BULLETIN OF THE POLISH ACADEMY OF SCIENCES
TECHNICAL SCIENCES, Vol. 62, No. 3, 2014
DOI: 10.2478/bpasts-2014-0045

The new, parameterised VT model for determining quality in the video-telephony service

T. UHL[1]* and S. PAULSEN[2]

[1] Institute of Transport Engineering, Maritime University of Szczecin, 11 Henryka Pobożnego St. ,70-507 Szczecin, Poland
[2] Institute of Communications Technology, Flensburg University of Applied Sciences, Kanzleistraße 91-93, D 24943 Flensburg, Germany

Abstract. This paper describes a new method for measuring and appraising the quality of the video stream in the video-telephony service in IP environments (VToIP). The method uses the so-called VS Model and belongs to the group of parameter-based measurement techniques (offline operation, i.e. without active measurement). It delivers results that come extremely close to those contained in the corresponding Perceptual Evaluation of Video Quality (PEVQ) curves. The method is quick and easy to use – a convincing argument for using this new QoS measuring method.

Key words: communication networks, communication services, IP transport platform, QoS measurement techniques, VToIP, G.1070, PEVQ.

1. Introduction

Quality of Service (QoS) plays a very important role in modern digital networks. The term is becoming a household word and can be found among other things in the definition of Next Generation Networks according to the ITU-T Standard Y.2001 [1]. In 2009 the European Parliament and European Council published directives for the standardisation of networks and services [2, 3], placing great priority on quality of service.

The QoS in modern networks should be measured continuously - preferably automatically. This makes specialised measurement systems and methods indispensable. There are, however, hardly any standardised QoS measuring methods for video-communications applications such as video-telephony. At present there are only two standards: ITU-T Rec. J.144 [4] and ITU-T G.1070 [5] to resort to. However, the QoS measuring methods mentioned so far are very complex and the licences expensive. Until now, there simply has been no simple, parameterised QoS measuring techniques that are quick and easy to use. The VT Model (VT = Video-telephony) described in this paper aims to close that gap.

The VToIP service operates according to Recommendation H.323 [6]. This Recommendation defines the encoding of audio and video signals. Codecs H.263, H.263+ and H.263++ are provided for video streaming. A VToIP connection can be established, controlled and terminated using a range of signalling protocols. In practice, however, the SIP [7, 8] protocol is by far the most widely used. Real-life measurements of the VToIP service have revealed that a refresh rate of 25 frames per second is widely used. Common formats for the service are: CIF, QCIF, QVGA and QQVGA. In order to be of practical value, this paper takes all these observations into consideration.

To begin with, the main features of video-telephonic communication in an IP environment will be explained. There will then be a brief introduction to the G.1070 Standard. The next chapter contains a formulation of the VT Model and goes on to describe the aspects to be considered when the Model is implemented. The usefulness of the new model will be examined and its efficiency tested in near-life scenarios. The results gained from these analyses will then be represented graphically, and interpreted. The work concludes with a summary and an outlook on future areas of work.

2. Main features of VToIP communication

In order to determine the main features of VToIP communication the system depicted in Fig. 1 was set up in the Triple Play Services Lab at the Flensburg University of Applied Sciences. The system consists of two VoIP clients between which the VoIP communication (including video-telephony) is supported by the 3CX IP-PBX/VoIP-Server [9]. In this configuration the system is also capable of remote communication using the VoIP provider SipGate [10]. After logon, the signalling is performed by a proxy server of the VoIP provider.

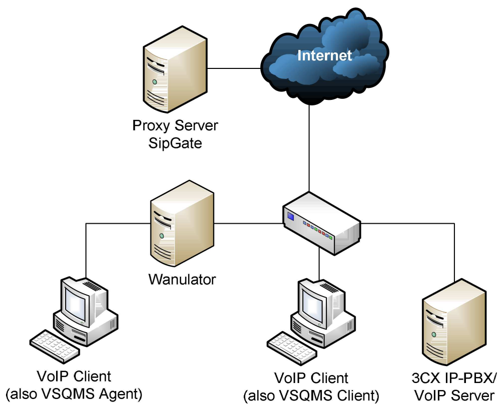

Fig. 1. Structure of the VoIP Systems in the Triple Play Services Lab

*e-mail: t.uhl@am.szczecin.pl

The video and audio data are exchanged directly between the participating stations in the lab using the redirect mode. The wanulator [11] ensures that the technician can realistically replicate and regulate the desired impairments. Trafficlyser®measurement systems from the company Nextragen [12] (a cooperation partner of the Flensburg University of Applied Sciences) are implemented for the purposes of protocol analysis in the lab. They have been installed on the VoIP clients.

