VOIP TECHNOLOGIES

Edited by **Shigeru Kashihara**

INTECHWEB.ORG

VoIP Technologies
Edited by Shigeru Kashihara

Published by InTech
Janeza Trdine 9, 51000 Rijeka, Croatia

Publishing Process Manager Jelena Marusic
Technical Editor Teodora Smiljanic
Cover Designer Martina Sirotic
Image Copyright WitR, 2010. Used under license from Shutterstock.com

First published February, 2011
Printed in India

A free online edition of this book is available at www.intechopen.com
Additional hard copies can be obtained from orders@intechweb.org

VoIP Technologies, Edited by Shigeru Kashihara
p. cm.
ISBN 978-953-307-549-5

Contents

Preface IX

Chapter 1 **VoIP Quality Assessment Technologies** 1
Mousa AL-Akhras and Iman AL Momani

Chapter 2 **Assessment of Speech Quality in VoIP** 27
Zdenek Becvar, Lukas Novak and Michal Vondra

Chapter 3 **Enhanced VoIP by Signal Reconstruction and Voice Quality Assessment** 45
Filipe Neves, Salviano Soares, Pedro Assunção and Filipe Tavares

Chapter 4 **An Introduction to VoIP: End-to-End Elements and QoS Parameters** 79
H. Toral-Cruz, J. Argaez-Xool, L. Estrada-Vargas and D. Torres-Roman

Chapter 5 **Influences of Classical and Hybrid Queuing Mechanisms on VoIP's QoS Properties** 95
Sasa Klampfer, Amor Chowdhury, Joze Mohorko and Zarko Cucej

Chapter 6 **VoIP System for Enterprise Network** 127
Moo Wan Kim and Fumikazu Iseki

Chapter 7 **An Opencores /Opensource Based Embedded System-on-Chip Platform for Voice over Internet** 145
Sabrina Titri, Nouma Izeboudjen, Fatiha Louiz, Mohamed Bakiri, Faroudja Abid, Dalila Lazib and Leila Sahli

Chapter 8 **Experimental Characterization of VoIP Traffic over IEEE 802.11 Wireless LANs** 173
Paolo Dini, Marc Portolés-Comeras, Jaume Nin-Guerrero and Josep Mangues-Bafalluy

Chapter 9 **VoIP Over WLAN: What About the Presence of Radio Interference?** 197
Leopoldo Angrisani, Aniello Napolitano and Alessandro Sona

Chapter 10 **VoIP Features Oriented Uplink Scheduling Scheme in Wireless Networks 219**
Sung-Min Oh and Jae-Hyun Kim

Chapter 11 **Scheduling and Capacity of VoIP Services in Wireless OFDMA Systems 237**
Jaewoo So

Chapter 12 **Reliable Session Initiation Protocol 253**
Harold Zheng, and Sherry Wang

Chapter 13 **Multi-path Transmission, Selection and Handover Mechanism for High-Quality VoIP 277**
Jingyu Wang, Jianxin Liao and Xiaomin Zhu

Chapter 14 **End-to-End Handover Management for VoIP Communications in Ubiquitous Wireless Networks 295**
Shigeru Kashihara, Muhammad Niswar, Yuzo Taenaka, Kazuya Tsukamoto, Suguru Yamaguchi and Yuji Oie

Chapter 15 **Developing New Approaches for Intrusion Detection in Converged Networks 321**
Juan C. Pelaez

Preface

Voice over IP (VoIP) is undoubtedly a powerful and innovative communication tool. Compared with the public switched telephone network (PSTN), VoIP offers the benefit of reducing communication and infrastructure cost. This makes it possible for everyone to easily keep in touch with family, friends, and clients around the world. Furthermore, VoIP has the potential to create new and attractive communication tools by integrating with various applications, such as Web systems, presentation software, and photo viewers. In the near future, we expect VoIP to provide a rich multimedia communication service.

