Where Packet Traces Meet Speech Samples: An Instrumental Approach to Perceptual QoS Evaluation of VoIP

Florian Hammer, Peter Reichl, Thomas Ziegler
Telecommunications Research Center (ftw.) Vienna
Email: hammer@ftw.at, reichl@ftw.at, ziegler@ftw.at

***Abstract*— With the increasing deployment of real-time Internet services, evaluating the user perception of quality of service (QoS) has gained rapidly increasing importance. In the case of Voice over IP (VoIP), the standard approach of listening-only tests for subjectively assessing a limited number of speech samples, which are supposed to be representative for selected network conditions, does in no way reflect the huge variability of packet loss patterns that may originate from the underlying network. Performing tests by employing objective (instrumental) evaluation methods in a live testbed environment is usually extensive and does not deliver reproducible results, moreover the measurement granularity is bounded by the length of the test speech samples.**

In this paper, we propose a methodology that circumvents these limitations by employing arbitrary packet traces and successively matching the encoded speech sample with all possible trace fragments. This approach allows for continuous perceptual evaluation of VoIP traffic carried over various QoS-enabled transmission technologies. Results based on traces from testbed measurements reflecting different Web-like cross traffic situations for both the G.729 and iLBC codecs validate our approach and allow interesting insights into the dependence of perceived VoIP quality on underlying technological conditions.

I. Introduction

The provision of Quality-of-Service (QoS) enabled services in the future Internet is of vital importance for the successful deployment of real-time applications like Voice-over-IP (VoIP) with their sometimes strict requirements on QoS parameters like packet loss rate, end-to-end delay or jitter. On the other hand, it has been argued [1] that providing QoS is not only a matter of guaranteeing bounds for these parameters, but eventually boils down to the question of whether the end user is satisfied by the subjective quality she does actually perceive. Among the different approaches to evaluate this perceptual QoS, subjective testing (i.e. confronting test persons with different delivered qualities and asking them to explicitly express their opinion about them) plays a central role, but at the same time is a very expensive and resource-consuming activity. In the case of VoIP, a well-known example are listening-only tests: here, the test subjects have to assess a limited number of speech sample they are listening to, where the samples are supposed to be representative for a relevant range of network conditions as characterized e.g. by the overall packet loss rate while the variety of potential loss patterns that may originate from the underlying network is not taken into account.

Instrumental tests (sometimes also termed "objective tests") have been introduced to facilitate the evaluation of perceptual quality. Intrusive methods like PESQ [2] or TOSQA [3] are based on automatically comparing the original speech signal with its degraded version (after having passed through the network), whereas non-intrusive methods do not depend on having knowledge about the original signal. In both cases, bridging the gap between subjective and objective QoS is crucial for the validity of the evaluation [1].

In the case of VoIP, an important step towards solid instrumental tests consists of integrating a real IP transport network into the test scenario. In a straightforward version, speech samples are encoded and sent one after the other through a live testbed with realistic cross traffic. Then, the degraded signal is decoded and evaluated using standard algorithms. This approach allows to reflect realistic network conditions to a great detail, but is not suitable for delivering exactly reproducible conditions. Moreover, the granularity of the method is restricted to the length of the individual test samples and thus does not allow for a continuous perceptual QoS evaluation.

In this paper, we present a new methodology for measurement-based instrumental QoS evaluation of VoIP which allows an efficient comparison of different packet loss situations (e.g., different burst cases), and additionally facilitates time-varying speech quality measurement on a perceptual basis. The main idea is to start with a packet trace based on measurements (e.g., in a testbed), and choose arbitrary trace fragments whose length corresponds exactly to the speech sample under test. Matching speech sample and trace fragment allows to derive a realistic signal degradation which eventually is subject to a standard non-intrusive perceptual QoS evaluation. In contrast to the straightforward live measurement approach sketched

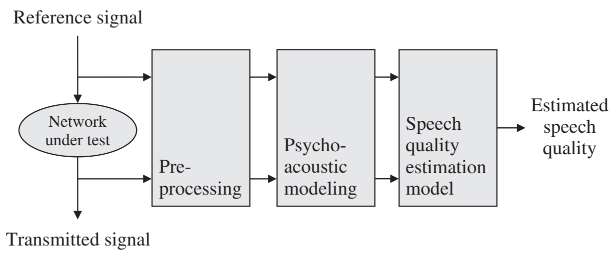

Fig. 1. Instrumental speech quality assessment: Structure

above which basically splits the packet trace into a series of subsequent fragments, we repeat this procedure for overlapping trace fragments (whose starting points differ, e.g., by only one packet). This allows a fine-granular, in fact continuous evaluation of time-varying QoS for different network load scenarios.

Summarizing, the main contributions of this papers are threefold: We introduce a new instrumental evaluation method for time-varying quality based on applying overlapping packet trace fragments on a suitable speech sample. Using this approach, we demonstrate that even for fixed packet loss rates, VoIP quality may vary significantly. Finally, our measurement results allow us some important conclusions for the case of bursty network traffic as well as a detailed comparison of the G.729 vs. the iLBC codec.