The measurements that were made can quite suitably be represented graphically. The diagram in Fig. 2 shows clearly the chronological sequences of the VToIP connection as three distinct phases. The first phase includes the signalling during connection establishment. As soon as the user is authenticated, four logical channels are opened (2 for voice and 2 for video) and the parameters for the RTP sessions (type of codec, bandwidth, encoding rate, etc.) are negotiated. In the second phase the user data are exchanged (encapsulated in RTP/UDP/IP). The third phase includes the signalling during connection termination.

The diagram in Fig. 2 underlines the conformity of protocol handling when SIP is used as the signalling protocol.

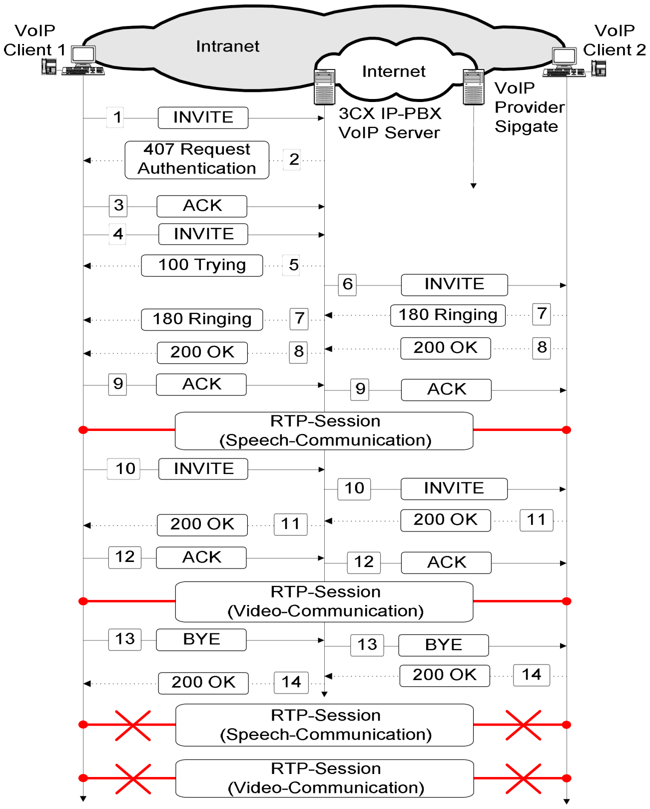

Fig. 2. Diagram of protocol handling in VToIP

A great many measurements (protocol application flow, network loading, determination of traffic characteristics, to name but a few) were made using the video telephony system described above and the measurement tool Trafficlyser®from the company Nextragen [12], and they led to the following assertions:

a) VToIP communication uses a unicast connection in full-duplex operation.
b) The traffic in VToIP communication features the CBR character.
c) Four logical channels per connection are used.
d) A voice connection with the G.711 codec uses approx. 160 kbps of the bandwidth.
e) A video connection with the H.263 codec (CIF with 25 images/s) uses approx. 200 kbps of the bandwidth.
f) Using the G.711 codec, the packet size for audio is approx. 214 bytes, of which 160 bytes are used for the data field (20 ms speech samples).
g) Using the H.263 codec, a transport packet for video is always 1,400 bytes long.
h) An increase in the efficiency of VToIP transmission can only be achieved if the capacity of the compensation buffer in the VToIP terminal is sized accordingly (in practice approx. 0.3–0.5 seconds).

This knowledge will be used to the full when the new model for evaluating QoS in the VToIP service is designed. The following chapter begins with a brief description of ITU-T's Standard G.1070.

3. The ITU-T recommendation G.1070

Figure 3 shows the framework of Recommendation G.1070 [5]. Its input parameters are video and speech quality parameters that are considered important in QoE/QoS planning. The model consists of three functions: video quality estimation, speech quality estimation, and multimedia quality integration functions. The degradation caused by delay alone is considered only in the multimedia quality integration function.

In the branch of the diagram representing audio quality estimation the G.1070 Model uses the well-known E Model (ITU-T G.107 [13]). It is common knowledge, and indeed has been proved in [14], that the E Model is only moderately suited to use on IP transport platforms. A number of changes will have to be made either to the E Model or to any other alternative models before they are fully adapted to IP-based environments.

In the branch of the diagram that represents video quality estimation the G.1070 Model uses its own parameterised model. Unfortunately, this model has been designed for only a limited number of codecs and image formats. It is in great need of updating to be capable of dealing with the latest video transmission types. It could prove wiser to switch to new, parameterised models that have been specifically designed to determine video quality in IP environments.

