Mean Opinion Score Measurements Based on E-Model During a VoIP call

A Single Comparison On-line and Off-line

Edgard Silva, Leandro Galvão, Edjair Mota
Institute of Computing – ICOMP
Federal University of Amazonas – UFAM
Manaus – Brazil
e-mail: {edgard, galvao, edjair}@icomp.ufam.edu.br

Yuzo Iano
Department of Communications – DECOM
University of Campinas – UNICAMP
Campinas – Brazil
e-mail: yuzo@decom.fee.unicamp.br

***Abstract*— This paper presents the Mean Opinion Score (MOS) measurements results of an input stage of an adaptive speech based on Quality of Experience (QoE) control during a Voice Over Internet Protocol (VoIP) call. QoE is periodically determined by means of a modified version of E-model and recorded in a log file. At the end, we compared the results to POLQA (Perceptual Objective Listening Quality Assessment), and prove that continuously measurements are correct and can be used in the future real-time QoE controller to be build.**

Keywords-Quality of Experience (QoE); Voice over Internet Protocol (VoIP); Adaptive control system; MOS; speech quality; Codec Switching.

I. INTRODUCTION

Present in our personal and professional activities, speech signals transmissions over computer networks (VoIP - Voice over Internet Protocol) have gained wide acceptance by general Internet users. In a VoIP application, the main goal is to transfer voice signals over the IP network. In order to achieve this, the voice is digitized and packetized at the sender, and next, the result of the packaging is transmitted over the IP network to the receiver. In the receiver, the voice signals are unpacked from the received packets and then are decoded. After this, the results are played out to the listener [1]. However, VoIP is subject to several degradations, both at the application layer or the network layer, such as compression of the encoder, the delay end-to-end packet loss, jitter and bandwidth levels. Thus, to keep and attract new users, the quality of the provision of VoIP services need to be measured and optimized to ensure user satisfaction [2].

Quality provision assurance is one of the problems not solved yet, although VoIP enjoys the progress made in the last two decades [3]. In recent years, researchers developed Quality of Service (QoS) Control mechanisms to improve the use of network resources and user's terminal to minimize speech quality degradation. Some of these mechanisms seek to adapt the voice stream or other VoIP parameters, according to significant changes in network end users preferences, or providing requirements of service providers.

The speech quality can be measured subjectively or objectively. Subjective evaluation involves 12-24 participants individually listening to an audio stream of several seconds and classifying the audio quality on a scale of 1 (poor) to 5 (excellent). These ratings form a single Mean Opinion Score (MOS), as specified in ITU-T Recommendation P.800 [4]. This evaluation is costly and time-consuming and it cannot be done in real-time if one considers these characteristics. Thus, various techniques have been proposed to estimate objectively MOS (without human perception), such as POLQA [5] and E-model [6].

POLQA [5] is a perceptual technique that compares off-line two signals to generate the MOS: a reference signal (for example, captured at the sender) and the degraded signal (for example, captured at the receiver). The requirement makes the approach unsuitable for live call monitoring. On the other hand, the technique of E-model specified in ITU-T Rec. G.107 [6] is a noninvasive method that uses the network metrics monitored locally and the equipment impairment factor to estimate the quality of the call, so it can be used for monitoring live calls. A problem with the E-model is that only the ITU-T provides the equipment impairment factor specified in ITU-T Rec. G.113 [7]. For a range of other commonly used codecs not specified by the ITU-T Rec. G.113 [7], the equipment impairment factor is not provided.

Adaptive control systems in general respond to changes in their internal state or external environment with guidance of an underlying control system. VoIP systems are likely to need dynamic adaptation to deal with the complex dilemma between voice quality and impairment. This is necessary due to the nature of decentralized control of IP networks and the stochastic nature of data packets delivery. While existing solutions for QoS control of VoIP show some performance improvement and have feedback, they do not provide explicit focus on the control loop [8]. Measurable QoS relates directly to the state of the network, while QoE relates directly the quality level perception that users have. This perception will play a key role in the decision of making a VoIP application success or failure [9].

In this paper, our intention is to obtain measurements in real-time, of QoE of a VoIP transmission. In the future, our intention is be to develop a robust controller for a codec in command of a bidirectional audio stream.

The outline of the paper is as follows. The next section briefly surveys the issue concerning the adaptation during a VoIP call. Section III presents the methodology we followed to get measures of the metrics of interest. This section also presents details about the scenarios for testing and the measurements results while Section IV discusses the findings. A brief concluding section presents our ongoing research.