However, voice communication over the Internet is inherently less reliable than PSTN. Since the Internet essentially works as a best-effort network without a Quality of Service (QoS) guarantee, and since voice data cannot be retransmitted, voice quality may suffer from packet loss, delay, and jitter due to interference from other data packets. In wireless networks, such problems may be compounded by various characteristics of wireless media, including reduction of signal strength and radio interference. Additionally, we need to pay closer attention to security issues related to VoIP communications. Thus, VoIP technologies are challenging research issues.

This book comprises 15 chapters and encompasses a wide range of VoIP research, from VoIP quality assessment to security issues. Much of the content is focused on the key areas of VoIP performance investigation and enhancement. Each chapter includes various approaches that illustrate how VoIP aspires to be a powerful and reliable communication tool. We hope that you will enjoy reading these diverse studies, and will find a lot of useful information about VoIP technologies. Finally, I would like to thank all authors of the chapters for their great contributions.

Shigeru Kashihara
Nara Institute of Science and Technology
Japan

1

VoIP Quality Assessment Technologies

Mousa AL-Akhras and Iman AL Momani
The University of Jordan
Jordan

1. Introduction

Circuit Switching technology has been in use for long time by traditional Public Switched Telephone Network (PSTN) carriers for carrying voice traffic. Before users may communicate in circuit switching network, a dedicated channel or circuit is established from the sender to the receiver and that path is selected over the most efficient route using intelligent switches. Accordingly, it is not necessary for a phone call from the same sender to the same receiver to take the same route every time a phone call is made.

During call setup once the route is determined, that path or circuit stays fixed throughout the call and the necessary resources across the path are allocated to the phone call from the beginning to the end of the call. The established circuit cannot be used by other callers until the circuit is released, it remains unavailable to other users even when no actual communication is taking place, therefore, circuit switching is carrying voice with high fidelity from source to destination (Collins, 2003). Circuit switching is like having a dedicated railroad track with only one train, the call, is permitted on the track at one time.

Today's commercial telephone networks that based on circuit switching technology have a number of attractive features, including: Availability, Capacity, Fast Response and High Quality (Collins, 2003). The quality is the main focus of this chapter.

One alternative technology to circuit switching telephone networks for carrying voice traffic is to use data-centric packet switching networks such as Internet Protocol (IP) networks. In packet switching technology, no circuit is built from the sender to the receiver and packets are sent over the most effective route at time of sending that packet, consequently different packets may take different routes from the same sender to the same receiver within the same session.

Transmitting Voice over IP (VoIP) networks is an important application in the world of telecommunication and is an active area of research. Networks of the future will use IP as the core transport network as IP is seen as the long-term carrier for all types of traffic including voice and video. VoIP will become the main standard for third generation wireless networks (Bos & Leroy, 2001; Heiman, 1998).

Transmission of voice as well as data over IP networks seems an attractive solution as voice and data services can be integrated which makes creation of new and innovative services possible. This provides promises of greater flexibility and advanced services than the traditional telephony with greater possibility for cost reduction in phone calls. VoIP also has other advantages, including: number portability, lower equipment cost, lower bandwidth requirements, lower operating and management expenses, widespread availability of IP, and other advantages (Collins, 2003; Heiman, 1998; Low, 1996; Moon et al., 2000; Rosenberg et al., 1999). VoIP can be used in many applications, including: call centre integration, directory services over telephones, IP video conferencing, fax over IP, and Radio/ TV Broadcasting (Collins, 2003; Miloslavski et al., 2001; Ortiz, 2004; Schulzrinne & Rosenberg, 1999).

VoIP technology was adopted by many operators as an alternative to circuit switching technology. This adoption was motivated by the above advantages and to share some of the high revenue achieved by telecommunication companies. However, to be able to compete with the highly reputable PSTN networks, VoIP networks should be able to achieve comparable quality to that achieved by PSTN networks. Although VoIP services often offer much cheaper solutions than what PSTN does, but regardless of how low the cost of the service is, it is the user perception of the quality what matters. If the quality of the voice is poor, the user of the traditional telephony will not be attracted to the VoIP service regardless of how cheap the service is. This comes from the fact that customers who are used to the high-quality telephony networks, expect to receive a comparable quality from any potential competitor.