For the purposes of this study the sectors in which the audio and video qualities are determined will be replaced by our own parameterised QoS models. This is the essence of the new, parameterised VT Model for determining quality in the video-telephony service over IP (VToIP).

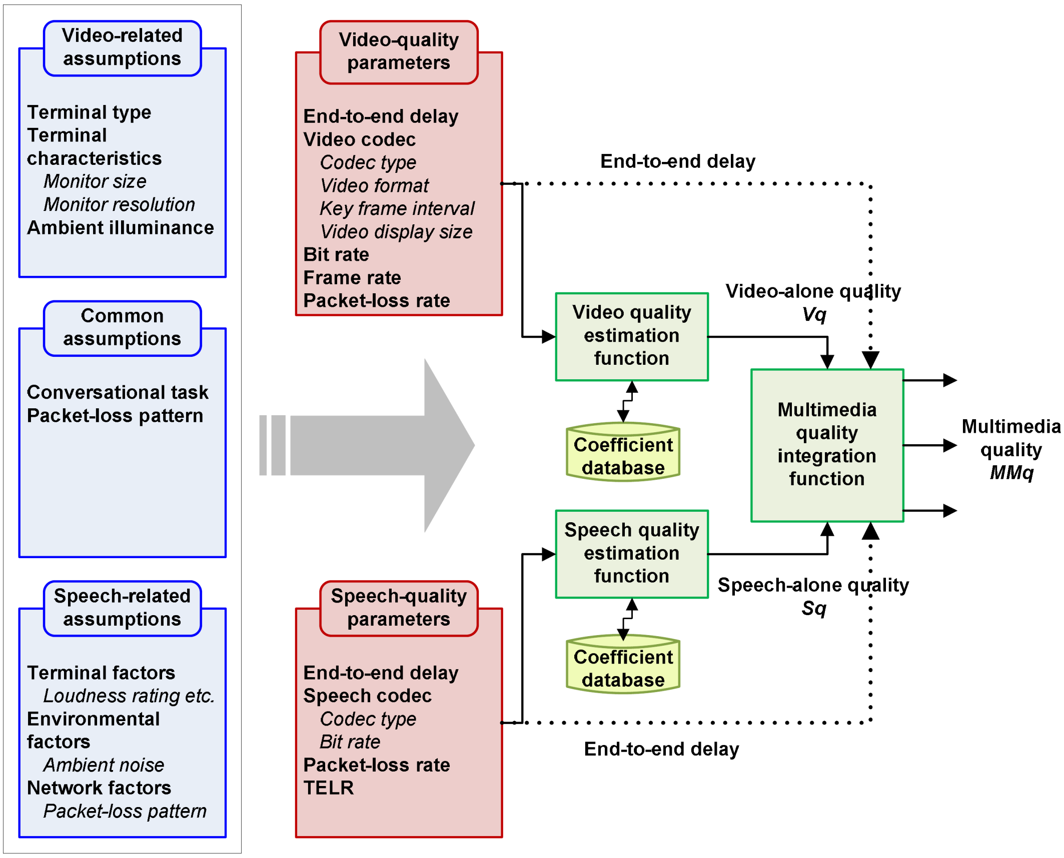

Fig. 3. G.1070 Framework after Ref. 15

4. The new, parameterised VT Model

The structure of the new, parameterised VT Model for determining the QoS of video-telephony is based on ITU-T Recommendation G.1070. Thus in contrast to the G.1070 Model the quality of video and audio streams is determined on the basis of our own parameterised models. The A Model [15] is used for audio and the VS Model [16] for video. Both models are described briefly in the following and explained in the light of typical VToIP codecs, i.e. G.711 [17] for audio and H.263 [18] for video.

The A Model is based on PESQ (Perceptual Evaluation of Speech Quality [19]) curves that have been reliably calculated as a function of packet losses and burst size for speech samples of various lengths and different audio codecs. In general, the A Model can be characterised by the following equation [15]:

$$A - Factor = A_1 e^{-0.015\alpha} + A_2 e^{-0.45\alpha} + A_3 \quad \text{[MOS-LQO]}. \tag{1}$$