 ISBN: 978-1-61208-411-4

II. Background

The studies conducted by Karapantazis [3], Manousos [10], Costa [11], Qiao [12], Myakotnykh [13] and Viana [14] show that adaptation during a VoIP call (e.g., codec, packing, redundancy) can significantly improve the speech quality. However, as pointed out by Carvalho [15], these works often focus on parameter setting and little or nothing are based on the advances of research in adaptive systems.

Aktas et al. [16] compares the speech quality of a set of standard VoIP codecs given different network conditions and propose an adaptive end-to-end based codec switching scheme based on packet loss, jitter, and available bandwidth as the factors that define the current network condition.

Costa and Nunes [17] describe an adaptive codec switching technique that starts to monitor and analyze the quality of the voice, changing to a lower or higher codec rate according to predefined threshold values for each codec.

Haytham Assem [1] presents and evaluates an algorithm that performs the selection of the most appropriate audio codec given prevailing conditions on the network path between the endpoints of a voice call.

Bringing together contributions from the fields of VoIP and adaptive control systems, this project proposes a well-founded solution to the problem of real-time control of the quality of speech on VoIP calls and introduces the method of measurement and results the input signal of the proposed controller. For this, it is necessary to define what to measure and how to measure these parameters of the controller input.

III. Methodology

The method to be employed in this work is to improve the diagnostic module developed in [15]. The diagnostic module is composed of agents of monitoring and analysis. It uses the RTCP XR reports (RTP Control Protocol Extended Reports) [18]. These reports carry information about the instantaneous quality of a VoIP call, such as loss, delay and ambient noise, codec impairment, all this used in a function that calculates the E-model for instant call quality.

As a testbed, we used 64-bit machines with Ubuntu 10.04 operating system (newer versions proved inefficient for preliminary tests and incompatible with the Intel IPP 7.7.1, a codec library used in the experiments) to implement the PJSIP 1.10, a free library of multimedia communication and open code written in the C language that implements the protocols based on standards such as SIP, SDP, RTP, among others [19].

We had in the first instance, a machine (sender) directly connected to another machine (receiver) as proof of concept for local testing and controlled conditions.

At all stages of this experiment, we recorded logs files of transmission (sent file) and received (received file) files for analysis in Section VII. We transmitted the same file multiple times in different test scenarios and test multiple files with different contents in order to compare the measurements of the systems (internal measurements) and the result of the global measurements (POLQA).

We have two kinds of moments of measurements: during a VoIP call (punctual) and the cumulative average at the end of a VoIP call (global).

This paper presents the results of measurements performed during a call, the procedure and the results graphically.

During a VoIP call, there are two applications for point measurements, which are results of the measurements provided by the E-model routine implemented in PJSIP via RTCP-XR:

- Communication Performance: the instantaneous value of the QoE during an analysis interval of a VoIP call.
- Control: (Specifically in this study, the control application is not enabled).

We have the E-model (Rec. ITU-T G.107 [6]) among the methods of timely measurement of quality of speech most used. Its measurement procedure consists of collecting parameters of voice stream, which serve as input to a set of equations that return as a result the R factor, whose value ranges from 0 (worst) to 100 (best) as a measure of quality speech evaluated. This result is mapping from R factor (0-100) to MOS (1-5) by (1).

$$MOS = \begin{cases} 1 & if\ R < 0 \\ 1 + 0.035R + 7\text{x}10^{-6}R\,(R-60)(100-R) & if\ 0 \leq R \leq 100 \\ 4.5 & if\ R > 100 \end{cases} \quad (1)$$

Typically, the result of E-model is transformed to a scale of MOS [4] as presented in Table I.

At the end of the VoIP call, we received a file (received file) for a given file transmitted (sent file) in a given situation, to be used for measuring the overall quality of the call during the analysis presented in Section IV, and a score between 1 and 5 is generated, including a confidence interval.

IV. Analysis of Results

Having two types of metrics (punctual/local and global/external) we had two stages of analysis.

Quality measurement based on QoE will be held along the VoIP call. The measurement data quality along the VoIP call will be determined by a series of calls, with the same uploaded file (sent file), with and without the controller application enabled. Then, the results will be compared each other. A graphic of the call quality over time will be generated (in seconds) (Q × t) for both situations, considering the uncertainty range for each measurement point in the various transmissions.

