

Assessment of objective voice quality over best-effort networks

Jari Turunen*, Pekka Loula, Tarmo Lipping

Tampere University of Technology, Pori Pohjoisranta 11, P.O. Box 300, FIN-28601 Pori, Finland

Received 14 May 2003; revised 21 January 2004; accepted 17 February 2004

Available online 22 December 2004

Abstract

VoIP calls transferred over dedicated bandwidth or QoS capable networks is a cost-effective alternative for PSTN in large enterprises. However, the calls made over the best effort network, such as the global Internet, suffer packet loss and jitter. In some VoIP-codecs, such as ITU G.723.1 and G.729a, there are built-in recovery mechanisms for concealing packet-based errors in the audio/speech stream. These recovery mechanisms can conceal up to 5% packet losses without significant quality degradation, as shown in this article. The 5% quality degradation is approximately within 0.5 MOS scale when compared to the original signal. Beyond 5%, the speech quality will drop gradually. The overall quality of MOS scale 3 can be maintained even with 14–17% packet loss rates. The influence of delay variation or jitter cannot be eliminated with the concealment algorithms unless the jitter time exceeds packet loss indication delay. The influence of jitter is not critical below 20 ms but beyond 20 ms limit its influence will decrease the speech quality very steeply. This suggests that packet losses can be recovered in normal conditions, but the influence of jitter must be eliminated somehow. The interleaving and piggybacking-based stream manipulation enhances speech quality in packet dropout situations. The guaranteed delay over the whole Internet would enhance the possibilities of VoIP to achieve success.

© 2004 Elsevier B.V. All rights reserved.

Keywords: Packet concealment; Jitter; Delay; VoIP

1. Introduction

The Voice over Internet Protocol (VoIP) calls were predicted to replace ordinary phone calls in all IP-based communication scenarios. However, it seems that the full-scale implementation and of the VoIP is still waiting its final success. Several reasons are pointed out in numerous research reports that have partially affected to the VoIP breakthrough as a communication channel. The main reason for the delayed success of the VoIP might be that the Internet was designed to be a fault tolerant data exchange/transmission medium and the traffic re-routing is the primary target in the case of Internet server additions and removals. The delivery time or transmission delay was never the primary goal in the design phase although the Internet Protocol (IP) has a support for real-time

transmission. The IP-protocol comprises User Datagram Protocol (UDP) and Real-Time Protocol (RTP) that are primarily designed for stream transmissions. Despite the benefits that UDP and RTP can offer to the stream transmission, VoIP has suffered speech quality problems over the Internet, as reported for example in [1–3]. The voice packets are lost or damaged in transmission, the transmission delay might be too long for interaction and the changes in transmission rate can make the interaction with VoIP very uncomfortable.

The average global packet loss is around 5% in the Internet core networks and peak losses are well beyond 10%. The average round-trip delay is below 200 ms with peaks up to 500 ms, as seen in [4]. The round-trip delay consists of transmission time from sender to destination and the echoed message return time to sender. The end-to-end transmission delay time is approximately half of the round trip delay time. These values vary from time to time due to network traffic, congestion and routing situations [5]. Although the traffic and its changes on the Internet can be

* Corresponding author. Tel.: +358 2 627 2748; fax: +358 2 627 2727.

E-mail addresses: jari.j.turunen@tut.fi (J. Turunen), pekka.loula@tut.fi (P. Loula), tarmo.lipping@tut.fi (T. Lipping).

predicted with some accuracy, this information does not help the caller who tries to establish an urgent VoIP call over a congested network.

Jitter, or delay variation, is a harmful characteristic of the Internet. There are several aspects that cause jitter, starting from the traffic re-routing and different packet queuing mechanisms over the Internet [5,6]. The influence of jitter in media streams can be smoothed using buffers, but there are some problems which will be discussed later.

VoIP problems have gained worldwide interest and several studies have been made to improve packet loss recovery methods in packet based communication networks [7–19]. Also, the network point of view has been studied, for example, in [20,21], and several solutions have been proposed for ensuring the quality of service over the congested network. But still there are major problems in speech transmission and decoded speech quality despite of the developed solutions.

When considering new inventions in speech coding from a practical point of view, the adaptation of new technology to the existing equipment is a slow process, no matter how clever and simple the methods are. Also, interoperability must be guaranteed with the older VoIP phone versions. In this paper, we compare two commonly used VoIP coders, ITU-T G.723.1 and G.729a, widely used in most VoIP systems, and their ability to handle lost packets with and without a lost packet concealment system. The main purpose is to evaluate the usefulness and appropriateness of the built-in error concealment system in the coders in question. We made the codec evaluations using Objective Mean Opinion Score (OMOS) that gives an opportunity to make a more detailed analysis when compared to ordinary subjective MOS [22]. Later, we discuss different aspects that may have influenced the VoIP penetration level.