The constant A_1 is responsible for the shape of the QoS curve when packet losses are high and the constant A_2 when they are low. Using constant A_3 ensures that the maximum QoS value for the speech codec under analysis (e.g. the value 4.4 MOS for G.711) is preserved. The coefficients -0.015 and -0.45 in the exponents of Eq. (1) were calculated to be optimum values. The variable α can be determined as a function of packet losses (in practice it can be determined with an appropriate measuring system) and of the *K-factor*, using the following equation:

$$\alpha = \frac{Packetloss}{K - Factor}. \tag{2}$$

The *K-factor* is generated by the linear approximation of the normed PESQ values (in relation to the curve for burst size equal to "1" and the smallest speech sample length) as a function of the BSLP (burst size sample length product; this parameter too is determined in practice by a suitable measuring system) and packet losses.

$$K - Factor = a \cdot \beta \cdot BSLP + b \tag{3}$$

All three coefficients from Eq. (3), i.e. a, b and β, depend on the speech codec under analysis. Values a and b are coefficients that are calculated by linear approximation. Using the scale parameter β ensures that with the aid of the *K-factor* the QoS curves can be emulated equally well for both small and large BSLP values.

So for the audio codec G.711, that is typically used in VToIP, the general equations assume the following concrete forms:

$$A - Factor = 9.5e^{-0.015\alpha} + 0.7e^{-0.45\alpha} - 5.8, \tag{4}$$

$$K - Factor = 0.0011 \cdot 2.5 \cdot BSLP + 0.9086. \tag{5}$$

The A Model was developed along the lines of ITU-T's Recommendation P.564 [20] with the result that *A-factor* values correspond to the MOS-LQO scale (Mean Opinion Score-Listening Quality Objective).

The VS Model is based on PEVQ (Perceptual Evaluation of Video Quality [21]) curves which have been calculated as functions of packet losses and burst sizes for various image formats, encoding rates and video codecs accordingly. Going off these curves, the first step is to approximate the PEVQ curves over packet losses for burst sizes "1" to "5" and selected encoding rates with the aid of Eq. (6) (See [16] for details):

$$VS - Factor = P \cdot e^{\frac{a \cdot Packetloss}{Burstsize}} + Q \cdot e^{\frac{b \cdot Packetloss}{Burstsize}} \quad \text{[MOS]}. \tag{6}$$

The constants a and b are selected so that they display values equal to or smaller than 0, with significantly smaller values being chosen for b. The result of this is that the two summands are responsible for the steepness of the curves when packet loss is small (second summand) or large (first summand). All constants (P, Q, a, and b) are now calculated iteratively as best possible values for each encoding rate.

The next step is to determine the equations for the constants P, Q, a and b. Once calculated, the constants are recorded as functions of the encoding rate. Their corresponding equations can then be calculated by means of polynomial approximation. The degree of the polynomial is caused by the complexity of the curve. Eqs. (7) to (10) show the simple relationship:

$$P = w_n \cdot Bitrate^n + w_{n-1} \cdot Bitrate^{n-1} + \ldots + w_1 \cdot Bitrate^1 + w_0, \tag{7}$$

$$Q = z_n \cdot Bitrate^n + z_{n-1} \cdot Bitrate^{n-1} + \ldots + z_1 \cdot Bitrate^1 + z_0, \tag{8}$$

$$a = x_n \cdot Bitrate^n + x_{n-1} \cdot Bitrate^{n-1} + \ldots + x_1 \cdot Bitrate^1 + x_0, \tag{9}$$

$$b = y_n \cdot Bitrate^n + y_{n-1} \cdot Bitrate^{n-1} + \ldots + y_1 \cdot Bitrate^1 + y_0. \tag{10}$$

For the video codec H.263 with 25 images per second and CIF format that is typically used for VToIP the general equations assume the following concrete form:

$$P = 3.54 \cdot 10^{-8} \cdot Bitrate^2 - 3.45 \cdot 10^{-4} \cdot Bitrate + 2.39, \tag{11}$$

$$Q = -7.02 \cdot 10^{-15} \cdot Bitrate^4 + 1.36 \cdot 10^{-10} \cdot Bitrate^3 - 9.66 \cdot 10^{-7} \cdot Bitrate^2 + 3.02 \cdot 10^{-3} \cdot Bitrate - 0.51, \tag{12}$$

$$a = -7.00 \cdot 10^{-10} \cdot Bitrate^2 + 8.00 \cdot 10^{-6} \cdot Bitrate - 2.39 \cdot 10^{-2}, \tag{13}$$

$$b = 3.68 \cdot 10^{-11} \cdot Bitrate^3 - 5.23 \cdot 10^{-7} \cdot Bitrate^2 + 1.94 \cdot 10^{-3} \cdot Bitrate - 2.80. \tag{14}$$