The signal received by the listener (received file) will be recorded and compared off-line with the transmitted signal (sent file) with a perceptual objective method of measuring quality of speech and should be as close as possible to the subjective quality scores obtained in subjective tests hearing.

TABLE I. MEAN OPINION SCORE VALUE [16]

MOS	Quality	Impairment
5	Excellent	Imperceptible
4	Good	Perceptible but not annoying
3	Fair	Slightly annoying
2	Poor	Annoying
1	Bad	Very annoying

This method uses the knowledge of the workings of the human auditory system to compare a reference signal with a degraded signal in order to compose a measure of voice distortion, the main representatives sit POLQA, also known as ITU-T Rec. P.863 [5], which includes a model for predicting voice quality through analysis of digital voice signal.

The data collected will be analyzed by statistical tools like Akaroa [20] and R [21], based on statistical hypothesis testing, longitudinal data analysis, among other techniques. In order to emulate the packet loss in our network scenario, we use the Netem (Network Emulation) [23].

In this work, voice files from the Open Speech Repository (OSR) were used [22]. Figures 1, 2 and 3 are related to the sent file, named osr_us_000_0010_8k.wav, a female voice file, 16-bit PCM, 8 kHz sample rate. The sender establishes a VoIP call to the receiver machine and sends this file. A total of 25 connections were established for the same file at the following network conditions controlled by Netem: 0%, 2.5%, 5%, 7.5% and 10% packet loss. In all cases, the G.711 codec was used. The log files generated by each transmission feed a Perl script that filters the instantaneous quality information. An R script determines and plots the average value of instantaneous measured signal quality. This average is compared to the result of the analysis with POLQA of received and transmitted audio files.

Instantaneous measurements of MOS quality were taken during the random interval of 0.5 to 1.5 seconds. As the packet loss increases, the instantaneous MOS quality has become more unstable, as presented in Figure 1. This occurs as the RTP packets of data travelling across a network fail to reach their destination, i.e., the receiver station.

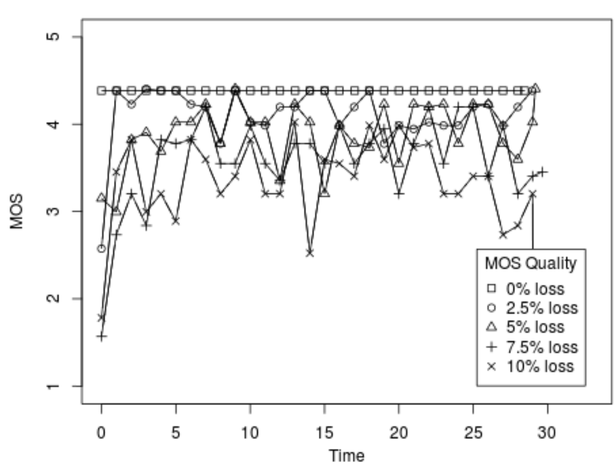

Figure 1. Instantaneous MOS Quality

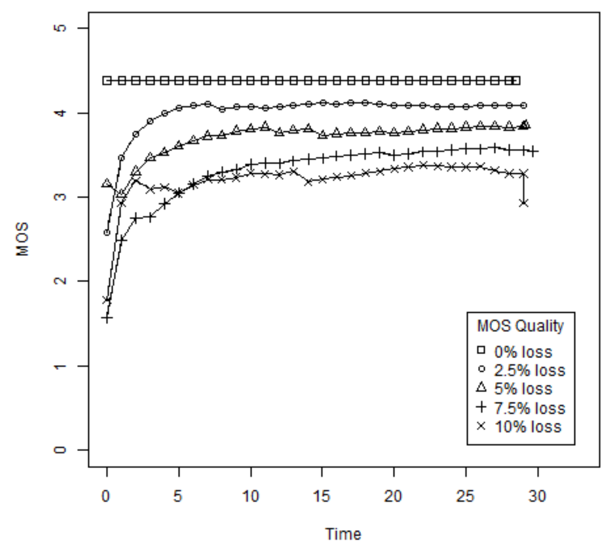

Figure 2. Average MOS Quality

Figures 1, 2 and 3 show the MOS in time considering some impairment. They show that the impairments result in MOS reduction.