2. Codecs and network

In VoIP-based calls, the speech is encoded and packed to a RTP/IP frame for transmission. Nowadays three codecs (codec = encoder/decoder) are common in VoIP, the first is ITU-T G.711, because it is mandatory in VoIP protocols, and the other two are optional codecs, G.723.1 and G.729a. The latter two codecs are more efficient than G.711 because they compress the speech before transmission in order to save bandwidth in the networks. There are several other voice codecs that can be implemented in the VoIP phones, but they are not addressed in this paper due to their limited success. The G.711 codec is a logarithmic (a-law/ μ -law) pulse code modulation codec operating at 64 kbit/s with toll quality. There is no packet error concealment algorithm in this codec, because the coding is performed on a sample-by-sample basis and the bitrate is absolutely too big in order to benefit from IP-networks. This codec will not be discussed later in this paper due to its simple mechanism without any built-in recovery mechanism. The G.723.1 is a two-mode

near-toll quality codec with 5.3 and 6.3 kbit/s bit rates. The G.729a is also a near-toll quality codec operating at 8 kbit/s.

Built-in error concealment mechanisms are used in both G.723.1 and G.729a codecs. The packet concealment mechanism in G.723.1 was designed for burst lost packet errors ignoring the random bit errors. In the case of a lost packet in G.723.1 the linear predictive filter coefficients are estimated from the past values, but the new information is set to zero. The excitation signal is copied from the last good packet information combined with its voiced/unvoiced classification. The unvoiced excitation is obtained from the random number generator with re-estimated gain and the voiced excitation is generated with a periodic pulse signal according to the pitch period and gain from the classifier. If the erasure lasts for the next two packets, the excitation is attenuated 2.5 dB for each packet. After three estimated speech frames the decoder and estimation are stopped [23].

In the case of G.729a, the linear predictive filter parameters are also estimated from the last good packet values. The excitation is constructed using attenuated adaptive and fixed codebook values from the previous good frame [24,25]. The overview of the packet concealment methods is presented in Fig. 1.

Speech encoding, transmission, buffering and decoding takes time. Contrary to the public switched telephone network (PSTN) phone call, the spoken VoIP message will be delayed, due to reasons mentioned earlier, before the listener can hear it at the other end. A delay exceeding a certain threshold will cause disturbances in interaction and will feel very uncomfortable during the call. That is why in ITU-T G.114 recommendation, the maximum end-to-end delay, also referred to as ‘mouth-to-ear’ delay, in real-time communications is recommended to be less than 150 ms. An end-to-end delay consists of data collection, encoding delay, network transmission delay, network buffering delays, and decoding delay. In ITU-T multimedia communication-related recommendation, H.225.0, it is recommended that the default-framing interval for audio should be 20 ms. The usage of longer packets than 20 ms is allowed unless the overall delay exceeds the end-to-end delay mentioned above. This means two 10 ms packets of speech for G.729a and one 30 ms speech packet for G.723.1. In the case of packet loss in transmission, the more speech information the lost packet will contain the greater the degradation will be in the decoded speech quality.

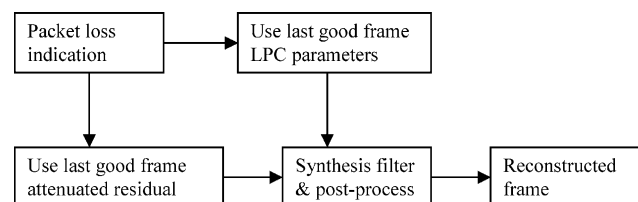


Fig. 1. The general schematic diagram of the packet reconstruction process.

3. Simulations

Real-time transmission with packet losses was simulated with G.723.1 and G.729a codecs. For evaluation 27 min of conversational speech data, consisting of 60% male and 40% female speech, sampled at 8000 Hz frequency was encoded and decoded. The same sampling frequency is used in the codecs mentioned above.

The speech quality assessment method was selected to be the ITU-T P.862 Perceptual Evaluation of Subjective Quality. This method is also called Objective MOS (OMOS), because it predicts the subjective quality of speech evaluated by humans with Mean Opinion Score (MOS) or Degradation Mean Opinion Score (DMOS) scale. The scale of these subjective methods is presented in Table 1. For example, number 3 defined by OMOS means ‘fair’ quality of analyzed speech and the distortions are ‘perceptible and slightly annoying’.