The last step is to determine the multimedia quality *MMq* (cf. the last block of the model in Fig. 3). The multimedia quality *MMq* is calculated using speech quality *Sq* = *A-factor*, video quality *Vq* = *VS-factor*, speech delay T_S and Video delay T_V. *MMq* is expressed as:

$$MM_q = m_1 MM_{SV} + m_2 MM_T + m_3 MM_{SV} MM_T + m_4, \tag{15}$$

where MM_{SV} represents audiovisual quality, MM_T represents the audiovisual delay impairment factor, and coefficients $m_1 - m_4$ are dependent on display size and conversational task. *MMq* values lie between 1 and 5.

$$MM_{SV} = m_5 S_q + m_6 V_q + m_7 S_q V_q + m_8, \tag{16}$$

$$MM_T = \max\{AD + MS,\ 1\}, \tag{17}$$

$$AD = m_9 (T_S + T_V) + m_{10}, \tag{18}$$

$$MS = \min\{m_{11}(T_S - T_V) + m_{12},\ 0\} \quad \text{if } T_S \geq T_V, \tag{19}$$

$$MS = \min\{m_{13}(T_V - T_S) + m_{14},\ 0\} \quad \text{if } T_S < T_V. \tag{20}$$

Coefficients $m_5 - m_{14}$ are also dependent on video display size and conversational task (see [5] for details).

Furthermore, multimedia quality will be determined using our own, innovative method. Equation (21) shows the mathematical relations involved.

$$VT - Factor = \max\{[\alpha S_q + (1-\alpha) V_q]\Psi \cdot \Phi;\ 1\}. \tag{21}$$

The values in this equation are to be interpreted as follows. The factor α determines the weighting function of the QoS values S_q to V_q. In practice it seems to make sense to set the value of α at 0.5, but other weight factors are conceivable.

The function Ψ represents the deterioration of the multimedia quality as a function of total delay, i.e. T_s plus T_V. This is a linear dependency (see formula (22)) and was calculated from the function *AD* using formula (18).

$$\Psi = \max[a_1 (T_S + T_V) + b_1;\ 0]. \tag{22}$$

The magnitude Φ represents the deterioration of the multimedia quality as a function of the asynchrony between the audio stream and the video stream. This dependency is also linear, see formulas (23) and (24), and was calculated from the function *MS* using Eqs. (19) and (20).

$$\Phi = \max[a_2 (T_S - T_V) + b_2;\ 0] \quad \text{if } T_S \geq T_V, \tag{23}$$

$$\Phi = \max[a_3 (T_V - T_S) + b_3;\ 0] \quad \text{if } T_S < T_V. \tag{24}$$

The coefficients for both linear equations Ψ and Φ (a_1, a_2, a_3, b_1, b_2, and b_3) are derived from the equivalent G.1070 equations for *AD* and *MS*. The values for both equations are normalised to a value range between "1" and "0" in order to adapt these coefficients for use in the VT model. Hence, conditions without a delay will result in a value of "1" having no influence on the actual QoS value, see formula (21). Table 1 shows the concrete coefficients for formulas (22) to (24), calculated for two display sizes (on the basis of the compatibility with G.1070).

Table 1
Coefficients for the formulas (2)–(24)

	Display 2.1 inch	Display 4.2 inch
α	0.5	0.5
a_1	-3.32E-05	-8.26E-05
b_1	1	1
a_2	-1.07E-03	-1.38E-03
b_2	1	1
a_3	-1.00E-03	-1.10E-03
b_3	1	1

5. Comparison study

Figure 1 shows the realistic analysis environment that was used for the comparison study. The Video Streaming Quality Measurement System (VSQMS), that was described in detail in paper [22], is the mainstay of the environment. With a few additions, the measurement system is capable of establishing two RTP connections between two VToIP stations and using them to transmit H.263-encoded video streams and G.711-encoded audio streams. Der VSQMS client functions here as a "mirror", that sends the incoming RTP packets back to the VSQMS agent. The agent implements both of the new QoS models, i.e. the A Model and the VS Model. These two QoS measuring methods, the equations for the multimedia quality MMq from Recommendation G.1070 and the equations for the new VT Model are all needed to calculate QoS values for the VToIP service.

