

Estimation of Media QoE without reference*

Péter Orosz, Zoltán Nagy

University of Debrecen
oroszp@unideb.hu, nagy@unideb.hu

Abstract

QoE analysis of media services over IP networks (e.g., VoIP, IPTV, OTT, etc.) is a complex task, since it is based on the subjective human opinion. The subjective assessment requires a large number of volunteers to periodically provide feedback about the perceived quality of a service. Accordingly, there is a requirement for an objective alternative to the inherently subjective, perception-based service quality assessment approaches. The pivotal point of the presented solution is that the actual service quality is calculated from the momentary value of the QoS metrics measured at packet-level and it is presented on the MOS scale. Using an estimation-based approach, we are working on an objective NR type QoE estimation mechanism that needs no reference content from the sender side. To determine an estimation function, we performed combined (subjective and objective) assessments to build a reference dataset of 3-tuples of MOS, jitter and loss values. Applying polynomial regression and investigating canonical correlation, we are searching a low-degree two-variable polynomial to hash objective QoS metrics (jitter and loss, respectively) to the subjective MOS score of the service quality.

Keywords: QoE estimation, QoS-QoE correlation, QoS metrics, subjective assessment, network measurement

1. Motivation

Nowadays a high percentage of the Internet traffic mix is real-time media service, such as VoIP, IPTV, OTT video, etc. The transmission of these media contents over the IP network can be performed effectively with network parameters in tight control. These network services require well-defined network resources, which is particularly critical for real-time media streams. The Internet Protocol provides

*The research was supported by the TÁMOP 4.2.2.C- 11/1/KONV-2012-0001 project. The project was implemented through the New Széchenyi Plan, co-financed by the European Social Fund.

best-effort service, which enables the routers to process and forward network traffic as soon as possible, but cannot guarantee the delivery in time, nor the arrival of the package, so that is not optimized for the transmission of time sensitive traffic. The important factors for the providers are the fulfillment of Service-level agreements (SLA), and the end user satisfaction. Therefore the objective measurement of the network metrics is performed occasionally or continuously. But objective packet-level measurement alone is not sufficient for evaluating the quality of media services, because the objective QoS is not always correlate with subscribers' perceived quality (subjective QoE). The perceptual quality can be monitored in subjective way. There are scales for QoE tests like MOS (Mean Opinion Score) which require to give each sentence a rating between one and five. However what's the problem with the subjective methods? The subjective assessment requires a large number of volunteers and it just periodically provide feedback about the perceived quality of a service. Because of these difficulties, there can be a requirement for an alternative objective method to measure media QoE with analyze the media performance metrics. In case of voice coding, there is a single frame-type. Within a stream one packet contains one audio frame, and the bitrate is constant. Bitrate of IPTV is also constant, but a video stream contains multiple frame-types (i.e., I-frame, P-frame, and B-frame), so the challenge is to identify the affected frame-type in the case of packet loss event, and integrating this information into the objective QoE estimator.

2. Objective

Our objective is designing and prototyping a method for real-time service quality evaluation for VoIP and IPTV services. Using this estimation-based approach, we would like to create an objective NR type QoE estimation mechanism that needs no reference content from the sender side. In the first phase of our work, we need to design a laboratory test-bed for emulating network perturbation, and performing media transmission in many QoS scenarios with packet-level QoS measurement. After the subjective evaluation of the results with a large number of volunteers, we would like to find a generic estimator function to map measured packet-level metrics to the subjective MOS scale. This paper presents our work in progress and shows some preliminary results of the voice quality assessment.

3. Related work

A Raake provided a basic guide to assessing the speech quality of VoIP networks [2]. Anssi Ramö and Henri Toukoma made subjective perception tests with Opus MDCT and LP modes, and they compared them with 3GPP AMR, AMR-WB and ITU-T G.718B [3]. They demonstrated that the Opus codec is a high quality alternative for the mentioned codecs. The papers of C. Hoene *et al.* includes different QoE tests and compares the codec to Speex, iLBC, G.722.1, G.722.1C

and AMR [4][5][6]. They showed that at lower rates the Opus codec performs better than most of the legacy codecs, but both AMR-NB and AMR-WB still outperform it. Jean-Marc Valin et al. minimized the impact of coding artifacts by improving the Opus encoder [7].