The remainder of the paper is structured as follows: after providing some material concerning background and related work in section 2, section 3 introduces the measurement method. The general methodology is presented and the employed IP testbed is described. Section 4 includes a variety of measurement results and their discussion. We provide a comparison of speech quality evaluation for consecutive vs. overlapping trace fragments and investigate burst performance, while comparing the two standardized codecs mentioned above. Section 5 closes the paper with summarizing conclusions and an outlook on further work.

II. Background and Related Work

VoIP QoS perceived by the user can generally be measured in two ways, i.e. subjectively or instrumentally. During subjective tests, test persons rate the quality of speech samples (listening-only tests) or the quality of the live-connection they have been using (conversational tests). These tests require lots of effort and may turn out to be quite expensive. Therefore, instrumental algorithms have been developed which are able to compare a degraded speech sample with its original in the perceptual domain and estimate the speech quality of the degraded sample.

One widely used example is the "Perceptual Evaluation of Speech Quality" (PESQ) algorithm which has been standardized by the International Telecommunication Union (ITU) as Recommendation P.862 [2]. The structure of PESQ is illustrated in Figure 1 and can be described as follows: After some preprocessing, such as level- and time-alignment, both the original and degraded speech signal are transformed into a psychoacoustic representation which models the properties of the human auditory system. In the perceptual domain, the signals are compared and a speech quality estimate is calculated which corresponds to a subjective mean opinion score (MOS) ranging from 1 (bad) to 4.5 (excellent) in most cases. We refer to the speech quality estimate as PMOS (PESQ-MOS). For a more detailed information on the psychoacoustic model of PESQ see [4]. PESQ has been validated for a wide variety of transmission impairments including packet loss and a variety of codecs like ITU Rec. G.729. Note that the speech source material is required to be at least 8 seconds long, and that PESQ reaches a correlation of up to 0.935 between subjective results and instrumental estimate [2].

For our investigations, we use English speech samples from the ITU-T speech database [5] which are 8 seconds long. The speech samples are coded using the ITU-T Rec. G.729 [6] speech codec which operates at 8 kbps and generates 10 ms frames with a size of 10 Bytes. In our study we put two speech frames into one IP/UDP/RTP packet such that eventually one such packet corresponds to 20 ms of speech. Therefore, 400 IP/UDP/RTP packets are required to send an 8 second speech sample. Furthermore, we compare the perceptual performance of the G.729 codec with the performance of the Internet Low Bitrate Codec (iLBC, [7]) which is about to become a standard within the Internet Engineering Task Force (IETF). The 20 ms mode of the iLBC works at a bitrate of 15.2 kbps.

It has been already mentioned in the introduction that the rate and structure of packet losses are of central importance for our investigations. Packet losses are usually modelled by the Gilbert model [8], [9] which is a 2-state Markov model as illustrated in Figure 2. p is the probability that a packet will be dropped given that the previous packet has been received, q is the probability that a packet will be received given that the previous packet has been lost. 1-q is termed the "conditional loss probability" (*CLP*), which serves as an indicator for the loss burstiness of the traffic:

$$CLP = 1 - q. \tag{1}$$

Moreover, the unconditional loss probability (*ULP*) represents the average packet loss rate and can be calculated [8] as:

$$ULP = \frac{p}{p+q} \tag{2}$$

So far, there is only a relatively small amount of related work available that directly touches our problem. Most notably, Furuya et al. [10] investigate the relationship between IP network performance and VoIP speech quality

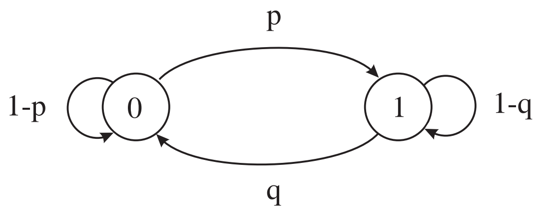

Fig. 2. 2-state Gilbert Model.

through extensive experiments in a testbed. The authors have evaluated the speech quality by cyclically sending an artificial speech signal [11] of 16 seconds over a bottleneck link that is shared with web traffic. The speech sample is coded using ITU-T G.711 Pulse Code Modulation and packetized into 20 ms packets. The speech sample sent and the degraded speech sample that has been received were recorded for quality evaluation using PESQ. In a second study, Conway [12] uses PESQ in a passive method for measuring and monitoring the speech quality in live VoIP calls. In this method, the actual packet loss pattern is extracted from the live stream and applied to an artificial speech signal [11], and the resulting speech quality is evaluated via PESQ (even if PESQ for artificial speech signals currently still lacks validation [2]). Beyond the usage of artificial speech signals, both approaches differ from our work also in terms of granularity, accuracy and efficiency.

III. Measurement Method

A. Paradigm

In contrast to previous work, our proposal is based on a clear separation between the network measurement process (including QoS-related signal processing) and the speech quality evaluation process. In the measurement phase, packet traces are collected from a particular network environment and situation. The speech quality resulting from the impairments subsumed within these traces is then evaluated off-line. This decoupling allows a substantial reduction of the evaluation effort (for example live recordings are no longer necessary) and eventually results in an efficient and accurate perceptual evaluation mechanism for VoIP.