The following parameters were assumed for the comparison study:

a) Nondeterministic distributed packet losses (properties of the tools [11]) of 0 to 20% and constant burst size "1".
b) Nondeterministic distributed delays T_S and T_V (properties of the tools [11]) with mean values of 0 ms and 500 ms respectively.
c) Nondeterministic distributed delays T_S (properties of the tools [11]) with a mean value of 150 ms and $T_V = 0$ ms.
d) Video codec H.263 with encoding rates of 1702 kbps and 4978 kbps.
e) Image format CIF (corresponds approximately to the resolution of the QVGA format).
f) Image refresh rate of 25 images/s.
g) Video display size = 4.2 inch.
h) Speech Codec G.711 with speech sample lengths of 20 ms.
i) 31 measurements per value of each of the variables used. This ensures that confidence intervals are achieved that are less than 10% of the mean values under analysis (with a probability of error of 5%).
j) Multimedia QoS values according to G.1070 (MM_q metric) and according and to the VT Model (*VT-factor*).

The results of the comparison study are presented graphically in Figs. 4 to 9. All the curves in Figs. 4 to 9 have an exponential character. Another thing that all figures have in common: using the metric MM_q from G.1070 always yields the poorest values for multimedia quality. This is quite unusual, and surprising. The individual S_q values and V_q values are frequently good, yet the resulting overall quality is very low. The slight rise in MM_q values when packet loss is high can be explained with Eq. (16). When there are large discrepancies between S_q and V_q, the first summand of the equation, i.e. $m_5 S_q$ (m_5 negative!), yanks the MM_q value down steeply. The ensuing decrease in this discrepancy weakens its influence, and a slight rise in MM_q values is the result. The VT Model delivers results that are far easier to understand, for the overall QoS always lies somewhere between the individual QoS values for S_q and V_q. In conclusion, it must be said that the metric MM_q from the G.1070 drastically underrates the quality of the service VToIP whereas the new VT Model seems perfectly suitable for the practical use.

Figures 4 and 5 show that in a loss-free environment higher encoding rates lead to an increase in the quality V_q. In a lossy environment, however, this advantage dwindles rapidly. Here it is much more beneficial to choose a lower encoding rate. The high encoding rate is only favourable in a loss-free environment.

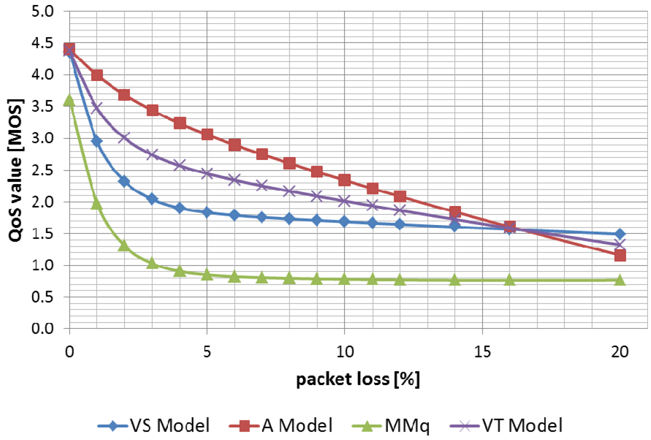

Fig. 4. QoS Values as a Function of Packet Losses, a Delay of 0 ms and an Encoding Rate of 1702 kbps

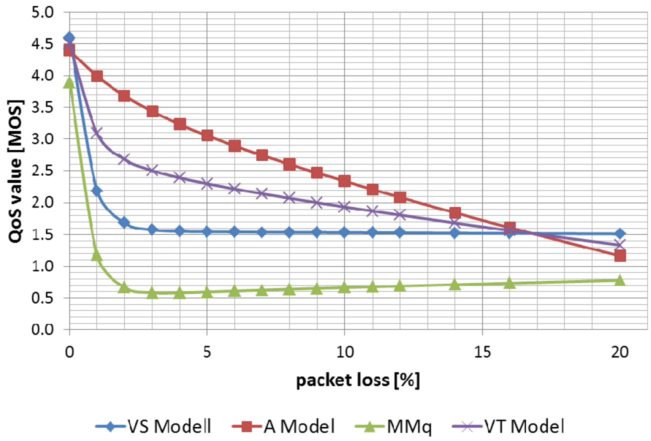

Fig. 5. QoS Values as a Function of Packet Losses, a Delay of 0 ms and an Encoding Rate of 4978 kbps

Figures 6 and 7 show clearly the relationships as they were described above. In addition, as total delay $(T_s + T_V)$ increases, multimedia quality in comparison to $(T_s + T_V) = 0$ declines as a matter of course.