IP networks were originally designed to carry non real-time traffic such as email or file transfer and they are doing this task very well, however, as IP networks are characterised by being best-effort networks with no guarantee of delivery as no circuit is established between the sender and the receiver, therefore they are not particularly appropriate to support real-time applications such as voice traffic in addition to data traffic. The best-effort nature of IP networks causes several degradations to the speech signal before it reaches its destination. These degradations arise because of the time-varying characteristics (e.g. packet loss, delay, delay variation (jitter), sharing of resources) of IP networks.

These characteristics which are normal to data traffic, cause serious deterioration to the real-time traffic and prevent IP networks from providing the high quality speech often provided by traditional PSTN networks for voice services. Sharing of resources in IP networks causes no resources to be dedicated to the voice call in contrast to what is happening in traditional circuit switching telephony such as PSTN where the required resources are allocated to the phone call from the start to the end. With the absence of resource dedication, many problems are inevitable in IP networks.

Among the problems is packet loss which occurs due to the overflow in intermediate routers or due to the long time taken by packets to reach their destinations (Collins, 2003). Real-time applications are also sensitive to delay since they require voice packets to arrive at the receiving end within a certain upper bound to allow interactivity of the voice call (ITU-T, 2003a;b). Also, due to their best-effort nature, packets could take different routes from the same source to the same destination within the same session which causes packets interarrival time to vary, a phenomenon known as jitter. Due to the problem of jitter, it is not easy to play packets in a steady fashion to the listener (Narbutt & Murphy, 2004; Tseng & Lin, 2003; Tseng et al., 2004). The above challenges cause degradation to the quality of the received speech signal before it reaches its destination. Many solutions have been proposed to alleviate these problems and the quality of the received speech signal as perceived by the end user is greatly affected by the effectiveness of these solutions.

Another approach is to reserve resources across the path from the sender to the receiver. A mechanism called Call Admission Control (CAC) is needed to determine whether to accept a call request if it is possible to allocate the required bandwidth and maintain the given QoS target for all existing calls, or otherwise to reject the call (Mase, 2004). Among the solutions that have been proposed to implement CAC and to manage the available bandwidth efficiently are: Resource Reservation Protocol (RSVP), Differentiated Service (DiffServ), MultiProtocol Label Switching (MPLS), and End-to-end Measurement Based Admission Control (EMBAC). Reserving resources is difficult and very expensive proposal as it requires changes to all routers across the network which is inapplicable in non-managed networks such as the Internet.

Therefore, it is important to measure the quality of VoIP applications in live networks and take appropriate actions when necessary. This importance comes from legal, commercial and technical reasons. Measurement of the quality would be a necessity as customers and companies are bound by a service level agreement usually requiring the company to provide a certain level of quality, otherwise, customers may sue the companies for poor quality. Also, measuring the quality gives the chance to network administrators to overcome temporal problems that could affect the quality of ongoing voice calls. Measurement of the quality also allows service providers to evaluate their own and their competitors' service using a standard scale. It is also a strong indicator of users' satisfaction of the service provided (Takahashi et al., 2004; Zurek et al., 2002).

To this end, a specialised mechanism is required for measuring the speech quality accurately. One of driving forces in the world of telecommunication is the International Telecommunication Union (ITU). ITU is the leading United Nations (UN) agency for information and communication technology. As the global focal point for governments and the private sector in developing telecommunication networks and services, ITU's role is to help the world communicate. ITU - Telecommunication Standardisation Sector (ITU-T, http://www.itu.int/ITU-T/) is a permanent organ of the ITU that plays a driving force role toward standardising and regulating international telecommunications worldwide. Toward this goal, ITU-T study technical, operating and tariff questions and produce standards under the name of Recommendations for the purpose of standardising telecommunications worldwide. ITU-T's Recommendations are divided into categories that are identified by a single letter, referred to as the series, and Recommendations are numbered within each series, for example P.800 (ITU-T, 1996b). ITU-T has a formal recognition as it is part of ITU which is a UN Organisation (UNO).

