

An Approach for Instrumental Evaluation of Time-varying Speech Quality in an IP-testbed

Florian Hammer

Telecommunications Research Center Vienna (ftw.), A-1220 Vienna, Email: hammer@ftw.at

Introduction

Voice over IP (VoIP) is about to become a reliable communication service. In order to be able to compete with traditional circuit-switched telephony services, VoIP service providers need to offer an acceptable level of Quality of Service (QoS). The time-varying nature of IP-networks due to non-stationary cross-traffic causes packetized voice-traffic to be exposed to changing conditions with regard to packet loss, and delay variation. Such a network behavior eventually ends up in a time-varying degradation of speech quality. In this paper, we present an instrumental method for perceptual speech quality measurement that allows a comprehensive analysis of packet loss traces collected from different network scenarios.

Perceptual Quality Evaluation

In general, perceptual speech quality is determined in *subjective tests*. Therefore, test persons rate the quality of either speech samples they are listening to or conversations they have with another test person in a reproducible environment. Subjective speech quality testing methods are standardized within ITU-T Rec. P.800 [1]. The major disadvantage of subjective testing is that it is time-consuming and costly.

To reduce this effort and get estimates for the perceptual speech quality more efficiently, *instrumental assessment* methods are being developed. Throughout the last decade, intrusive methods have been developed which are based on the comparison of a degraded signal with its original in the perceptual domain. The ITU-T has standardized such a method as Rec. P.862, “Perceptual Speech Quality Evaluation” (PESQ, [2]). PESQ allows for an efficient evaluation of various degradations like transmission channel errors or packet losses and can be applied for speech codec evaluation and selection, and testing of live- or emulated networks. Thus, PESQ is a useful tool for evaluating the quality of a large amount of speech samples.

Trace-based Quality Assessment Paradigm

Our aim is to study the perceptual speech quality resulting from arbitrary packet loss traces. Loss traces are lists containing information about whether a voice packet has been dropped or not. Traces can result from the real Internet, Testbed measurements or simulations. Before the speech data is decoded, we obtain a so-called “Effective Loss Pattern” (ELP) at the receiver as shown in Figure 1. The ELP depends on the QoS of the network, and the QoS-related signal processing like jitter-buffering or po-

tential perceptual packet marking algorithms. Furthermore, the coding algorithm will influence the resulting quality which is determined using PESQ.

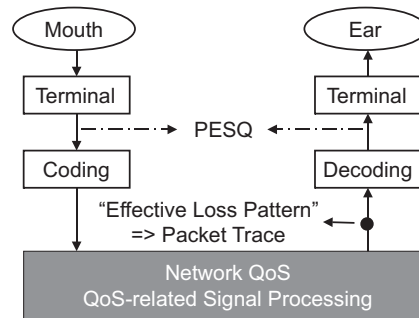


Figure 1: General measurement paradigm.

Evaluation Method

After obtaining the ELP of a certain network scenario, we figuratively slide a speech sample along the trace as depicted in Figure 2. Therefore, loss trace fragments are continuously extracted following a certain step-size and applied to the bitstream of the speech sample. If a packet is lost at a certain position within the fragment, the corresponding position within the bitstream is manipulated as to indicate the loss of a speech frame to the speech decoder.

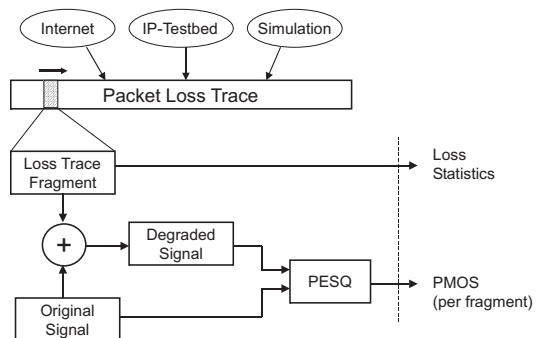


Figure 2: Packet trace evaluation method.

Testbed Measurements

As a first application of our method, we have evaluated packet loss traces resulting from measurements in our IP-testbed. The testbed structure is shown in Figure 3. A voice flow shares a 10 Mbps bottleneck link with cross-traffic resulting from different numbers of Web-users. For this work, we have chosen load situations of 150, 200, and 250 Web-users resulting in average packet loss rates (PLR) of 0.9%, 5.3%, and 13.7%, respectively, for a trace length of 60400 packets.

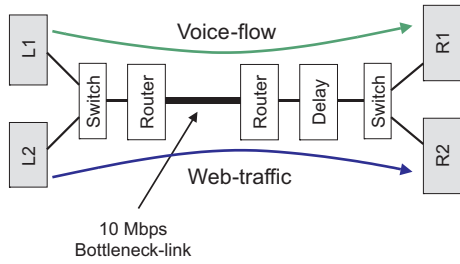


Figure 3: Structure of the IP-Testbed.