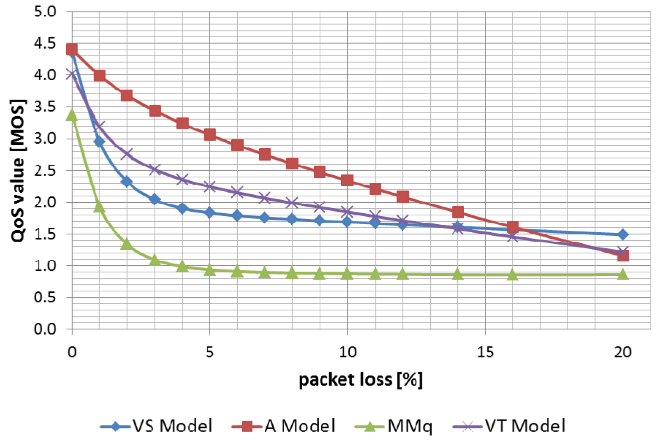

Fig. 6. QoS Values as a Function of Packet Losses and Delays T_S and T_V of 500 ms respectively and an Encoding Rate of 1702 kbps

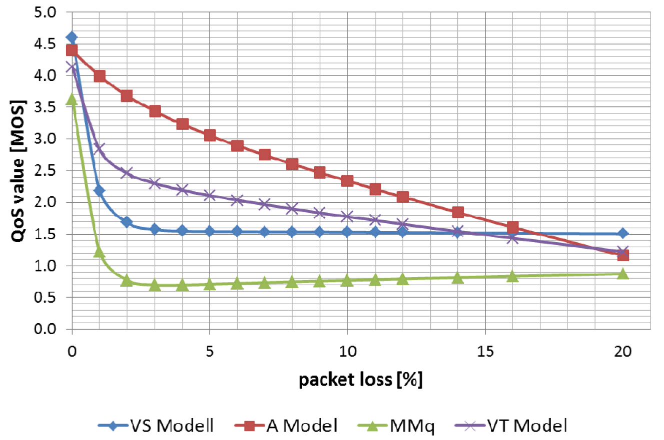

Fig. 7. QoS Values as a Function of Packet Loss and Delays T_S and T_V of 500 ms respectively and an Encoding Rate of 4978 kbps

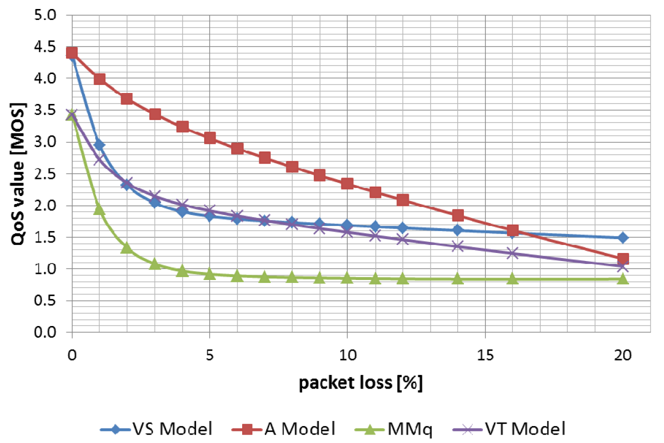

Fig. 8. QoS Values as a Function of Packet Loss and Delays of T_S =150 ms and T_V = 0 ms and an Encoding Rate of 1702 kbps

Figures 8 and 9 also show the dependency described above. It must, however, be pointed out that asynchrony has a far greater influence on multimedia quality than is the case with large values for T_s and T_V when synchrony is present. This seems reasonable enough since increases in asynchrony between audio and video will obviously lead to a reluctance to award good QoS values. Time lapses between the delivery of audio output and video output have a much greater influence on the user's perception of quality than a lengthy delay in the reaction of the dialogue partner.

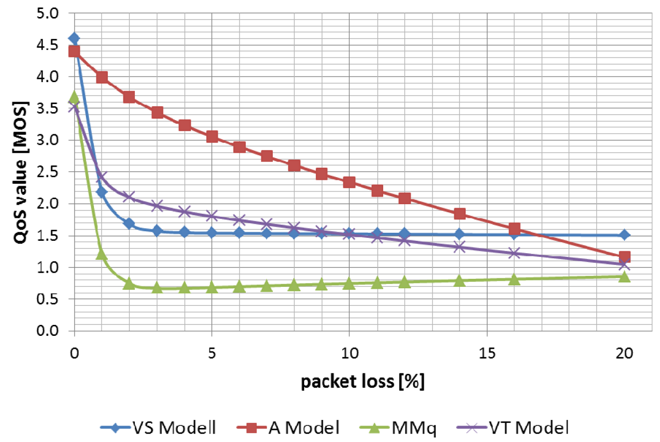

Fig. 9. QoS Values as a Function of Packet Loss and Delays of T_S = 150 ms and T_V = 0 ms and an Encoding Rate of 4978 kbps

6. Summary and outlook

In the course of the work described in this paper a new, parameterised model was developed to evaluate the quality of the video-telephony service, and its functionality was tested in an exhaustive study. All parameters associated with the VToIP services were taken into consideration when the model was defined. The new VT Model is based on the PESQ and PEVQ curves, which is of great practical significance since both of these algorithms are considered to be the most objective methods of measuring the QoS of both audio and video content. The comparison study proved unequivocally that the VT Model can be used in everyday practice. By contrast, the study laid doubt on the G.1070 Model. Consequently, the new VT Model is a simple and economical replacement for the G.1070 method.

