

RESEARCH

Open Access



New approach for determining the QoS of MP3-coded voice signals in IP networks

Tadeus Uhl^{1,2*} , Stefan Paulsen³ and Krzysztof Nowicki⁴

Abstract

Present-day IP transport platforms being what they are, it will never be possible to rule out conflicts between the available services. The logical consequence of this assertion is the inevitable conclusion that the quality of service (QoS) must always be quantifiable no matter what. This paper focuses on one method to determine QoS. It defines an innovative, simple model that can evaluate the QoS of MP3-coded voice data transported through an IP environment. It describes tests of the model's practicability that were conducted in a comprehensive comparison study. The so-called MP3 Model is one of a number of parameter-based measuring techniques and delivers results that come very close to the corresponding perceptual evaluation of speech quality (PESQ) curves. This is one of the features that make this new QoS measuring method so attractive.

Keywords: Communication networks, Communication services, IP transport platform, QoE/QoS measurement techniques, MP3 encoder, PESQ

1 Introduction

Quality of service (QoS) and quality of experience (QoE) play a very important role in modern digital networks. The terms are becoming household phrases and can be found among other things in the definition of next-generation network according to the ITU-T Standard Y.2001 [1]. November 25, 2009, the European Parliament and the European Council adopted the so-called Communications Package that includes Directive 2009/136/EC [2] and Directive 2009/140/EC [3] that both place great value on quality of service.

The quality of service in modern networks should be measured continuously—preferably automatically—which makes specialised measurement systems and methods indispensable. They are generally classified into two groups: signal-based and parameter-based measuring methods [4]. In the dual-ended model, two signals are used: original signal and degraded signal. These two signals are available uncompressed, allowing measurements to be carried out for both quality of experience (subjective evaluation) and quality of service (objective

evaluation). In the single-ended model, only the impaired signal (often compressed) is available. This allows only an objective evaluation (QoS) to be made. QoS measurement is referred to as “intrusive measurement” (online) in the case of the dual-ended model, and as “non-intrusive measurement” (offline) in the case of the single-ended model.

When it comes to the audio-type service (voice, music), the PESQ algorithm [5] and the PEAQ algorithm [6] are very often mentioned in connection with the dual-ended model. It is a time-consuming (active measurement) and expensive (licence) technique. It does, however, yield objective and extremely accurate results. The PESQ algorithm can be classed as a QoE technique inasmuch as it comprises a model of the human auditory system. The model was developed on the basis of extensive tests conducted on real test persons. That is why this paper has classified PESQ values as QoE values and uses them as a reference. This saves a great deal of time and trouble when it comes to developing new QoS models.

A further signal-based QoS method suitable for audio applications using the MP3 encoder (MPEG-1 layer III) is to be found in paper [7]. The technique is called the “audio watermarking procedure”. Specifically, a fragile watermark is embedded in an MP3-type host data audio transport stream using a spread-spectrum approach. At the receiver, the watermark is extracted and compared

* Correspondence: tadeus.uhl@hs-flensburg.de; t.uhl@am.szczecin.pl

¹Flensburg University of Applied Sciences, Kanzlei St. 91-93, 24943 Flensburg, Germany

²Maritime University of Szczecin, H. Pobożnego St. 11, 70-507 Szczecin, Poland

Full list of author information is available at the end of the article

Table 1 MPEG Audio Layer III versions [19]

Version	International standard	Supported sampling rate frequencies
MPEG-1 Audio Layer III	ISO/IEC 11172-3 (MPEG-1 Part 3)	32,000 Hz/44,100 Hz/48,000 Hz
MPEG-2 Audio Layer III	ISO/IEC 13818-3 (MPEG-2 Part 3)	16,000 Hz/22,050 Hz/24,000 Hz
MPEG-2.5 Audio Layer III	Nonstandard, proprietary	80,00 Hz/11,025 Hz/12,000 Hz

with its original counterpart. Since the alterations undergone by the watermark can be expected to be those experienced by the MP3 file, since they have followed the same communication link (including coder and transport connection), the watermark's degradation can be used to estimate the overall alterations experienced by the MP3 data. The QoS assessment is based on the evaluation of both objective (mean square error) and subjective (International Telecommunication Union Standard for Objective Measurement of Audio Quality) indicators. It is also a complex, elaborate and time-consuming QoS-measuring technique.

