

IF Language = 3 THEN MOS= Poor3

IF Packet_Loss = 32 THEN MOS= Poor3

IF Packet_Loss = 35 THEN MOS= Poor3

IF Packet_Loss = 31 THEN MOS= Poor3

IF Packet_Loss = 27 THEN MOS= Poor3

IF Packet_Loss = 37 THEN MOS= Poor3

IF Packet_Loss = 34 THEN MOS= Poor3

IF Packet_Loss = 46 THEN MOS= Poor2

IF Language = 18 THEN MOS= Poor3

IF Packet_Loss = 43 THEN MOS= Poor2

IF Packet_Loss = 44 THEN MOS= Poor2

IF Language = 17 THEN MOS= Poor3

IF Packet_Loss = 40 THEN MOS= Poor2

IF Packet_Loss = 42 THEN MOS= Poor2

IF Packet_Loss = 47 THEN MOS= Poor2

IF Packet_Loss = 41 THEN MOS= Poor2

IF Packet_Loss = 50 THEN MOS= Poor2

IF Language = 13 THEN MOS= Poor3

IF Packet_Loss = 39 THEN MOS= Poor2

IF Packet_Loss = 48 THEN MOS= Poor2

IF Packet_Loss = 45 THEN MOS= Poor2

IF Packet_Loss = 38 THEN MOS= Poor2

IF Packet_Loss = 49 THEN MOS= Poor2

IF Burst_Ratio = 5 THEN MOS= Poor2

(b) CAnt Miner

IF Packet_Loss $\geq$ 26.5 THEN MOS= Poor2

IF Burst_Ratio $\geq$ 1.5 AND Packet_Loss $\geq$ 14.5 THEN MOS= Poor3

IF Codec = 1 AND Burst_Ratio <1.5 AND Packet_Loss $\geq$ 11.5 THEN MOS= Poor2

IF Packet_Loss <5.5 THEN MOS= Good1

IF Gender = 2 AND Packet_Loss $\geq$ 8.5 THEN MOS= Poor4

IF Gender = 1 AND Packet_Loss <10.5 THEN MOS= Fair3

IF Gender = 1 AND Burst_Ratio $\geq$ 1.5 THEN MOS= Fair1

IF Codec = 2 AND Packet_Loss <17.5 THEN MOS= Fair2

IF Language = 4 AND Burst_Ratio $\geq$ 1.5 THEN MOS= Fair4

IF Packet_Loss $\geq$ 7.5 THEN MOS= Poor4

IF Burst_Ratio <1.5 THEN MOS= Poor3

IF Language = 1 THEN MOS= Fair3

IF Language = 9 THEN MOS= Fair3

IF Language = 14 AND Burst_Ratio $\geq$ 2.5 THEN MOS= Fair3

IF Language = 2 THEN MOS= Fair3

IF Packet_Loss <6.5 THEN MOS= Fair1

IF Burst_Ratio <4.5 THEN MOS= Fair1

IF Gender = 2 THEN MOS= Poor4

(c) CAnt Miner type2

IF Burst_Ratio $\geq$ 1.5 AND Packet_Loss $\geq$ 21.5 THEN MOS= Poor2

IF Codec = 1 AND Packet_Loss $\geq$ 21.5 THEN MOS= Poor1

IF Packet_Loss $\geq$ 10.5 THEN MOS= Poor3

IF Burst_Ratio $\geq$ 1.5 AND Packet_Loss $\geq$ 4.5 THEN MOS= Fair2

IF Codec = 1 AND Packet_Loss $\geq$ 4.5 AND Gender = 2 THEN MOS= Poor3

IF Gender = 2 AND Codec = 1 THEN MOS= Fair1

IF Gender = 1 AND Burst_Ratio $\geq$ 1.5 THEN MOS= Good1

IF Gender = 2 THEN MOS= Fair4

IF Packet_Loss <5.5 AND Codec = 2 THEN MOS= Good2

IF Codec = 1 AND Packet_Loss $\geq$ 5.5 THEN MOS= Fair2

IF Gender = 1 AND Codec = 2 THEN MOS= Fair3

IF Packet_Loss $\geq$ 3.5 THEN MOS= Fair3

IF Packet_Loss <1.5 THEN MOS= Good1

IF Packet_Loss $\geq$ 2.5 THEN MOS= Good1

IF Codec = 1 THEN MOS= Fair4

إستخدام مملكة النمل الإصطناعية لتقييم جودة الصوت المنقول عبر شبكة الإنترنت

إعداد

رامي سليمان عايد الخوالدة

المشرف

الدكتور موسى توفيق الأخرس

المشرف المشارك

الدكتور خليل محمد علي الهندي

قدمت هذه الرسالة استكمالا لمتطلبات الحصول على درجة الماجستير في علوم الحاسوب

الملخص

الإشارات الصوتية تقليديا تُنقل عبر شبكة اتصالات التوزيع العام. حتى نستغل شبكة البروتوكولات المعتمدة على الإنترنت، حزمة الصوت يمكن أن تُنقل عبر شبكة البروتوكولات المعتمدة على شبكة الإنترنت والتي أصلاً خصصت لنقل البيانات، لهذا شبكة البروتوكولات