The delay distribution of the transmitted speech was selected to follow exponential distribution corresponding the actual delay distribution in best-effort networks, as stated in [26]. The shape of the distribution is adjusted using the estimate of the expected mean value. The jitter was defined for the test cases using the exponential random data where the expected mean corresponds to the wanted (average) jitter. In Fig. 2, two cases of exponential distributions are presented.

The baseline codec tests are made with a built-in recovery method for lost packets. The G.723.1 test was performed with 6.3 kbit/s in this baseline simulation. The median results of the baseline simulations with 0–10% packet losses and jitter from 0 to 50 ms are presented in Fig. 3. In the case of G.729 two consecutive packets that are delivered with one IP-frame are lost at once if the IP-frame loss occurs and only one G.723.1 packet will be lost in the IP-frame loss case.

The overall quality of speech is almost similar in both cases and the zero-jitter/zero-loss results are close to ITU-T reference results (G.723.1 3.9, G.729a 3.7). The average saturation level over 20 ms jitter is slightly higher with G.723.1 than with G.729. The significance of the packet loss is smaller when compared to jitter, as seen in Fig. 3. This is obvious because the codec can conceal few consequent packet losses by estimating a replacement for them, but the influence of jitter cannot be concealed unless it exceeds the packet loss indication delay. The quality decreases very

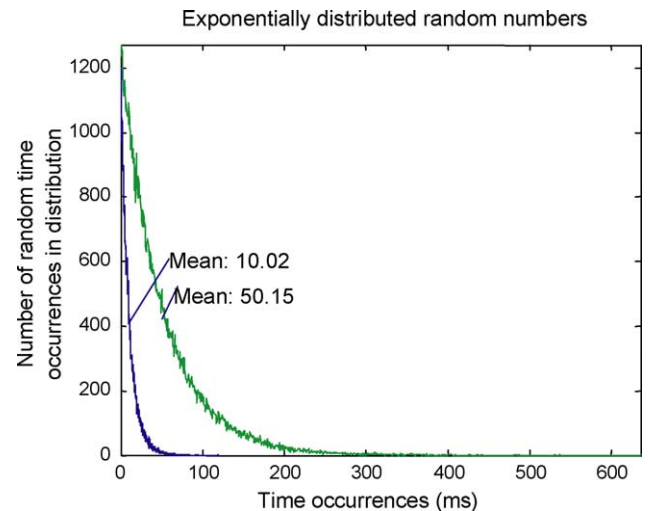


Fig. 2. Two distributions of exponential random numbers. Each distribution histogram is plotted from 100,000 data points using mean values of 10 and 50 ms, respectively.

rapidly even with small jitter value increments below OMOS value 3.

In Fig. 4, the experimental speech quality evaluation results of packet losses between 0 and 95% are presented. The presented results are the median of the data set. Jitter is in this case 0 ms and the packet concealment is enabled in both codecs.

In Fig. 4, the degradation of the speech quality between 20 and 70% packet losses is almost linear in the case of G.723.1. G.729a Fig. 4 shows rapid quality degradation in the beginning and later it settles to a more steady downhill. The close-to-linear behavior can be explained by the fact that in conversational speech, there are silence sections between the talk spurts, so part of the random losses will automatically hit the silence sections. However, the speech quality will be maintained at 0.5 MOS level in 9% packet loss limits with G.723.1 and in 6% limits with G.729. The ‘fair’ quality will be still acquired with 17% packet losses in the case of G.723.1 and with 14% in the case of G.729.

In the recovery process, the influence of the recovery mechanisms on the speech quality was measured. The same data was fed into the codec twice, with and without packet concealment. In Fig. 5, both G.723.1 and G.729a results are presented.

The recovery mechanism can conceal the packet losses effectively in small jitter cases. However, the influence of the jitter cannot be recovered with the recovery mechanism and the recovery efficiency will drop dramatically as the jitter increases. The behavior of 20% packet loss curves for the two codecs is due to the fact that in case of G.729a two speech frames are packed into a single RTP/IP frame while in the case of G.723.1 each RTP frame contains only one speech frame. The extrapolation of the packet recovery beyond 20% packet loss will not raise the quality in the case of G.723.1 codec. The quality will start to weaken also in the case of G.723.1 if over 20% of the packets are lost.