For standardized evaluation of speech quality, the International Telecommunication Union (ITU) has its own recommendation. After they developed PEAQ, PSQM, PESQ and PESQ-WB algorithms, the POLQA (ITU-T P.863) is able to perform speech assessment using the MOS metrics for sampling rates up to 14 kHz [8]. Cherif W. *et al.* created a perceived speech quality prediction method, called A_PSQLA [9]. This non-intrusive method relies on Random Neural Network (RNN) approach. Fei L. *et al.* suggest a QoE-based carrier scheduling scheme for the advancement of long-term evolution (LTE) networks [10]. In their investigation packet delay and bandwidth are the essential factors affecting QoE. Jelassi S. *et al.* summarized and analyzed objective QoE assessment methods [11]. We made an extensive survey on objective voice quality evaluation methods with subjective approaches. In our previous paper we examined the correlation of objectively measured QoS metrics (packet loss and jitter) with subjectively evaluated QoE values in the case of VoIP transmission [1]. For transmitting human voice streams in emulated WAN environment we used pre-defined QoS parameters. After the packet-level QoS measurements, the output audio files were evaluated by a number of volunteers. Based on these results we started to work on an objective QoE estimator function.

4. Methodology

4.1. Rnetprobe QoS module

Our QoE estimation method will be dedicated as a feature of a hardware accelerated service monitoring system that provides:

- Lossless packet-level capturing
- High precision timestamping
- Demultiplexing RTP flows
- Extracting codec related metadata
- Real-time measurement of network metrics
 - Packet inter-arrivals
 - Packet delay variation (jitter)
 - Packet reordering
 - Packet loss

4.2. Correlation analysis for voice QoS and QoE

The measurement was performed with controlled network perturbation in a laboratory configuration to investigate error-tolerance of VoIP voice transmission. We used the Netem Linux kernel module, which was configured symmetrically on both directly connected interfaces to emulate a WAN connection and produced various network anomalies that affect QoS (i.e., packet loss and variation of network delay) [12]. Table 1 contains the emulated transmission parameters for each case.

| Netem parameters | Set of values |
|---|--|
| jitter | 1 ms, 2 ms, 3 ms, ..., 20 ms |
| packet loss | 1%, 2%, 3%, ..., 40% |
| Combination of jitter packet loss | 1 ms, 2 ms, 3 ms, ..., 10 ms 1%, 2%, 3%, ..., 10% |

Table 1: Pre-defined QoS values for network

After the measurements, we got 160 audio clips with iterated perturbation values. With zero jitter and packet loss values an initial measurement were run to get a reference audio clip. The audio files, what were one minute long, got subjective QoE values by volunteers in a calm environment. To minimize the monotony each person has done the tests in her own tempo. Each test was started with listening of the reference audio clip. The QoE analysis is based on the average of the QoE ratings. The combined measurements determined a surface, therefore we have performed a deeper investigation.

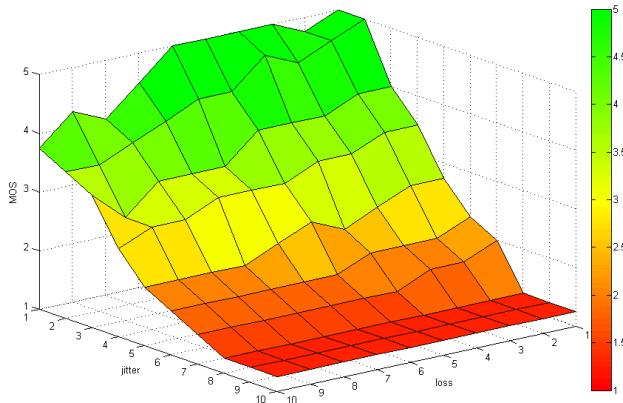


Figure 1: MOS values under mixed network conditions

Rated packet loss, jitter and MOS values are presented in a 3-axis graph on Figure 1. With using polynomial regression we determined a two-variable polyno-

mial that generates a fitted surface. This can be calculated using class of functions (4.1).

$$F = \{p_n(x) = a_0 + a_1x + \dots + a_nx^n\} \tag{4.1}$$

A matched polynomial is the result class of polynomials of (4.2).

$$M(\eta - f^*(\xi))^2 = \min_{\forall f \in F} M(\eta - f(\xi))^2 \tag{4.2}$$

Coefficients are calculated by solving the linear simultaneous equations in (4.3).