Many ITU-T Recommendations are concerned with standardising the measurement of speech quality for voice services, many of these standards are considered in this chapter. Speech quality in ITU-T standards is expressed as Mean Opinion Score (MOS) which ranges between 1 and 5, with 1 corresponds to poor quality and 5 to excellent quality.

Some standards measure the speech quality or the MOS **subjectively** by setting lab conditions and asking subjects to listen to the speech signal and give their estimation of the quality in terms of MOS. This method is standardised in ITU-T Recommendation P.800 (ITU-T, 1996b). Other methods are **objective** that depend on comparison of the received signal with the original signal to measure the perceived quality in terms of MOS, these methods are known as **intrusive** methods as they require the injection of the original signal to analyse the distortion of the received signal. The most recent method for measuring the speech quality intrusively is known as Perceptual Evaluation of Speech Quality (PESQ). PESQ is standardised as ITU-T Recommendation P.862 (ITU-T, 2001). Yet another **objective** category depends on either the received signal or the networking parameters to estimate the quality **non-intrusively** without the need for the original signal. The two main methods in this category are Recommendation P.563 (ITU-T, 2004) and the E-model as defined in ITU-T Recommendation G.107 (ITU-T, 2009). Many other standards and methods have been proposed by other organisations, other researchers, and the authors of this chapter independent of the ITU-T, these attempts will be discussed in detail later in the chapter.

The selection of a method for VoIP quality assessment should take the characteristics of IP networks and voice calls into consideration. Such characteristics that affect the selection include the requirement to measure the quality of live-traffic while the network is running in a real environment during a voice call. To able to do this, an objective solution that measures the quality without human interference and depending on the received signal at the receiver side without the need for the original speech signal at the sender side; i.e. a non-intrusive measurement is needed.

This chapter aims to serve as a reference and survey for readers interested in the area of speech quality assessment in VoIP networks. The rest of this chapter is organised as follows: Section 2 categorises speech quality assessment techniques and discusses the main requirements of an applicable technique in VoIP environment. Sections 3 and 4 discusses subjective and objective quality assessment technologies, respectively. To avoid ambiguity, different qualifiers are used to distinguish between different quality measurement methods and presented in section 5. Conclusions and possibilites for future work are given in section 6.

2. Categories of VoIP quality assessment technologies

VoIP quality assessment methods can be categorised into either subjective methods or objective methods. Objective methods can be either intrusive or non-intrusive. Non-intrusive methods can be either signal-based or parametric-based. Figure 1 depicts different classifications.

The primary criterion for voice and video communication is subjective quality, the user's perceptions of service quality. A subjective quality assessment method is used to measure the quality. Subjective quality factors affect the quality of service of VoIP, among those factors are: packet loss, delay, jitter, loudness, echo, and codec distortion. To measure the subjective quality, a subjective quality assessment method is used, the most widely accepted metric is the Mean Opinion Score (MOS) as defined by ITU-T Recommendation P.800 (ITU-T, 1996b).

However, although subjective quality assessment is the most reliable method, it is also time-consuming and expensive as any other subjective test. Thus other methods to automatically estimate quality objectively should be considered. This can be done intrusively by comparing the reference signal with the degraded signal or non-intrusively utilising physical quality parameters or the received signal without using the reference signal.

The applicability of any solution for measuring the speech quality in VoIP networks should take into consideration the nature of IP networks and the characteristics of voice traffic. Among the desired features for a VoIP speech quality assessment solution are:

1. Automatic: It should provide measurement of speech quality online while the network is running.
2. Non-intrusive: It should be able to provide measurement of the speech quality depending on the received speech signal or network parameters without the need for the original signal.
3. Accurate: It should provide accurate measurement of speech quality to reflect how the quality is perceived by the end-user.
4. With the changing world, it should be applicable to new and emerging applications and networking conditions. As such it should avoid the subjectivity in estimating parameters. The E-model (section 4.2) for example depends on subjective tests to estimate packet loss parameters which hinders its applicability for new networking conditions.