Results

The average speech quality for the different load situations using the G.729 codec [3] (two frames per packet = 20 ms of speech) is presented in Table 1. These values stem from fragments of traces containing 60400 packets which corresponds to a measurement duration of about 20 minutes. The trace fragment step-size was set to one packet, and the English female speech sample we have used was 400 packets (8 seconds) long (taken from the ITU-T speech database (Series P Supplement 23)). The

# Users	Avg. PLR [%]	PMOS
150	0.9	3.37 ± 0.13
200	5.3	2.95 ± 0.21
250	13.7	2.44 ± 0.20

Table 1: Average packet loss probabilities and average PMOS values for G.729 and iLBC speech coding.

graphs in Figure 4 present cut-outs of the mean PLR per trace fragment and the resulting time-varying speech quality for the various load situations. We can clearly observe the correlation between the packet loss rate and the perceptual quality. Moreover, this figure indicates the high variability of the perceptual quality over time. Figure 5 presents the distribution of the PMOS values resulting from the testbed traces and the distribution of the differences between two consecutive PMOS values. These graphs result from 9000 samples per condition which corresponds to a transmission time of 3 minutes. Note that the standard deviation of the PMOS difference of two consecutive samples (shifted by one packet) is 0.15 which reflects the influence of the loss location within a sample

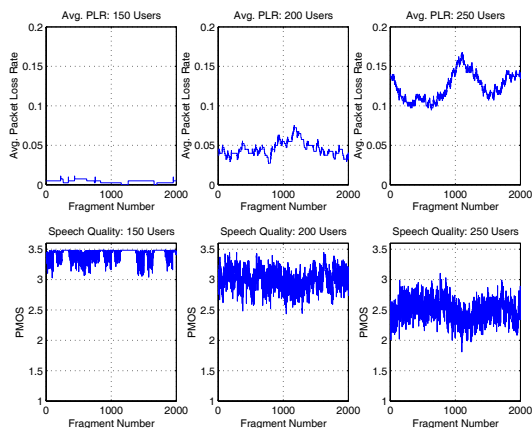


Figure 4: Time-varying speech quality for 150, 200, and 250 users.

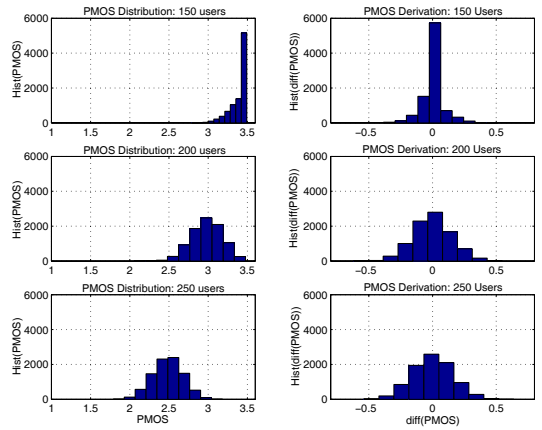


Figure 5: Distribution of PMOS values and their first deviation (step-size = 1 packet).

on the perceptual quality (cf. Sun et al. [4]). The PMOS difference varies between ± 0.5 for 150 users and between about ± 0.63 for 200 and 250 users.

Conclusions

We have presented a method for the instrumental evaluation of perceptual speech quality based on packet loss traces collected in an IP-testbed. By separating the quality assessment procedure into a measurement phase and an evaluation phase, we obtain more flexibility and efficiency compared to online measurements in which speech samples are periodically sent over a network [5]. We have used this method to accomplish a fine-grained analysis of the continuous speech quality at different cross-traffic load situations.

Further work will focus on the application of this method to the evaluation of perceptual packet marking algorithms, and future network QoS mechanisms. In addition, we are interested in the relation of this approach to continuous subjective testing of both microscopic and macroscopic packet loss (cf. Gros and Chateau [6]).

References

- [1] International Telecommunication Union. Methods for subjective determination of transmission quality. ITU-T Recommendation P.800, Aug. 1996.
- [2] International Telecommunication Union. Perceptual evaluation of speech quality (PESQ), an objective method for end-to-end speech quality assessment of narrow-band telephone networks and speech codecs, ITU-T Recommendation P.862, Feb. 2001.
- [3] International Telecommunication Union. Coding of speech at 8 kbit/s using conjugate-structure algebraic-code-excited linear-prediction (CS-ACELP), ITU-T Recommendation G.729, Mar. 1996.
- [4] L. Sun, G. Wade, B. M. Lines, and E. C. Ifeachor. Impact of packet loss location on perceived speech quality, IPTEL 2001, New York City, NY, Apr. 2001.
- [5] H. Furuya, S. Nomoto, H. Yamada, N. Fukumoto, and F. Sugaya. Experimental investigation of the relationship between IP network performances and speech quality of VoIP, Int. Conf. Telecommunications, vol. 1, pp. 543-552, Feb. 2003.
- [6] L. Gros and N. Chateau. Instantaneous and Overall Judgements for Time-Varying Speech Quality: Assessments and Relationships, Acta Acustica united with Acustica **87**, pp. 367-377, 2001.