$$\begin{pmatrix} 1 & M\xi & \dots & M\xi^n \\ M\xi & M\xi^2 & \dots & M\xi^{n+1} \\ \vdots & \vdots & \ddots & \vdots \\ M\xi^i & M\xi^{i+1} & \dots & M\xi^{i+n} \\ \vdots & \vdots & \ddots & \vdots \\ M\xi^n & M\xi^{n+1} & \dots & M\xi^{2n} \end{pmatrix} * \begin{pmatrix} a_0 \\ a_1 \\ \vdots \\ a_i \\ \vdots \\ a_n \end{pmatrix} = \begin{pmatrix} M\eta \\ M\eta\xi \\ \vdots \\ M\eta\xi^i \\ \vdots \\ M\eta\xi^n \end{pmatrix} \tag{4.3}$$

With *poly{ij}* Matlab function, we performed polynomial regression. By changing the degree of input parameters the surface model can be tuned. For example, *poly{21}* model does the polynomial regression using second degree for the first variable and first degree for the second one (4.4).

$$Z = p_{00} + p_{10}x + p_{01}y + p_{20}x^2 + p_{11}xy \tag{4.4}$$

Sum of Squares due to Error (SSE) determines the standard deviation between a set of points and the matched surface and is calculated using (4.5).

$$SSE = \sum_{i=1}^n (y_i - \hat{y}_i)^2 \tag{4.5}$$

The surface provides lower match to the specified set of points if SSE value falls more far from zero.

4.2.1. Matched surfaces – poly11

Figure 2 plots the matched surface for this purely linear model, where the surface is determined by (4.6) and its coefficients.

$$f(x, y) = p_{00} + p_{10}x + p_{01}y \tag{4.6}$$

$$\begin{aligned} p_{00} &= 5.151 \\ p_{10} &= -0.07513 \\ p_{01} &= 0.4111 \end{aligned}$$

Value of goodness of fit with SSE is 16.24, and thus the MOS scores fall far from the surface shown on Figure 2.

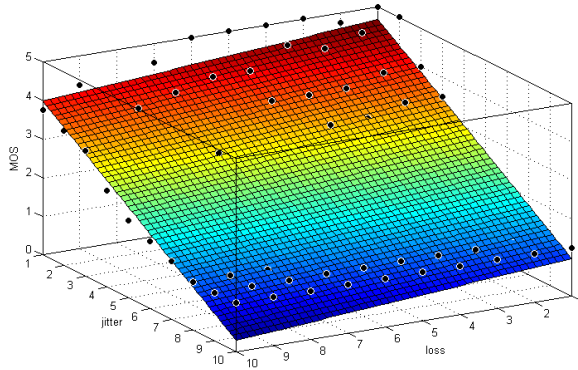


Figure 2: The surface determined by (4.6) and MOS values from the subjective evaluation

4.2.2. Matched surfaces – poly12

If we use a polynomial with second degree jitter (y) variable, the describing function will be (4.7).

$$f(x, y) = p_{00} + p_{10}x + p_{01}y + p_{11}xy + p_{02}y^2 \quad (4.7)$$

$$\begin{aligned} p_{00} &= 7.067 & p_{10} &= -0.2382 \\ p_{01} &= -1.07 & p_{02} &= 0.04787 \\ p_{11} &= 0.02217 \end{aligned}$$

Goodness of fit using SSE is 2.336. Introducing a second degree jitter variable resulted in a significant improvement in fitting as presented on Figure 3.

4.2.3. More matched surfaces

We calculated a number of matched surfaces of higher degree functions. Figure 4 and Table 2 show matching efficiency using SSE goodness of fit depending on the degree of the variables. In our arrangement, SSE values get closer to zero by increasing the degree of jitter (y), but by increasing the degree of loss (x) does not result in better fit.

5. Conclusion

Applying polynomial regression we searched a low-degree two-variable polynomial to hash objective QoS metrics (jitter and loss, respectively) to the subjective MOS score of the service quality. Our goal is dual, i.e., a good fitting surface and a fast algorithm. By increasing the degree of jitter (y) and loss (loss) values resulted in

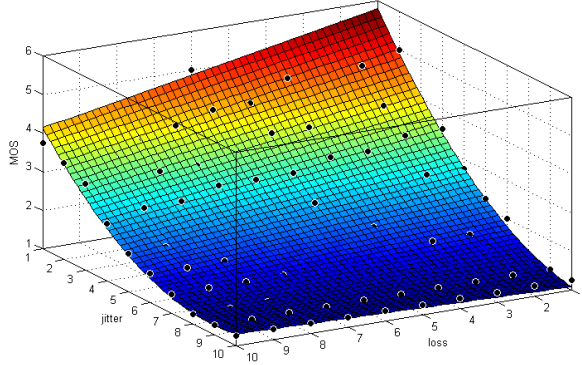


Figure 3: The surface determined by (4.7) and MOS values from the subjective evaluation