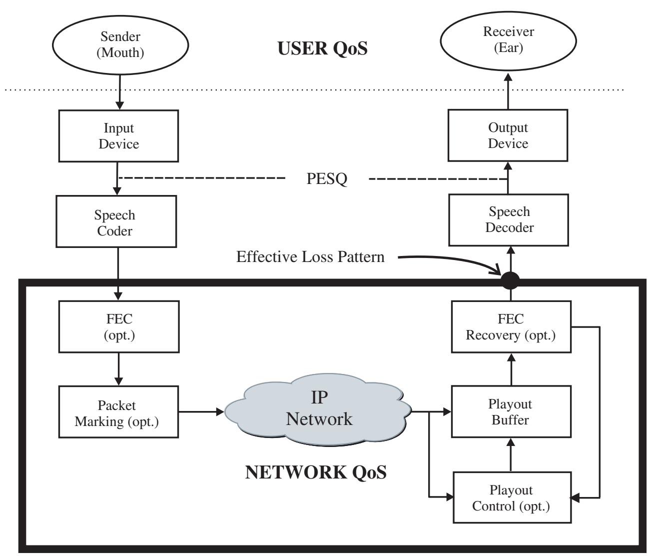

Fig. 3. Paradigm

Perceptual QoS of VoIP is influenced by a lot of factors. As already motivated in the introduction, we can divide QoS into network QoS and QoS perceived by the user (see Figure 3). Between the user and the network, the speech signals/packets need to be processed by algorithms such as speech coding, Forward Error Correction (FEC, optional), packet marking (optional), and jitter buffering which is usually adaptive/controlled. We refer to these algorithms as *QoS-related Signal Processing*.

The main assumption driving our work is based on the idea that the losses caused by the network, either due to congestion or due to bit errors in wireless links or due to jitter-buffering[1], minus the recovery of packets due to FEC result in a so-called “effective loss pattern”, which is finally forwarded to the speech decoder (see Figure 3). Obviously, any change in the loss pattern may influence the resulting perceptual speech quality.

Also the influence of the coding/loss concealment algorithms is subject to our evaluation method, whereas we do not include issues like acoustic echo cancellation, the quality of the input/output devices, and the level of background noise.

B. Evaluation Procedure

Based on the paradigm described above, we have developed a new methodology that, in principle, evaluates the perceptual speech quality resulting from arbitrary packet traces. These traces can be obtained from real Internet transmissions, testbed measurements, or simulations (e.g., using the ns-2 network simulator [13]).

The evaluation procedure works as follows: A speech sample is coded and the resulting bitstream is applied to a fragment of the packet trace under consideration, such that for each packet loss exhibited in the trace fragment, the corresponding part of the speech sample’s bitstream is deleted. This procedure ends up with a degraded bitstream. Decoding this bitstream, we obtain a degraded speech sample that is finally evaluated with PESQ by comparing it with the original speech sample. Now, we repeat this procedure using a trace fragment that is shifted by one packet, and so on. Viewing this loop from a different

[1] If a packet arrives too late at the receiver due to delay variation, it needs to be dropped.

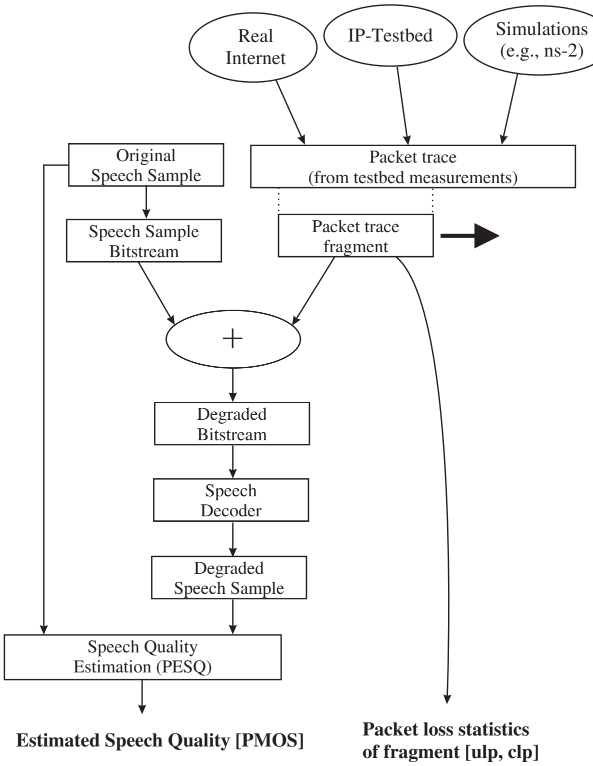

Fig. 4. Structure of the methodology.

perspective, we "slide" the speech sample along the packet trace in steps of one packet.