Table 1
The speech quality evaluation

Rating	Speech quality (MOS)	Level of distortion (DMOS)
5	Excellent	Imperceptible
4	Good	Just perceptible but not annoying
3	Fair	Perceptible and slightly annoying
2	Poor	Annoying but not objectionable
1	Unsatisfactory	Very annoying and objectionable

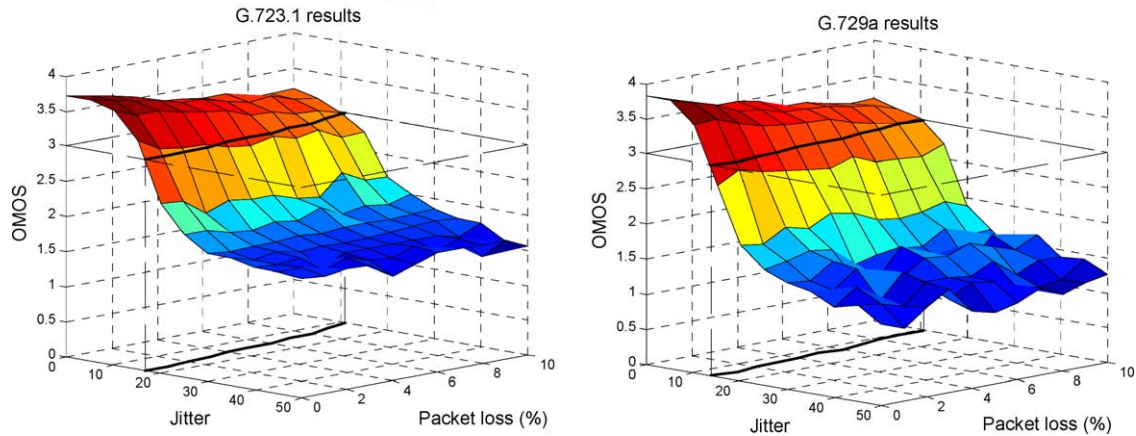


Fig. 3. Speech quality comparison between G.723.1 and G.729a codecs. The OMOS speech quality limit 3 is marked in the figures with dashed lines.

Several possibilities are reported in literature to reduce the effect of packet loss with packet transmission manipulation. The two well-known methods are packet interleaving [27] and transmitted information multiplying that is also called piggybacking. The basic idea of interleaving is that a sequence of samples or packets, ordered as: (1,2,3,4) (5,6,7,8) (9,10,11,12) (13,14,15,16), etc. is re-arranged into the sequence of (1,5,9,13) (2,6,10,14) (3,7,11,15) (4,8,12,16), etc. This means that in the case of one IP-frame loss only every 4th sample or speech packet is lost and the lost item is easy to interpolate from the arrived items. However, this arrangement cannot be made for real-time packet-based communications. If the frame takes 10 ms, the above arrangement will take 160 ms solely for buffering. However, a more simplified mechanism can be tested. In the case of real-time transmissions, the packet interleaving cannot be used over a wide range of packets in time. In this test, the encoded speech packets are arranged and interleaved from the sequence (1,2) (3,4) (5,6) (7,8), etc. to the sequence (1,3) (2,4) (5,7) (6,8), etc. This arrangement will add a 20 ms

delay to the stream that can be eliminated by the jitter buffer. The piggybacking meant originally that the packet in question would include some information from the past and/or future packets. In this test, this structure is simplified in a way that the packets are sent twice. In the case of G.729a piggybacking it means 16 kbit/s transmission instead of 8 kbit/s. The packet stream is converted from (1,2) (3,4) (5,6) (7,8), etc. into (The first packet optionally (0,1)) (1,2) (2,3) (3,4) (4,5) (5,6) (6,7) (7,8), etc. In Fig. 6, the packet interleaving and piggyback results are presented.

As shown in Fig. 6, interleaving will increase the speech quality only approximately 0.1 OMOS when compared to conventional speech transmission at 0–20% packet loss rates. Piggybacking will increase speech quality by 0.2–0.4 OMOS at 0–20% packet loss rates, and maintain the OMOS quality of 3 at over 25% loss rates while the conventional speech can hold the OMOS quality of 3 with a 15% loss rate. The influence of simple interleaving is merely marginal taking into account the extra delay and effort required by interleaving.