As the value of MOS is calculated from the R factor and this is directly related to the network parameters, one being the loss of packets, by varying the value of the R factor, the instantaneous value of the MOS also varies. A variation on the R factor necessarily implies a variation of the MOS. Note that the best case for the MOS is when there is no packet loss.

The average value of the MOS Quality approaches to the value of the global MOS as time passes. In Figure 2, the MOS value for a given situation of packet loss gets worse as the loss increases.

In Figure 2, the average MOS Quality presents a little difference before 5 seconds between the values of packet loss of 5% to 10%. The initial values have wide oscillations that are minimized by the averaging as the time elapses.

As the packed loss increases, MOS Quality decreases, as shown in Figure 3.

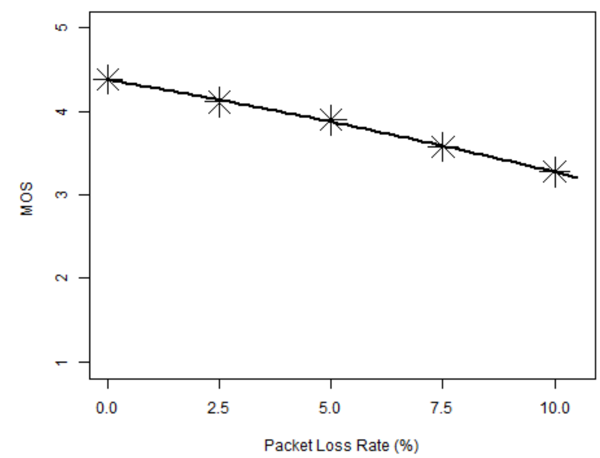

Figure 3. Average MOS versus packed loss rate - measured approximated values

V. CONCLUSIONS AND FUTURE WORK

This paper presented the results of measurements of input parameters of an adaptive control system of quality of speech based on QoE during a VoIP call.

The QoE varies along a voice transmission depending on factors like packet loss, jitter and bandwidth in addition to the equipment impairment factor. It is possible to measure these values and transform them into a single value (R factor). The R factor and the value of MOS Quality can be used as input variables in our adaptive control system. They vary over time and its measurement was easily implemented. This will lead our control system to make a decision to change or not the encoder and when this change will be made in order to impact as little as possible the communication. The current encoder itself represents one of the input parameters in the control system; all of that seeks to minimize the encoder changes in order to maintain the best quality experience for the users.

The system will choose a new codec according to the trend of monitored variables. Machine learning can be used as new encoders appear on the market. Three factors must be considered: a) when to switch the codecs, b) the codec used and the codec chosen to replace it and c) the reasons that led to the decision by the choice of the new codec.

The next steps of this work will be the extent of equipment impairment factor for no ITU codecs, such as Speexs, for example. Moreover, add to library encoders available, the Opus [8][24] wideband codec.

ACKNOWLEDGMENT

We would like to thank the School of Technology (EST) of the State University of Amazonas (UEA), the Foundation for Research Support of the State of Amazonas (FAPEAM), the Department of Communications of the Faculty of Electrical and Computer Engineering of the University of Campinas (UNICAMP), FAPESP, CAPES and CNPq for support in developing this work.

REFERENCES

[1] A. Haytham, M. Adel, D. Malone, B. Jennings, J. Dunne, and P. O'Sullivan, "A Generic Algorithm for Mid-Call Audio Codec Switching," in *IFIP/IEEE International Workshop in Quality of Experience Centric Management (QCMAN 2013)*, Ghent, Belgium, May 2013.

[2] D. Rodrigues, E. Cerqueira, and E. Monteiro, "QoE Assessment of VoIP in Next Generation Networks," in *Proceedings of the 12th IFIP/IEEE International Conference on Management of Multimedia and Mobile Networks and Services: Wired-Wireless Multimedia Networks and Services Management*, Venice, Italy, October, 2009.

[3] S. Karapantazis and F.-N. Pavlidou, "VoIP: A comprehensive survey on a promising technology," *Computer Networks,* vol. 53, 2009, pp. 2050–2090.

[4] ITU-T, "Rec. P.800: Methods for subjective determination of transmission quality," Geneva (Switzerland), 08/1996.

[5] ITU-T, "Rec. P.863: Perceptual objective listening quality assessment (POLQA)," Geneva (Switzerland), 01/2011.

[6] ITU-T, "Rec. G.107: The E-model, a computational model for use in transmission planning," Geneva (Switzerland), 07/2002.