Figure 4 may serve as an instructive illustration of where the packet traces meet the speech samples. Note that the plus sign represents the matching process between speech sample bitstream and trace fragment, while the sliding process is indicated by the bold black arrow. In this way, we receive values for the perceptual speech quality represented by the PESQ-MOS which correspond to the packet loss statistics for the individual trace fragments. Figure 5 shows how a speech sample, containing important packets ("A") and less important packets ("B"), is sliding along a packet trace, with dark fields within the trace indicating individual lost packets.

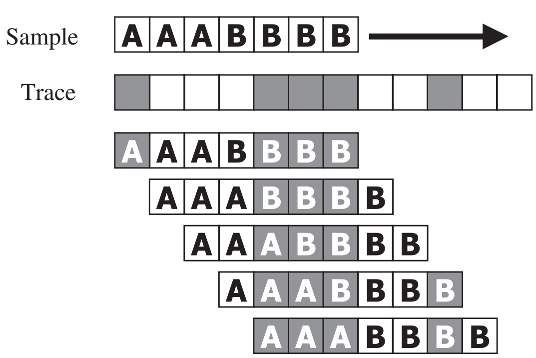

Fig. 5. Sliding of a speech sample along a packet trace. "A" and "B" denote important and unimportant packets, respectively.

C. IP-Testbed Measurements

As an example for the measurement phase described in the previous section, we have started with testbed experiments, because a testbed provides well-defined load situations with realistic traffic patterns, and therefore allows the emulation of realistic Internet scenarios.

Our testbed [14] is sketched in Figure 6 and consists of

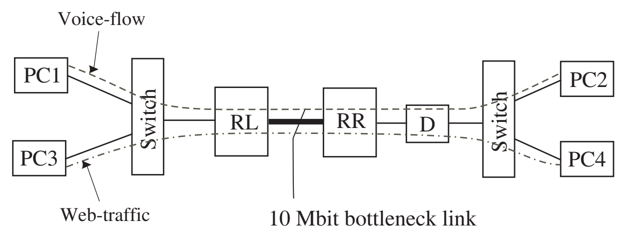

Fig. 6. Overview: Testbed

the following parts: Two PC-based routers RL and RR are connected by a 10 Mbit High-Speed Serial Link (HSSI). This link constitutes the bottleneck of the system. Voice packet losses happen due to congestion at the output queue of router RL towards the HSSI link. The buffersize at the routers is set to 200 packets. Two PCs, intended for traffic generation, are connected to each of the routers by a switch. Links between the PCs and the switches are 1 GB Ethernet, links between the switches and routers are 100 MBit Ethernet. To emulate realistic scenarios, a delay emulator D is connected in between RR and a switch in order to be able to delay packets as if they would traverse a long distance link.

We are using Web-like background traffic between servers emulated at PC3 and clients at PC4. Then, a voice flow is started between PC1 and PC2, carrying 20 Bytes of dummy data as RTP payload (resulting in a total IP/UDP/RTP packet size of 60 Bytes). In this study, we focus mainly on tracing packet losses and therefore incorporate also delayed packets through a large jitter buffer. The information about a packet being received or lost is stored at the sink PC2 in a file from which we can obtain the desired packet trace. For our measurements, the transmission delay is set to 100 ms.

The background traffic is generated according to the SURGE traffic model [15]. Based on an evaluation of realistic Internet traffic traces, SURGE models HTTP1.0 based client-server interactions. The client requests Web pages, and the server replies by opening a TCP connection for each object in the page. After downloading the page, the client changes into a state of inactivity before requesting the next page. We refer to [15] for the specific distributions and their parameter settings for client inactive time, object size, inter-object time, and the number of objects per web page. Our SURGE-based traffic generation tool has been extensively validated in prior investigations [16]. Note that a specific network load

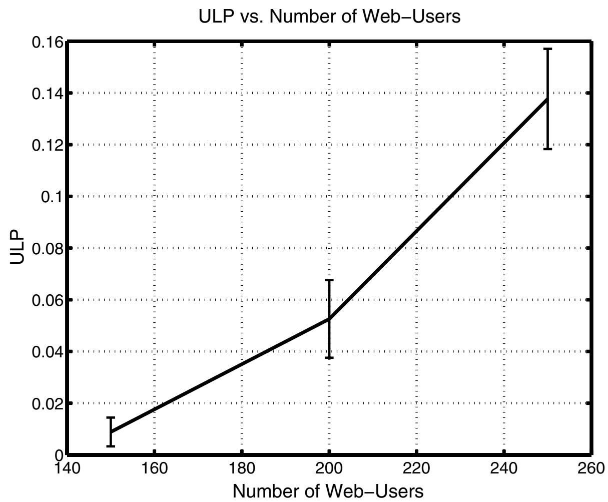

Fig. 7. Average ULP of 400-packet trace fragments vs. number of Web-Users.