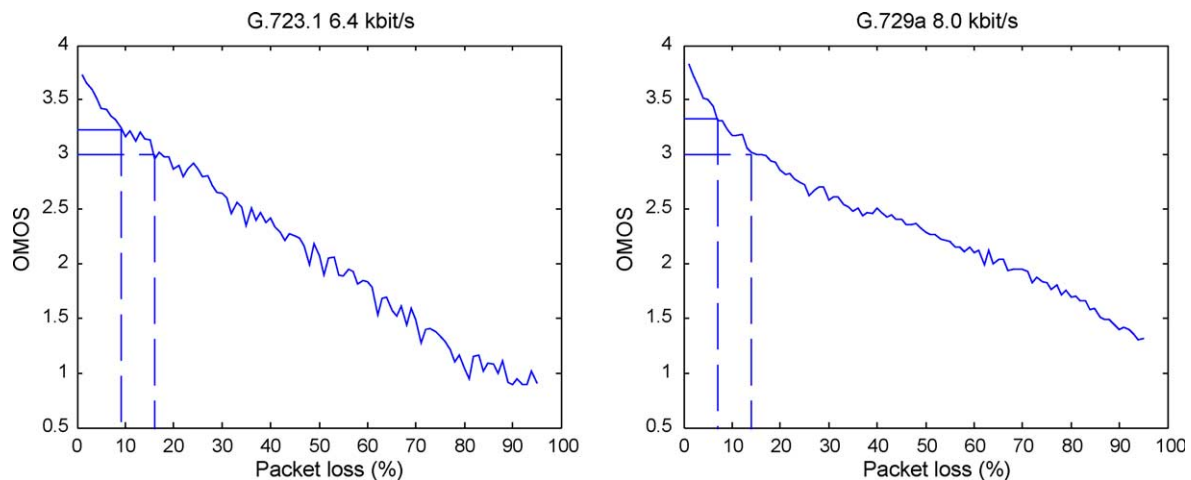


Fig. 4. The speech quality with packet losses 0–95% in the case of G.723.1 and G.729a codecs. The first dashed line shows the 0.5 MOS degradation limit and the second one the ‘fair’ quality limit against the packet losses.

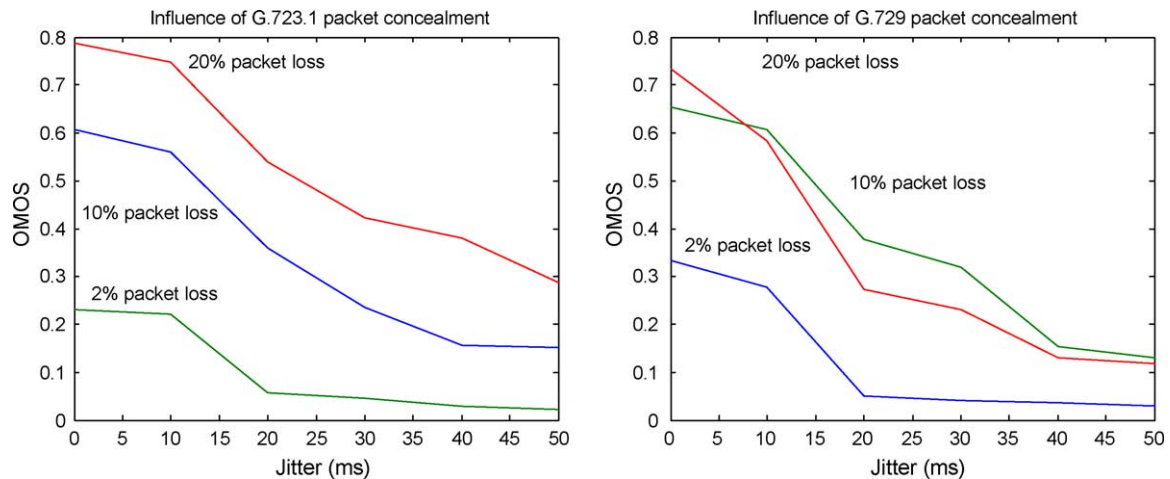


Fig. 5. Influence of the packet recovery mechanism in G.723.1 and G.729a codecs.

4. Discussion

In individual calls, packet losses in the network are not as significant as the effect of jitter in VoIP calls. The built-in recovery mechanisms are capable to maintain the perceived quality up to 5% packet losses with 0.5 MOS degradation. This means that the codec built-in recovery methods are usually appropriate for the VoIP calls. The significance of jitter is also important, but the effect of jitter can be ignored as far as it remains well below the jitter buffer size in question. The jitter buffer size can be either adaptive or fixed. In both cases, the smaller the buffer size is the shorter the delay is in the transmitted speech. On the other hand, the bigger the buffer is the less breakups there will be in the decoded speech. Somewhere between these two extremes is the optimum buffer size that minimizes the effects of delay and breakups.

Statistically, jitter and delay values between different network backbone servers are usually well below the size of, for example, 50 ms fixed jitter buffer. The effect of jitter

can be eliminated with the jitter buffer and packet losses can be covered up to 5–10% without significant degradation. More than that, piggybacking will enhance the quality in the case of even bigger packet loss rates. Beyond these well-known problems and solution approaches, there are several other tools, studies and algorithms available considering packet losses and network, where the different problems considering the whole VoIP chain are pointed out, explained and solved. Why has the VoIP not gained better success than it has nowadays?