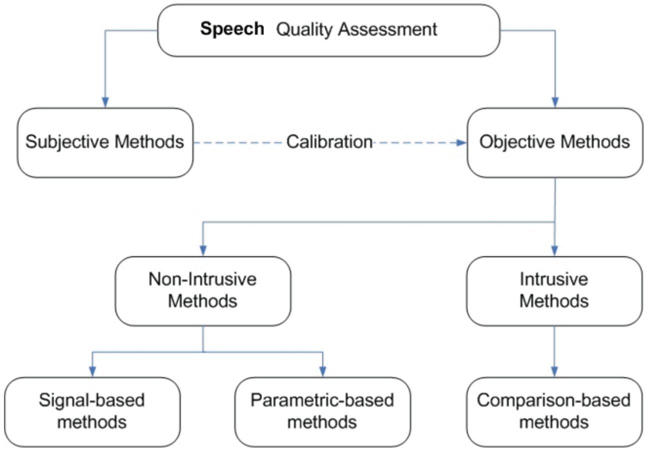

Fig. 1. Overview of VoIP measurement methods (Sun, 2004)

Based on the above requirements and from the previous discussion, the subjective and intrusive solutions that will be discussed in sections 3 and 4.1 respectively cannot be used for such task as they are manual and intrusive, respectively. The non-intrusive objective solutions that will be discussed in section 4.2 are candidates for such task. The most famous and widely used non-intrusive subjective solutions for measuring the speech quality are P.563 (ITU-T, 2004) and the E-model (ITU-T, 2009).

3. Subjective assessment of quality

The most widely used subjective quality assessment methodology is opinion rating defined in ITU-T Recommendation P.800 in which a panel of users (test subjects) perform the subjective tests of voice quality and give their opinions on the quality (ITU-T, 1996b). Subjective tests could be conversational or listening-only tests. In conversational test, two subject share a conversation 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 is calculated. In listening tests, one subject is listening to pre-recorded sentences (ITU-T, 1996b).

To conduct a subjective experiment according to the ITU-T Recommendation P.800, strict lab conditions should be in place. 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 case of 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). The recording system must be of high quality and recorded voice signals must consist of simple, meaningful, short phrases, taken from newspapers or non-technical lectures, randomly 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 (ITU-T, 1996b).

Recommendation P.800 also specifies 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. In case of listening-test they have never heard the same sentence lists before (ITU-T, 1996b).

In opinion rating methodology the performance of the system is rated either directly (Absolute Category Rating, ACR) or relative to the subjective quality of a reference system as in (Degradation Category Rating, DCR), or Comparison Category Rating (CCR) (ITU-T, 1996b; Takahashi, 2004; Takahashi et al., 2004).

The most common metric in opinion rating is Mean Opinion Score (MOS) which is an ACR metric with five-point scale: (5) Excellent, (4) Good, (3) Fair, (2) Poor, (1) Bad (ITU-T, 1996b). MOS is internationally accepted metric as it provides direct link to the quality as perceived by the user. A MOS value is obtained as an arithmetic mean for a collection of MOS scores (opinions) for a set of subjects. When the subjective test is listening-only, the results are in terms of listening subjective quality; i.e. MOS - Listening Quality Subjective or MOS_{LQS}. When the subjective test is conversational, the results are in terms of conversational subjective quality; i.e. MOS - Conversational Quality Subjective or MOS_{CQS} (ITU-T, 1996b; 2006). Although the overall quality of VoIP must be discussed in term of conversational quality, listening quality assessment is also quite helpful in analysing the effect of individual quality factors such as distortion due to speech coding and packet loss.

In DCR test two samples (A and B) are present: A represents the reference sample with the reference quality, while B represents the degraded sample. The subjects are instructed to acoustically compare the two samples and rate the degradation of the B sample in relation to the A sample according to the following five-point degradation category scale: degradation is (5) inaudible, (4) audible but not annoying, (3) slightly annoying, (2) annoying, and (1) very annoying. The samples must be composed of two periods, separated by silence (for example 0.5 seconds), firstly sample A then sample B.