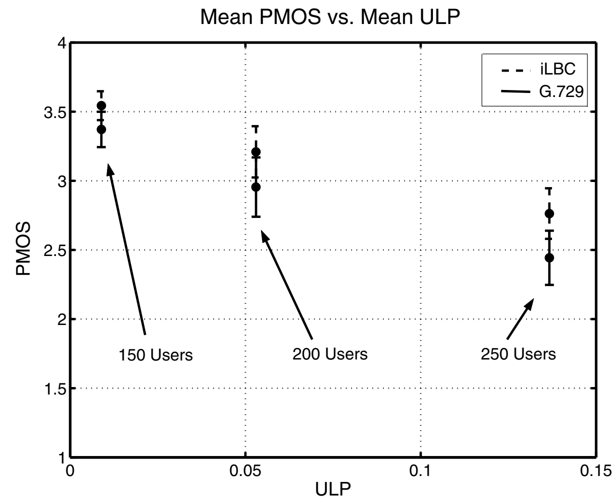

Fig. 8. PMOS vs. ULP: Comparison of G.729 and iLBC for different numbers of Web-users.

situation (i.e. overall packet loss probability) can simply be achieved by starting the corresponding number of client-server connections.

For the demonstration of our methodology, we have collected voice packet traces at the receiver for 150, 200 and 250 emulated Web-users. In order to demonstrate the relation between Web-users and overall loss probability, the ULP values (including their standard deviations) are depicted in Figure 7. The respective means and standard deviations of the three conditions are shown in Table I. These numbers show that the traffic load resulting from our selection of user numbers cover a relevant spectrum of load situations.

TABLE I

MEANS AND STANDARD DEVIATIONS OF THE ULP FOR DIFFERENT NUMBERS OF WEB-USERS.

No Users	ULP	
	$\bar{x}$ [%]	σ [%]
150	0.89	0.56
200	5.26	1.50
250	13.77	1.94

Based on these traces, in section 4 we present evaluation results for the three load scenarios related to 150, 200 and 250 users, resp., where the scenarios are characterized in terms of the corresponding ULP's throughout.

IV. RESULTS AND DISCUSSION

A. Perceptual Quality vs. ULP

The average speech quality (including standard deviation) that corresponds to the three different load situations is depicted in Figure 8. All of our results are based on 60000 packet trace fragments and their corresponding PMOS values.

Due to the varying network behavior, the actual ULP is subject to sudden changes in each of the scenarios. Therefore, the scatter plots of Figures 9–11 provide a much more detailed insight into how the perceptual speech quality is distributed over the PMOS-ULP plane. Each figure presents the performance of a certain number of Web-Users. For comparison, the left column presents the results for the G.729 codec, whereas the right column depicts the corresponding results for the iLBC codec. Each dot represents a PESQ result based on a specific position of the sample's bitstream when sliding along the packet trace. As a major result with regard to the performance of the codecs, we clearly observe the superior coding quality and graceful degradation of the iLBC. The ULP is bounded to [0.00%;2.87%] for 150 Web-Users, [0.87%;10.49%] for 200 Web-Users, and [7.74%19.60%] for 250 Web-Users. Thus, the range of packet loss within the trace fragments of 400 packets overlap to a significant extent.

B. Burst Performance

An important issue regarding the relation of packet loss and perceptual quality is the number of packets which are lost in sequence. Note that our definitions of a burst includes all events where two or more packets are loss immediately one after the other. Bursts of losses increase the impairment since the loss concealment processing algorithm synthesizes a signal which is based on the latest received packet. Thus, it produces a static, sometimes “robot”-sounding output signal. In the Gilbert Model, loss burstiness is represented by the CLP, as we have presented in section II. Figure 12 illustrates the perceptual performance (including standard deviations) with regard to burstiness as indicated by the CLP (depicted along the x-axis).

It is remarkable that for both codecs the PMOS does not

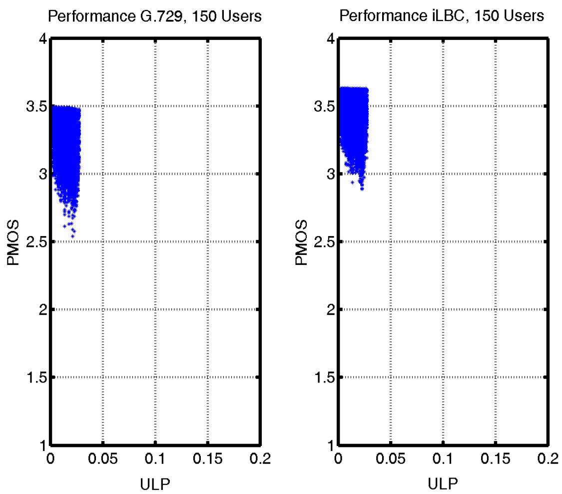

Fig. 9. PMOS vs. ULP for G.729 and iLBC for 150 Web-users.