The quality of transmitted speech over VoIP systems depends on many factors. The traffic between wide area network servers can be statistically steady and suitable for real-time purposes, but the rush hour traffic and other unexpected increases in the network traffic may even stop the individual call transmission totally. In very rare cases, the individual VoIP caller is directly connected to the backbone server. Instead of that, the VoIP connection is taken from LAN, ISDN or dial-up modem connection and it is impossible to predict how much latency, jitter or other

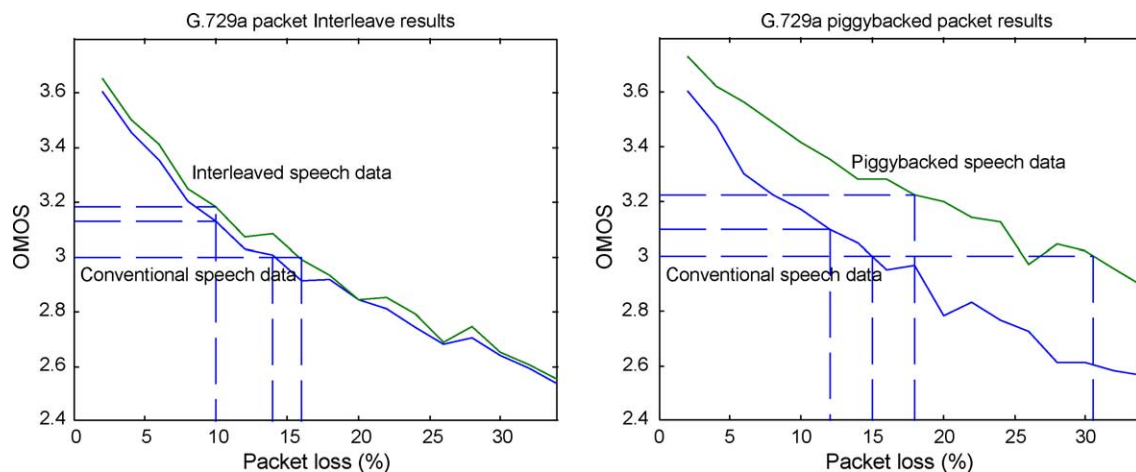


Fig. 6. The comparison results for packet interleaving and piggybacking against conventional speech data transmission. The dashed lines show the 0.5 MOS degradation limits and 'fair' (3) quality limits.

degradation these network connections will add unless there is a dedicated bandwidth available for real-time connections.

VoIP calls have replaced the ordinary telephones in many enterprises considering company-related intra calls. Usually, these companies already have an intranet and a decent Internet connection for data transfers. Turning the ordinary telephones to VoIP therefore reduces the double network investment costs to one network investment for the company. Also, the maintenance will drop from two to one-network annual maintenance costs. From these points of view VoIP turns out to be cost-effective in company intra calls. The workers need to make calls to ordinary telephones, portable phones and the enterprise's subsidiary numbers. Intranet-based VoIP must have a possibility to be connected to PSTN and portable phone network via dedicated gateway. Usually, large enterprises will have these facilities and guaranteed data connections, such as QoS capable networks or a fixed bitrate connection, with their subsidiaries, so the VoIP calls will use the same connection and the call costs can be included in the data connection costs. The calls between enterprises can be operated using, for example, networks that will give the highest priority to real time connections, such as telephone calls. The only problem is the enterprise calls over the Internet to another VoIP interface, where the quality of the transmitted speech will be defined based on overall traffic and router-to-router connections. However, the ordinary callers must usually also make a VoIP call under these circumstances all the time.

There are also weaknesses in the VoIP environment. For example, incompatibilities between two pieces of different vendor equipment have forced the callers to use the mandatory G.711 codec in some VoIP cases. This is equal to putting ISDN calls into the local network and increasing the data traffic by extra 128 kbit/s (64 kbit/s send and 64 kbit/s receive) per caller. Therefore, the use of compression methods is more than welcome in VoIP speech transmission. In addition, special expertise is needed in some cases to handle real-time connection maintenance that differs from data connection maintenance.

Guaranteed data connection usually enhances data transmission, but the fact is that the bandwidth consumer buys only a guaranteed bandwidth or certain guaranteed delay to the data from the operator. Naturally, the costs are higher when the user demands are tighter. The operator guarantees certain delivery time or throughput, but nothing will control the average delay or delay variation in guaranteed networks [28]. Fortunately, if the bought delivery time is very tight, say, below 50 ms, the influence of the delay and jitter can be eliminated with appropriate jitter buffer.