It would be interesting to discover why the G.1070 Model delivers such poor values for multimedia quality as a function of packet losses. Since all of the parameters $m_1 - m_{14}$ in this model were derived from subjective studies, it would be necessary to construct a similar evaluation environment in which a similar study could be conducted with groups of test persons. Although studies of this kind notoriously consume a great deal of time and resources, it would constitute a possible area of research and the subject of a further study of the topic QoS in the VToIP service.

The final series of tests, that were run under practical conditions, showed that modern codecs such as G.711.1 for audio and H.264 for video are being implemented more and more frequently in the VToIP service. That is what prompted the authors to anchor these new codecs within the VT Model. New equivalents for S_q and V_q will have to be calculated for the new codecs though. Studies have already begun at the Maritime University of Szczecin and at the Flensburg University of Applied Sciences to achieve that goal.

REFERENCES

[1] *Definition von NGN*, http://www.itu.int/rec/T-REC-Y.2001 (May 2014).

[2] *Directives 2009/140/WE for the Creation of Engineering Standards in Networks and Electronic Services*, http://eurlex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2009:337:0037:0069:PL:PDF (May 2014).

[3] *Directives 2009/136/ WE for the Creation of Engineering Standards in Networks and Electronic Services*, http://eurlex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2009:337:0011:0036:PL:PDF (May 2014).

[4] *ITU-T Recommendation J.*144, http://www.itu.int/rec/T-REC-J.144/en (May 2014).

[5] *ITU-T Recommendation G.1070*, http://www.itu.int/rec/T-REC-G.1070/en (May 2014).

[6] *ITU-T H.323 Protocol Suite*, http://www.openh323.org/standards.html (May 2014).

[7] *IETF SIP Protocol Suite*, http://www.ietf.org/rfc/rfc3261.txt (May 2014).

[8] P.S. Gutkowski and S. Kaczmarek, "The model of end-to-end call setup time calculation for Session Initiation Protocol", *Bull. Pol. Ac.: Tech.* 60 (1), 95–101 (2012).

[9] *3CX IP-PBX/VoIP Server*, http://www.sipcenter.com/files/sip2.pdf (May 2014).

[10] *The Service Provider "Sipgate"*, http://www.sipgate.de (May 2014).

[11] *Wanulator*, http://wanulator.de (April 2014).

[12] *Trafficlyser® Suite*, http://www.nextragen.de (May 2014).

[13] *Recommendation G.107*, http://www.itu.int/rec/T-REC-G.107/en (May 2014).

[14] S. Paulsen and T. Uhl, "Adjustments for QoS of VoIP in the E-Model", *Proc. World Telecommunications Congress* 1, 69–74 (2010).

[15] S. Paulsen and T. Uhl, "Das neue, parametrisierte A-Modell für QoS beim Dienst VoIP" (in Germany), *J. Praxis Informationsverarbeitung und Kommunikation* 34 (1), 23–30 (2011).

[16] S. Paulsen and T. Uhl, "The new, parameterized VS Model for determining the quality of video streams in the video-telephony Service", *J. Telecom. Review.* 8–9, 1155–1166 (2012).

[17] *Codec G.711*, http://www.itu.int/rec/T-REC-G.711.0/en (May 2014).

[18] *Recommendation H.263*, http://www.itu.int/rec/T-REC-H.263/en (May 2014).

[19] *Recommendation P.862 (PESQ)*, http://www.itu.int/rec/T-REC-P.862/en (May 2014).

[20] *Recommendation P.564*, http://www.itu.int/rec/T-REC-P.564/en (May 2014).

[21] *Recommendation J.247 (PEVQ)*, http://www.itu.int/rec/T-REC-J.247/en (May 2014).

[22] T. Uhl and Ch. Hoppe, "Neues Messsystem zur Beurteilung der Qualität des Videostroms beim Dienst Video Telephony in IP-Umgebung", *Electronic Communications EASST* 37, 197–207 (2011), (in German).