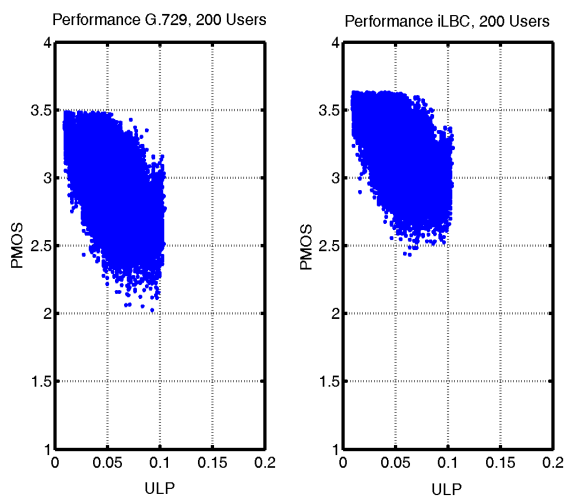

Fig. 10. PMOS vs. ULP for G.729 and iLBC for 200 Web-users.

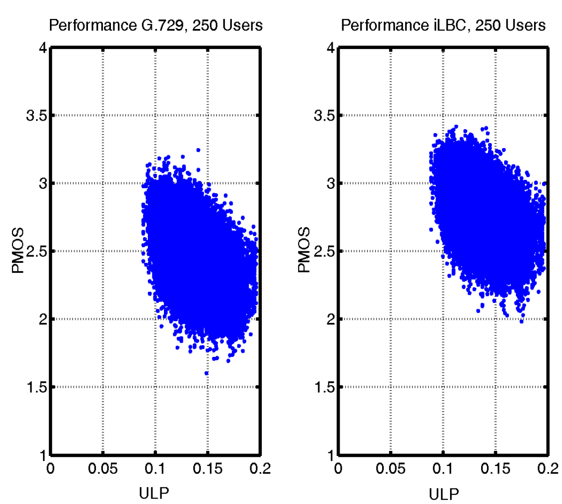

Fig. 11. PMOS vs. ULP for G.729 and iLBC for 250 Web-users.

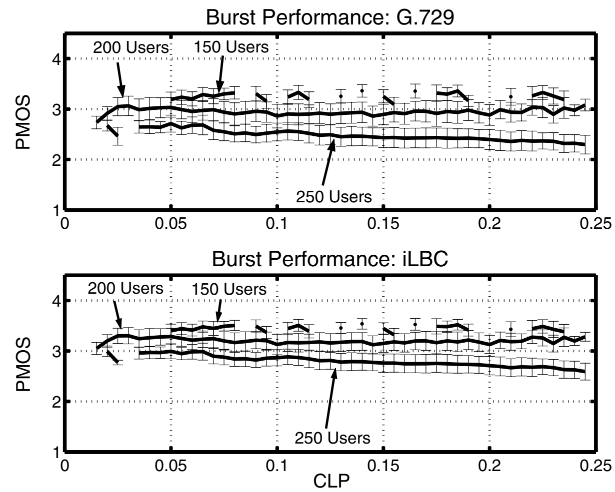

Fig. 12. Burst loss performance: Comparison of G.729 and iLBC with respect to the CLP.

decrease with growing CLP for 150 and 200 users, whereas for a number of 250 users, the PMOS decreases by 0.38 points for G.729 and 0.40 points for iLBC. Except for the difference in performance that we have demonstrated already in the previous section, the CLP performance numbers suggest both codecs to have the same behavior in bursty situations. In fact, high speech quality at high CLP, i.e. high burstiness, seems to be a contradiction.

In the following paragraphs, we will explain why the use of the CLP as a burstiness metric is not appropriate in a performance assessment environment using speech samples of 8 seconds (400 packets of 20 ms speech information). In situations of low ULP, consecutive packet losses result in higher CLP values. As an example, if only two packets are lost consecutively out of 400, the CLP results in 50 % whereas the ULP equals 0.5%. Therefore, we introduce a new metric which better reflects the relation between bursty losses and the perceptual speech quality, i.e. the "Effective Burst Probability" (EBP):

$$EBP = CLP * ULP. \tag{3}$$

In analogy to the coefficient of variation, which relates the standard deviation to the mean, the EBP establishes a comparable relationship between the CLP and the ULP by

TABLE II

MEANS AND STANDARD DEVIATIONS OF THE EBP FOR DIFFERENT NUMBERS OF WEB-USERS.

Number of Users	EBP $\bar{x}$ [%]	EBP σ [%]
150	0.0631	0.1014
200	0.5547	0.3239
250	2.2436	0.6935

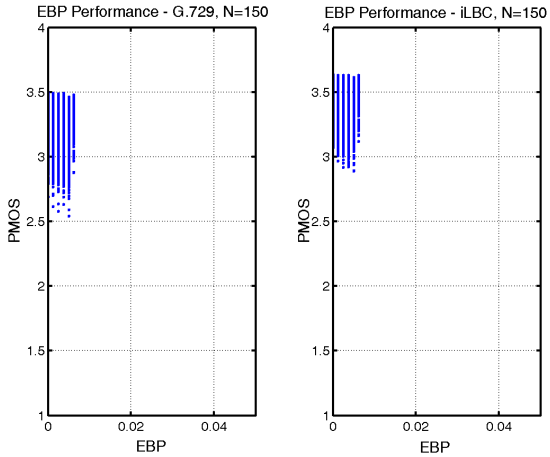

Fig. 13. PMOS vs. EBP for G.729 and iLBC (150 Web-Users).