VoIP is a successor but also a competitor for the PSTN-based calls. The PSTN is a robust and existing network with a very simple user interface and the call establishing

procedure is compatible around the globe. Telephone operators have an existing medium, system and connections to all other operators and from the operators' point of view this might have caused the lack of interest to develop a new system parallel to the existing one. For the individuals and the companies the price is the only way how the VoIP can compete against the PSTN.

Several deficiencies can be seen in the user interfaces and in the usability of the terminals, but the constantly evolving design and development work will bring practical solutions to usability problems. However, in the case of incompatibility between different VoIP terminals forces to use the mandatory G.711 call, which is impossible to use over analogue modem data line.

Several aspects should be considered in VoIP-based calls with the coding. For example, RTP and its Real time Transmission Control Protocol (RTCP) have several properties that can be utilized to improve VoIP calls. The RTCP will frequently send information on the arrived frames, lost frames and jitter to the transmitter, so this information can be and also has been used for adjusting, for example, the size of the jitter buffer. RTP has the payload type field defined in the RTP header, and for example in RFC1890 the predefined audio and video types for RTP are presented. In addition, there are dynamic payload type numbers that can be selected freely by the application. In VoIP protocols, some of these properties are already reserved to be used in protocol. The G.711 is the only mandatory codec. Both Session Initiation Protocol (SIP) and H.323 support the change of the speech codec 'on-the-fly', but despite the mandatory and 'de-facto' codecs, such as G.723.1 and G.729, the number of supported codecs is solely based on the proprietary solutions.

The VoIP offers a framework to make adjustments during a call to enhance the stream quality for the terminal programs and agents. The primary purpose of the terminal agent is to maintain the speech quality at least at OMOS 3 level. For example, as seen in our studies, jitter reduction plays more significant a role than packet loss concealment. By the help of RTCP, the stream transmission information about jitter and packet losses, the agent can tell the terminal program to adjust the jitter buffer size automatically. According to the packet loss information, the terminal program can be asked to change the current codec to a low-bit rate codec and in severe packet loss moments, the piggybacked or interleaved encoded speech transmission can be taken in to use.

To conclude this paper, it seems that VoIP will be an alternative for speech transmission, which is based on corporate intranet. Also in these corporate networks low-bit rate codecs are capable of handling under 5% packet loss rates maintaining good quality and jitter can be eliminated with appropriate buffers. However, calls over the Internet can be guaranteed with this generation of protocols and networks using QoS services in call transmission. Also, by using the agents in call monitoring the quality can be

improved even more and the parameter adjustment is totally negligible for the caller.