The results (opinions) are averaged as Degraded MOS (DMOS). Each configuration is evaluated by means of judgements on speech samples from at least four talkers. DCR test affords higher sensitivity and used with high-quality voice samples, this is especially useful when the impairment is small and a sensitive measure of the impairment is required as ACR is inappropriate to discover quality variations as it tends to lead to low sensitivity in distinguishing among good quality circuits (ITU-T, 1996b; Takahashi et al., 2004).

The CCR method is similar to the DCR method as subjects are presented with a pair of speech samples (A and B) on each trial. In the DCR procedure, a reference sample is presented first sample (A) followed by the degraded sample (B). In the DCR method, listeners always rate the amount by which sample B is degraded relative to sample A. 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 following scale: (3) Much Better, (2) Better, (1) Slightly Better, (0) About the Same, (-1) Slightly Worse, (-2) Worse, and (-3) Much Worse (ITU-T, 1996b). 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 is represented as Comparison MOS (CMOS).

Results of MOS scores should be dealt with care as results may vary depending on the speaker, hardware platform, listening groups and test data and slight variation between different subjective tests should be expected although the above rigid conditions should guarantee

minimisation of such cases.
Although opinion rating methods are the most famous subjective quality assessment methodology, but other methods have also been proposed. Diagnostic Rhyme Test (DRT) is an intelligibility measure where the subject task is to recognise one of two possible words in a set of rhyming pairs (e.g. meat-beat). Diagnostic Acceptability Measure (DAM) scores are based on results of test methods evaluating the quality of a communication system based on the acceptability of speech as perceived by a trained normative listener (Spanias, 1994). Li (2004) proposed the use of intelligibility index as an additional parameter that can be used along with the commonly used MOS score. Opinion rating methods are still the most famous and widely used method.
Although subjective quality measurement is the most accurate and reliable assessment method to measure the quality as it reflects the user's perceptions of service quality, but there are few problems associated with subjective tests. It is apparent from the strict conditions associated with opinion rating methods as mentioned above that the inherent problems in subjective MOS measurement are that it is: time-consuming, expensive, lacks repeatability , and inapplicable for monitoring live voice traffic as commonly needed for VoIP applications. This has made objective methods very attractive to estimate the subjective quality for meeting the demand for voice quality measurement in communication networks to avoid the limitations of the subjective tests.

4. Objective assessment of quality

Objective speech quality assessment simulates the opinions of human testers algorithmically or using computational models to automatically evaluate the transmitted speech quality over IP networks to replace the human subjects, where the aim is to predict MOS values that are as close as possible to the rating obtained from subjective test and to avoid the limitations of subjective assessment methods. However, as subjective methods are the most accurate and reliable methods for measuring speech quality, they are used to calibrate objective methods. Therefore the accuracy, effectiveness and performance evaluation of objective methods are determined by their correlation with the subjective MOS scores.
Objective assessment of speech quality is based on objective metrics of speech signal or properties of the carrier network. Objective quality assessment methodologies can be categorised into two groups: Intrusive speech-layer models and Non-Intrusive models (Signal-based and parametric-based). Figure 2 shows the three main types of objective measurement.

4.1 Intrusive objective assessment of quality

Intrusive measures, often referred to as input-to-output measures or comparison-based methods, base their quality measurement on comparing the original (clean or input) speech signal with the degraded (distorted or output) speech signal as reconstructed by the decoder at the receiver side, this is shown in Figure 2 (a). Intrusive objective assessment of speech quality or speech-layer objective models are full-reference methods for measuring the quality. They provide an accurate method for measuring speech quality as they require the original or reference speech signal as input and produce measurement of listening MOS by comparing the post-transmitted signal with the original one (double-ended) using a distance measure, based on this comparison the quality of the degraded signal is measured in comparison with the quality of the original signal. However, such methods are inapplicable in monitoring live traffic because it is difficult or impossible to obtain actual speech samples as the