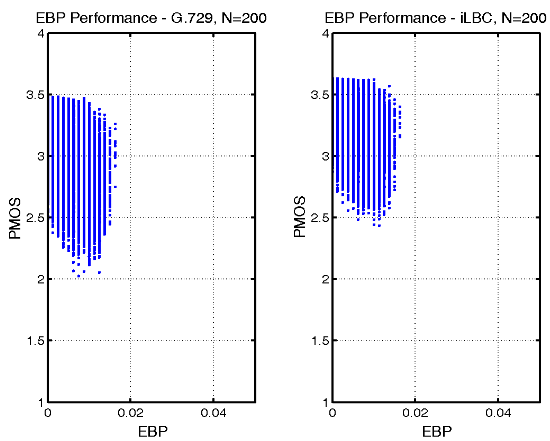

Fig. 14. PMOS vs. EBP for G.729 and iLBC (200 Web-Users).

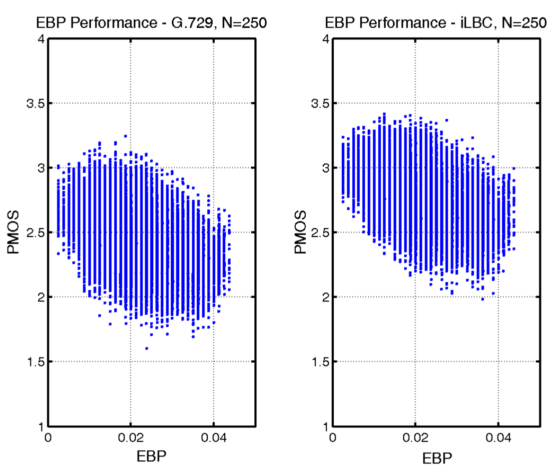

Fig. 15. PMOS vs. EBP for G.729 and iLBC (250 Web-Users).

appropriately normalizing the former one. Table II depicts the mean and standard deviation for the different numbers of Web-Users.

For illustration, Figures 13–15 present the PMOS performance of both codecs with regard to the EBP. These figures demonstrate the impact of increasingly bursty losses on the perceptual quality. While Figure 13 shows that a network load of 150 Web-Users does hardly introduce bursty losses to the voice stream, Figures 14 and 15 indicate the perceptual quality resulting from a broad range of burst situations.

Figure 16 presents the distribution of the EBP for the three scenarios. The first graph suggests as conclusion that 150 Web-Users hardly introduce burst losses to the voice-flow, whereas the effective burstiness EBP rises with an increasing number of users as indicated by the graphs below.

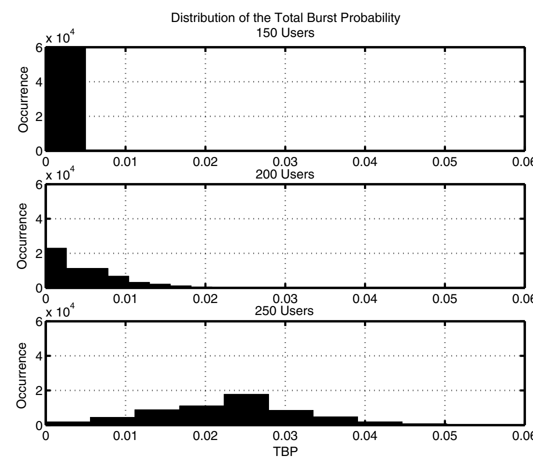

Fig. 16. Distribution of the Effective Burstiness Probability.

C. *Time-Varying Quality*

As already mentioned above, one of the major advantages of our approach concerns the fine-granularity of the QoS evaluation results, which for our proposals corresponds essentially to the length of one packet or 20 ms only. This immediately triggers the question to which extent a loss pattern may cause variations in PMOS on such a small time-scale. Figure 17 depicts the difference of the PMOS values of G.729-coded consecutive speech samples at a load situation of 250 users. As a surprising result, we observe that the PMOS value can change by up to 0.64 within 20 ms. This effect can be explained from the fact that not every packet possesses equal perceptual importance (cf. [17], [18]). Thus, our results confirm the large influence of the location of packet losses within the phonetical structure of the spoken word that has already been observed e.g. by [19].

V. Conclusions

This paper proposes a new, efficient and accurate methodology for the instrumental QoS evalution of VoIP. The approach is based on a two-step procedure: first, we

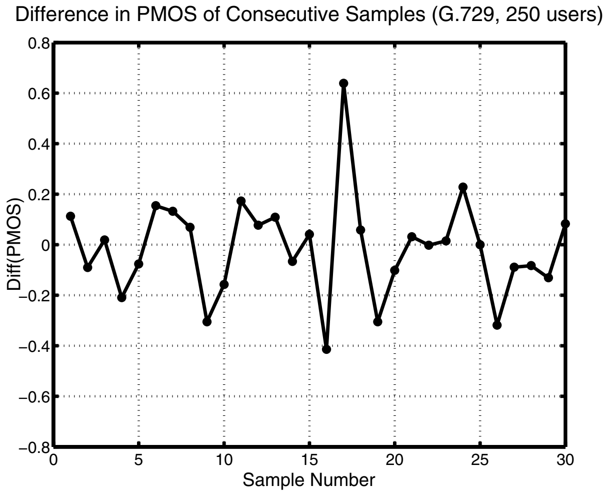

Fig. 17. Difference between consecutive PMOS values.