References

- [1] B. Duysburgh, S. Vanhastel, B. De Breese, C. Petrisor, P. Demeester, On the influence of best-effort network conditions on the perceived speech quality of VoIP connections, in: Proceedings of the 10th International Conference on Computer Communications and Networks, 2001, pp. 334–339.
- [2] J. Janssen, D. De Vleeschauwer, G.H. Petit, Assessing voice quality in packet-based telephony, *IEEE Internet Comput.* May–June (2002) 48–56.
- [3] J.-S. Kang, Y.-D. Kang, I.-H. Na, Y.-S. Choi, J.-H. Kim, A study of subjective quality measurement over VoIP network, in: Proceedings of IEEE International Conference on Info-Tech and Info-Net, 2001, pp. 311–316.
- [4] url: <http://www.internettrafficreport.com>, 13 Feb. 2003.
- [5] G. Held, *Voice Over Data Networks*, McGraw-Hill, New York, 1998, 238 pp.
- [6] F. Halsall, *Multimedia Communications*, Addison-Wesley/Pearson Education, Harlow, England, 2001, p. 1034.
- [7] D. Goodman, G. Lockhart, O. Wasseem, W.-C. Wong, Waveform substitution techniques recovering missing speech segments in packet voice communications, *IEEE Trans. Acoust. Speech Signal Process.* 34 (6) (1986) 1440–1448.
- [8] A. Husain, V. Cuperman, Classification and spectral extrapolation based packet reconstruction for low-delay speech coding, in: Proceedings of IEEE ICASSP, 1995, pp. 848–852.
- [9] C. Montminy, T. Aboulnasr, Improving the performance of ITU-T G.729A for VoIP, in: Proceedings of IEEE International Conference on Multimedia and Exposition, 2000, pp. 433–436.
- [10] T. Morinaga, S. Sasaki, K. Mano, T. Kaneko, Robust speech coding under packet-loss conditions using recovery sub-codec for broad-band IP network, in: Proceedings of IEEE ICASSP, 2002, pp. 2713–2716.
- [11] W. Jiang, A. Ortega, Multiple description of speech coding for robust communication over lossy packet networks, in: Proceedings of IEEE International Conference on Multimedia and Exposition, 2000, pp. 444–447.
- [12] J.-F. Wang, J.-C. Wang, J.-F. Yang, J.-J. Wang, A voicing-driven packet loss recovery algorithm for analysis-by-synthesis predictive speech coders over internet, *IEEE Trans. Multimedia* 3 (1) (2001) 98–107.
- [13] E. Gündüzhan, K. Momtahan, A linear prediction based packet loss concealment algorithm for PCM coded speech, *IEEE Trans. Speech Audio Process.* 9 (8) (2001) 778–785.
- [14] Y.-L. Chen, B.-S. Chen, Model-based multirate representation of speech signals and its application to recovery of missing speech packets, *IEEE Trans. Speech Audio Process.* 5 (3) (1997) 220–231.
- [15] J. Turunen, P. Loula, Assessment of various packet recovery mechanisms for G.728, in: Proceedings of ISCOM, 1999, pp. 57–61.
- [16] J. Turunen, P. Loula, Missing packet recovery method for G.723.1 based on neural networks, in: N. Mastroakis (Ed.), *Advances in Neural Networks and Applications*, World Scientific/Engineering Society Press, Pireus, Greece, 2001, pp. 372–377.
- [17] E. Ginters, V. Zagursky, Speech signal recovery in packet-switched communication networks, in: Proceedings of IEEE 10th Mediterranean Electrotechnical Conference, 2000, pp. 519–521.
- [18] J. Wang, J. Gibson, Parameter interpolation to enhance the packet erasure robustness of CELP coders in packet networks, in: Proceedings of IEEE ICASSP, 2001, pp. 745–748.
- [19] W.-T. Liao, J.-C. Chen, M.-S. Chen, Adaptive recovery techniques for real-time audio streams, in: Proceedings of IEEE Infocom, 2001, pp. 815–823.
- [20] H. Sanneck, *Packet Loss Recovery and Control for Voice Transmission over the Internet*, Dissertation, Technical University of Berlin, Berlin, 2000, 207 pp.
- [21] B. Wah, X. Su, D. Lin, A survey of error-concealment schemes for real-time audio and video transmissions over the Internet, in: Proceedings of International Symposium on Multimedia Software Engineering, 2000, pp. 17–24.
- [22] ITU-T P.862, Perceptual evaluation of speech quality (PESQ), an objective method for end-to-end speech quality assessment of narrow-band telephone networks and speech coders, ITU, Geneva, February 2001, 24 pp.
- [23] ITU-T G.723.1, Dual rate speech coder for multimedia communications transmitting at 5.3 and 6.3 kbit/s, ITU, Geneva, November 1996, 27 pp.
- [24] ITU-T G.729, Coding of speech at 8 kbit/s using conjugate-structure algebraic-code-excited linear-prediction, ITU, Geneva, March 1996, 35 pp.
- [25] ITU-T G.729 Annex A, Coding of speech at 8 kbit/s using conjugate-structure algebraic-code-excited linear-prediction, Annex A: reduced complexity 8 kbit/s CS-ACELP speech codec, ITU, Geneva, November 1996, 8 pp.
- [26] L. Huang, K. Sezaki, End-to-end Internet delay dynamics, in: Proceedings of Sixth Asia-Pacific Conference on Communications, TSR-15-2000 06, 2000, 6 pp.
- [27] N. Matsuo, M. Yuito, Y. Tokunaga, Packet interleaving for reducing speech quality degradation on packet voice communications, in: Proceedings of IEEE GLOBECOM, 1987, pp. 1787–1791.
- [28] P. Ferguson, G. Huston, *Quality of Service, Delivering QoS on the Internet and Corporate Networks*, Wiley, New York, 1998.

Jari Turunen received his MSc degree in 1998 and PhD degree in 2003 from Tampere University of Technology. His main research interests are in telecommunication and speech processing.

Pekka Loula received his MSc degree in 1987 and the PhD degree in 1994 from Tampere University of Technology. Nowadays, he holds a telecommunication professorship at Tampere University of Technology, Pori. He is the author of over 100 publications from the field of telemedicine, telecommunication and signal processing. His current research interests cover topics such as IP based networks, broadband telecommunication, QoS aspects and telecommunication applications.

Tarmo Lipping received his MSc degree in 1993 and the PhD degree in 2001 from Tampere University of Technology. Nowadays, he holds a signal processing professorship at Tampere University of Technology, Pori. His main research interests are in physiological signal processing, analysis and modeling.