Quick and easy QoS methods that do not require a reference signal are indispensable for routine practical tests; however, they do require parametrised QoS models. The E Model is the best-known QoS measuring technique in voice applications [8]. The E Model was originally developed for circuit-switched voice networks and does not really come up to scratch in IP environments [9]. Since its inception, the E Model has undergone further development to take account of broadband codecs, which have regained importance due to modern VoIP applications. That is why the new E(IP) Model was created [10]. Neither the E Model nor the E(IP) Model can handle the MP3 codec. This paper will include a presentation of the development of one such method of measuring the QoS of the service voice streaming with MP3-coded data in an IP environment.

To begin with, the MP3 codec will be presented. Then, the numerical analysis environment used for this paper will be described. The paper goes on to present the results of quality of service evaluations gained in the

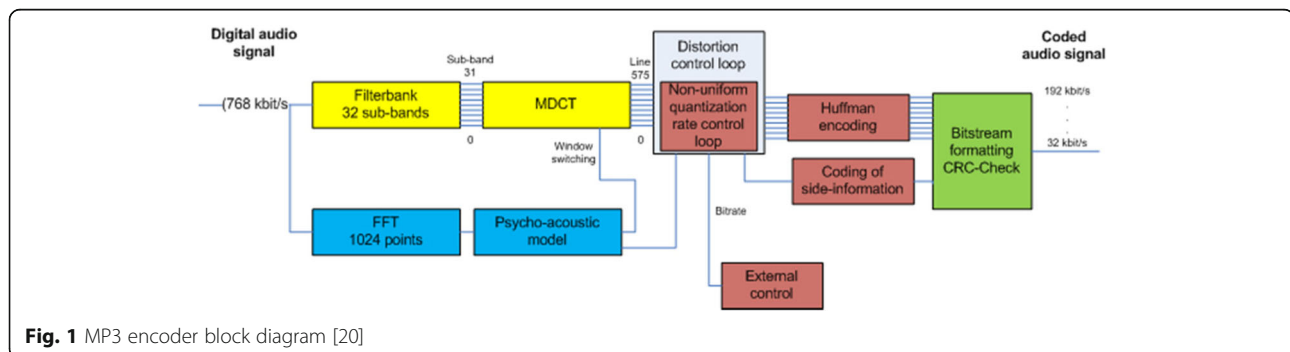
environment for MP3-coded voice streams. Following that, the new MP3 Model for voice streams in the IP environment will be formulated. After that, the practicability of the MP3 Model will be examined and its suitability for practical use will be put to test in a comparison study. The results gained from the study will be represented graphically and interpreted. The paper concludes with a summary and an outlook on future work.

2 The MP3 codec

The MP3 audio codec (MPEG Audio Layer III) developed in Germany by the Fraunhofer Institute is a standard for lossy audio compression and is a part of the Motion Pictures Expert Group (MPEG) standards [11]. MPEG specifies three classes of audio compression known as layer I, II and III, a higher number meaning an improved coding efficiency syntax and increased coding complexity. The compression is based on psycho-acoustic models in which the precision of audio components that are less audible to the human ear are reduced. Table 1 shows all MP3 versions: as part of the MPEG-1 standard, as part of its successor, MPEG-2, and as a proprietary solution (MPEG-2.5). Wideband reference signals (16 kHz sampling rate) and the MPEG-2 Audio Layer III compression were used throughout this paper.

The encoding process consists of the following four parts (see Fig. 1): First, the audio signal is divided into frames (1152 samples each), filtered and transformed by a modified discrete cosine transform (MDCT). The FFT and psycho-acoustic blocks comprise the calculation of the psychoacoustic model. Now, the signal is quantised and coded using the Huffman algorithm. Finally, the output stream is formatted, and a CRC error detection and correction is added, if necessary.

The MP3 codec is now the most frequently used codec for audio streaming over IP networks. Development of modern Video Telephony over IP (VToIP) systems strives for a considerable improvement of quality for the transmitted voice signals. This in particular can be achieved through deployment of broadband-voice codecs. Is the MP3 codec suitable for this? This is discussed in detail in this paper.