obtain off-line measurements of packet loss traces from arbitrary IP-based networks, and in a second step we match trace fragments against the bitstream of the speech sample in an overlapping manner and evaluate the resulting degraded speech sample using PESQ. We have validated our approach by evaluating the performance of VoIP transport in an IP-testbed for a relevant range of load situations and two different speech codecs. Due to the fine-granular evaluation results presented in this paper, we consider this methodology to be useful for speech quality assessment also in a highly time-varying environment such as mobile VoIP. The comparison of G.729 and iLBC (including their behavior under bursty loss patterns) with regard to a new proposal for a burstiness metric and an interesting result on potential PMOS changes on very small time-scales round off this study. Current and future work includes the evaluation of perceptual packet marking algorithms, the comparison of results for artificial voice vs. samples in different languages, the evaluation of further state-of-the-art codecs, and the integration of future network QoS mechanisms.

ACKNOWLEDGEMENTS

This work has been funded under the Austrian government's Kplus Competence Center Program. The authors would like to thank Eduard Hasenleithner for implementing the voice traffic generator and for his help regarding various aspects of the IP-testbed.

REFERENCES

[1] A. Bouch, M. A. Sasse, and H. Demeer, "Of packets and people: A user-centered approach to quality of service," in *IWQoS 2000*, Pittsburgh, June 2000.

[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] ——, "TOSQA - Telecommunication objective speech quality assessment," *ITU-T COM 12-34*, Dec. 1997.

[4] J. G. Beerends, A. P. Hekstra, A. W. Rix, and M. P. Hollier, "Perceptual evaluation of speech quality (PESQ): The new ITU standard for end-to-end speech quality assessment part II - psychoacoustic model," *Journal Audio Eng. Soc.*, vol. 50, no. 10, pp. 765–778, Oct. 2002.

[5] International Telecommunication Union, "ITU-T coded-speech database," *ITU-T Series P, Supplement 23*, Feb. 1998.

[6] ——, "Coding of speech at 8 kbit/s using conjugate-structure algebraic-code-excited linear-prediction (CS-ACELP)," *ITU-T Recommendation G.729*, Mar. 1996.

[7] S. V. Andersen *et al.*, "Internet low bit rate codec," *Internet Draft, draft-ietf-avt-ilbc-codec-04.txt*, Nov. 2003.

[8] W. Jiang and H. Schulzrinne, "Modeling of packet loss and delay and their effect on real-time multimedia service quality," in *Proc. Int. Workshop Network and Operating Systems Support for Digital Audio and Video NOSSDAV*, Chapel Hill, NC, June 2000.

[9] H. Sanneck and G. Carle, "A framework model for packet loss metrics based on loss runlengths," in *SPIE/ACM SIGMM Multimedia Computing and Networking Conference (MMCN 2000)*, San Jose, CA, Jan. 2000, pp. 177–187.

[10] 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," in *Int. Conf. Telecommunications*, vol. 1, Feb. 2003, pp. 543–552.

[11] International Telecommunication Union, "Artificial voice," *ITU-T Recommendation P.50*, Sept. 1999.

[12] A. E. Conway, "A passive method for monitoring voice-over-IP call quality with ITU-T objective speech quality mesurement methods," in *Proc. IEEE Int. Conf. Communications*, vol. 4, New York City, NY, 2002, pp. 2583–2586.

[13] NS Simulator Homepage, http://www.isi.edu/nsnam/ns/.

[14] E. Hasenleithner, H. Weisgrab, and T. Ziegler, "Final testbed report phase 1," Telecommunications Research Center Vienna, Tech. Rep. A0 Deliverable No. 5, Feb. 2002.

[15] P. Barford and M. Crovella, "Generating representative web workloads for network and server performance evaluation," in *Proc. ACM SIGMETRICS*, Madison, WI, July 1998.

[16] E. Hasenleithner and T. Ziegler, "Comparison of simulation and measurements using state of the art web traffic models," in *Proc. IEEE Symposium on Computers and Communications (ISCC'2003)*, June 2003.

[17] C. Höne, B. Rathke, and A. Wolisz, "On the importance of a voip packet," in *Proc. of ISCA Tutorial and Research Workshop on Auditory Quality of Systems*, Mont-Cenis, Germany, April 2003.

[18] F. Hammer, P. Reichl, T. Nordström, and G. Kubin, "Corrupted speech data considered useful: Improving perceived speech quality of voip over error-prone channels," to appear in *Acta Acustica*, 2004.

[19] L. Sun, G. Wade, B. M. Lines, and E. C. Ifeachor, "Impact of packet loss location on perceived speech quality," in *IPTEL 2001*, New York City, NY, Apr. 2001.