المعتمدة على الإنترنت يُمكن أن تستخدم لنقل نوعين من الحزم البيانات النصية والصوت كما في تطبيق الصوت المنقول عبر بروتوكولات الإنترنت. على الرغم من أن هذه الشبكة غير مصممه لدعم الاتصالات الصوتية في الوقت الحقيقي أو المتزامن لأن خصائص متغيراتها (مثل التأخير ، التنوع في التأخير ، و الحزم المفقودة) تؤدي إلى تدهور في نوعية الصوت. تحديات رئيسية في هذا النوع من الشبكات وهي في كيفية قياس أو التنبؤ بنوعية الصوت بشكل دقيق وفعال لأهداف المراقبة و التحكم بنوعية الخدمات وذلك لنؤكد متطلبات قانونية و تقنية و تجارية.

يوجد هنالك طريقتين حتى نقوم بقياس نوعية الصوت المنقول طريق موضوعية و هدفيه. الطرق الموضوعية تعتمد على المستخدم لقياس نوعية الصوت. لذلك نوعية الكلام هو بحد ذاته موضوعي حيث يعتمد على إدراك المُستمع، حيث تعتبر الفحوصات الموضوعية من أفضل الطرق الموثوق فيها لتقييم جودة الكلام. تلعب طرق تقييم نوعية الكلام الهدفية دور هام في الأبحاث الحالية لمعالجة محددات الفحص الموضوعي من خلال محاكاة آراء الفاحصين البشريين عل شكل خوارزميات أو استخدام التقييم الأوتوماتيكي الآلي. تقييم الجودة الهدفي ما زال يحتل انتباه مقدر. الأغلبية من هذه الطرق الهدفيه تعتمد على مقارنة الإشارات الصوتية الداخلة (الأصلية) والخارجة (المتلقاة من الشبكة) أنا أعني الطرق الاقتحامية والتي تقييم نوعية الكلام من خلال قياس الخلل بين الإشارات الصوتية الداخلة والخارجة من الشبكة أو الطرق غير الاقتحامية والتي تقييم نوعية الكلام معتمدة فقط على الإشارات الصوتية الخارجة أو الإشارات المختلة من الشبكة، و تحول القيم التي فيها خلل إلى قياس نوعية متوقع.

سلوك النمل الحقيقي في العثور على المأكل يمكن أن يُنمذج بشكل خوارزمي لحل بعض المشكلات التي تتطلب حلول أمثل. لذلك استخدمت طريقة مملكة النمل الاصطناعية في هذه الرسالة لإنشاء طريقة هدفيه غير مقتحمة مبتكرة لتقييم جودة الصوت المنقول عبر شبكة البروتوكولات المعتمدة على الإنترنت. طريقة مملكة النمل الاصطناعية التي استخدمت لهذا الهدف كانت طريقة النمل الاصطناعية المنقبة والتي تنظر في بيانات مدربة لإيجاد أفضل القواعد. هذه القواعد الأفضل تمثل العلاقة الداخلة في البيانات المدربة والتي سوف تستخدم كنموذج، للتنبؤ بقيمة نوعية الصوت الغير معروفة في المستقبل من مجموعة من قيم الخصائص الملاحظة. البيانات المدربة تُبنى من خلال التقاط مجموعة متغيرات من شبكة البروتوكولات المعتمدة على الإنترنت و من ثم تأخذ نوعية الصوت المحسوب اعتماداً

على هذه المتغيرات كملصق صفي، نتيجة لذلك القواعد الأفضل يمكن أن تمثل كقواعد (إذافإن) والتي تتكون من رابطين القاعدة السابقة (جزء إذا) والقاعدة التابعة (جزء فإن) حيث أن القاعدة السابقة تتضمن المتغيرات مع قيمها والقاعدة التابعة تتضمن نوعية الصوت كملصق صفي. هذه الرسالة تنجز نوعين من التجارب الأولى تدرس تأثير المتغيرات بشكل مستقل أو مع بعضها البعض على نوعية الصوت. دقة الطريقة المقترحة قورنت مع مهمات تصنيفية أخرى مثل طرق الانحدار الخطي وغير الخطي والشبكة العصبية الإصطناعية والتي أظهرت أن طريقة الشبكة العصبية الاصطناعية أكثر دقة من هذه الطرق وطرق أخرى مثل النمل الاصطناعي المنقب، حيث أنّ معامل الارتباط بيرسون للشبكة العصبية الإصطناعية ٩٥٥، ٠ ، يبين هذه الدقه، حيث دقة الانحدار الغير خطي ، النمل المنقب المستمر ، النمل المنقب المستمر النوع ٢، الانحدار الخطي ، النمل المنقب هي ٠،٩١٧، ٠،٨٧٥، ٠،٨٦٦، ٠،٨٤٨ و ٠،٤٧٠ على التوالي، غير أنّ القواعد المنتجة من النمل الاصطناعي المنقب تستخدم نموذج (إذافإن) والتي لها أهمية لأنها مدركة أكثر ومقروءة بشكل مناسب للمستخدم أكثر من طرق الانحدار والشبكة العصبية الاصطناعية.