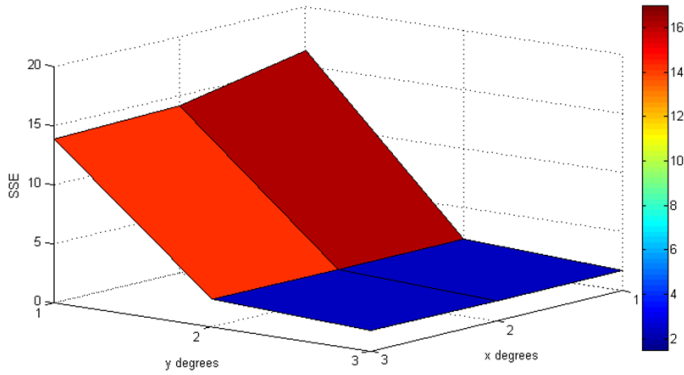


Figure 4: SSE goodness of fit for surface matching using polynomial regression

more accurate MOS estimation, but also adds significant amount of complexity to the calculation. Since we are going to optimize the method to hardware-based implementation, keeping the complexity of the algorithm low is critical. Introducing a second degree jitter variable resulted in a significant improvement in fitting. It is reasonable to ask: what is the optimal complexity-error trade-off for a real-time NR estimation method? In the next stage of our investigation, we are going to focus on the evaluation of the polynomials.

| Degree of x | Degree of y | SSE | No of coefficients |
|-------------|-------------|---------|--------------------|
| 1 | 3 | 1.6905 | 7 |
| 2 | 3 | 1.7138 | 9 |
| 3 | 3 | 1.7445 | 10 |
| 2 | 2 | 2.2895 | 6 |
| 3 | 2 | 2.3346 | 9 |
| 1 | 2 | 2.3363 | 5 |
| 3 | 1 | 13.8641 | 7 |
| 2 | 1 | 14.1052 | 5 |
| 1 | 1 | 16.2389 | 3 |

Table 2: SSE GOF values for different degrees of the variables

References

- [1] Orosz, Péter, Nagy, Zoltán, "Performance Analysis of the Opus Codec in VoIP Environment Using QoE Evaluation." ICSNC 2013, The Eighth International Conference on Systems and Networks Communications. 2013.
- [2] Raake, Alexander. Speech Quality of VoIP: Assessment and Prediction. John Wiley & Sons, 2007.
- [3] A. Ramö and H., "Voice Quality Characterization of IETF Opus Codec" INTERSPEECH, pp. 2541-2544. ISCA, 2011
- [4] C. Hoene, Ed., JM. Valin, K. Vos, and J. Skoglund, "Summary of Opus listening test results draft-valin-codec-results-01", May, 2012
- [5] C. Hoene, Ed., JM. Valin, K. Vos, and J. Skoglund, "Summary of Opus listening test results draft-valin-codec-results-02", May, 2012
- [6] C. Hoene, Ed., JM. Valin, K. Vos, and J. Skoglund, "Summary of Opus listening test results draft-valin-codec-results-03", November, 2013
- [7] JM. Valin, K. Vos, and J. Skoglund, "Summary of Opus listening test results draft-valin-codec-results-00" June, 2011
- [8] ITU-T P.863: Perceptual objective listening quality assessment, January 2011
- [9] Cherif, Wael, et al. "A_PSQL: PESQ-like non-intrusive tool for QoE prediction in VoIP services." Communications (ICC), 2012 IEEE International Conference on. IEEE, 2012.
- [10] Liu, Fei, et al. "A novel QoE-based carrier scheduling scheme in LTE-Advanced networks with multi-service." Vehicular Technology Conference (VTC Fall), 2012 IEEE. IEEE, 2012.
- [11] Jelassi, Sofiene, et al. "Quality of experience of VoIP service: A survey of assessment approaches and open issues." Communications Surveys & Tutorials, IEEE 14.2 (2012): 491-513.
- [12] netem: Linux Networking Emulator, <http://www.linuxfoundation.org/collaborate/workgroups/networking/netem>, [retrieved: August, 2013]