[7] ITU-T, "Rec. G.113: Transmission impairments due to speech processing Appendix I: Provisional planning values for the equipment impairment factor Ie," Geneva (Switzerland), 10/2001.

[8] J.-M. Valin, K. Vos, and T. B. Terriberry, "Definition of the Opus Audio Codec, RFC6716," September 2012. [Online]. Available: http://www.rfc-base.org/txt/rfc-6716.txt. [retrieved: Feb. 2014].

[9] M.-D. Cano and F. Cerdan, "Subjective QoE analysis of VoIP applications in a wireless campus environment," *Telecommunication Systems,* vol. 49, 01 January 2012, pp. 5-15.

[10] M. Manousos, S. Apostolacos, I. Grammatikakis, D. Mexis, D. Kagklis, and E. Sykas, "Voice-quality monitoring and control for VoIP," *Internet Computing,* vol. 9, no. 4, July-Aug 2005, pp. 35-42.

[11] N. Costa and M. S. Nunes, "Dynamic Adaptation of Quality of Service for VoIP Communications," vol. 2, no. 2 and 3, 2009, pp. 155-166.

[12] Z. Qiao, L. Sun, N. Heilemann, and E. Ifeachor, "A new method for VoIP quality of service control use combined adaptive senter rate and priority marking," in *IEEE International Conference on Communications*, 2004.

[13] E. S. Myakotnykh and R. A. Thompson, "Adaptive Rate Voice over IP Quality Management Algorithm," *International Journal on Advances in Telecommunications,* vol. 2, 2009, pp. 98-110.

[14] B. d. A. Vianna, N. T. Moura, C. V. N. d. Albuquerque, V. E. F. Rebello, and C. Boeres, "adaMOS: Algoritmo MOS-Adaptativo para Fontes VoIP," in *WebMedia '06 Proceedings of the 12th Brazilian Symposium on Multimedia and the web*, New York, NY, USA, 2006.

[15] L. S. G. d. Carvalho, "Adaptive Management of Speech Quality Between VoIP Terminals," PhD thesis, Universidade Federal do Amazonas (UFAM), Manaus, 2011.

[16] I. Aktas, F. Schmidt, E. Weingärtner, C.-J. Schnelke, and K. Wehrle, "An Adaptive Codec Switching Scheme for SIP-Based VoIP," in *12th International Conference, NEW2AN 2012, and 5th Conference, ruSMART 2012*, St. Petersburg, Russia, Jan. 2012.

[17] N. Costa and M. S. Nunes, "Adaptive Quality of Service in Voice over IP Communications," *Proceedings of the 2009 Fifth International Conference on Networking and Services (ICNS 2009)*, Abril 2009, pp. 19-24.

[18] T. Friedman, R. Caceres, and A. D. Clark, "RTP Control Protocol Extended Reports (RTCP XR). Request for Comments (RFC) 3611," November 2003. [Online]. Available: https://tools.ietf.org/html/rfc3611.

[19] B. Prijono and P. Ismangil, "PJSIP," [Online]. Available: http://www.pjsip.org/. [retrieved: Feb. 2015].

[20] G. Ewing, K. Pawlikowski, and D. McNickle, "Akaroa-2: Exploiting network computing by distributing stochastic simulation," in *Proceedings of the 13th European Simulation Multi-conference*, Warsaw, Poland, 06/1999.

[21] R Core Team, "R: A Language and Environment for Statistical Computing," R Foundation for Statistical Computing, 2013. [Online]. Available: http://www.R-project.org. [retrieved: Feb. 2015].

[22] O. Project, "Open Speech Repository," [Online]. Available: http://www.voiptroubleshooter.com/open_speech/american.html. [retrieved: Feb. 2015].

[23] Linux Foundation, "Netem: Network Emulator," 19 November 2009. [Online]. Available: http://www.linuxfoundation.org/collaborate/workgroups/networking/netem. [retrieved: Feb. 2015].

[24] IETF, "Definition of the Opus Audio Codec, RFC6716," IETF, 09 2012. [Online]. Available: http://www.rfc-base.org/txt/rfc-6716.txt. [retrieved: Feb. 2015].

[25] H. Müller, M. Pezzè, and M. Shaw, "Visibility of control in adaptive systems," in *Proceeding ULSSIS '08 Proceedings of the 2nd international workshop on Ultra-large-scale software-intensive systems*, New York, NY, USA, 2008.