**Fig. 1** MP3 encoder block diagram [20]

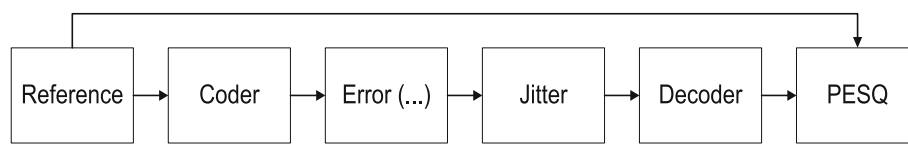


Fig. 2 Block diagram of the numerical software measurement tool

3 The analysis environment

Analyses of QoS and QoE can be conducted in two ways: (a) in real environments and (b) in emulated environments. The first case involves making measurements in the networks while the applications under analysis are actually being implemented. It is a very elaborate procedure and it seldom produces repeatable results. Besides, it is very difficult to create a particular measurement scenario using this method because regrettably, a lot must be left to chance in a real environment. The second method is more suitable for such analyses. It is possible to duplicate a defined measurement scenario and to obtain statistically meaningful results. In emulated environments, on the other hand, the impairment parameters can be implemented according to well-defined distributions. This proves to be of great benefit in research work. That is why a numerically emulated IP environment was used for the course of this study.

Figure 2 shows the block diagram of the numerical tool used here to determine the quality of service delivered by MP3-encoded audio streams in the IP environment. The detailed description and an operation of the tool can be found in the work [12].

The following parameters of the tool are assumed for the analyses here:

- Nondeterministic (binominal distribution) distributed packet loss rate from 0 to 10%.
- Nondeterministic (exponential distribution) distributed burst size with mean values from 1 to 5.
- Signals: “p564_speech_ukf1_wb.bin”, “p564_speech_spm2_wb.bin”, “p564_speech_jpf2_wb.bin” and “p564_speech_dem1_wb.bin” as the 16-kHz (wideband) reference signals (according to Rec. P.564 [13]).
- MP3 audio codec in the mode MPEG-2 Audio Layer III with an encoding rate of 64, 80 and 96 kbps and 1, 2 and 3 frame(s) per RTP packet [14].
- Thirty-one measurements for each determined performance value. This number of tests ensures a confidence interval of less than 10% of the estimated average (with a probability of error of 5%).
- PESQ in MOS(LQO) scale (according to the Rec. P.564) as resulting QoE values.
- “Silence insertion” method in cases of packet loss in accordance with ITU-T G.113 [15].

The following sections present some typical examples of the measurement results gained from the numerical measurement environment that was used. Here, it is the MP3 audio codec with an encoding rate von 80 kbps (see Figs. 3 and 4).

The curves in Figs. 3 and 4 show almost linear progressions with somewhat higher values for larger burst sizes. With a 3 frame pro RTP packet, the PESQ values decrease slightly more slowly than with

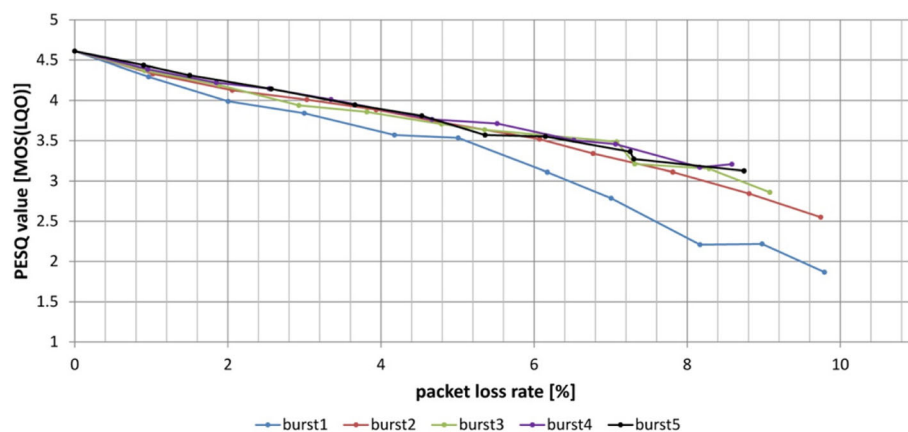


Fig. 3 PESQ values as a function of packet loss rate and burst size with a 1 frame pro RTP packet for the codec MP3 (80 kbps)

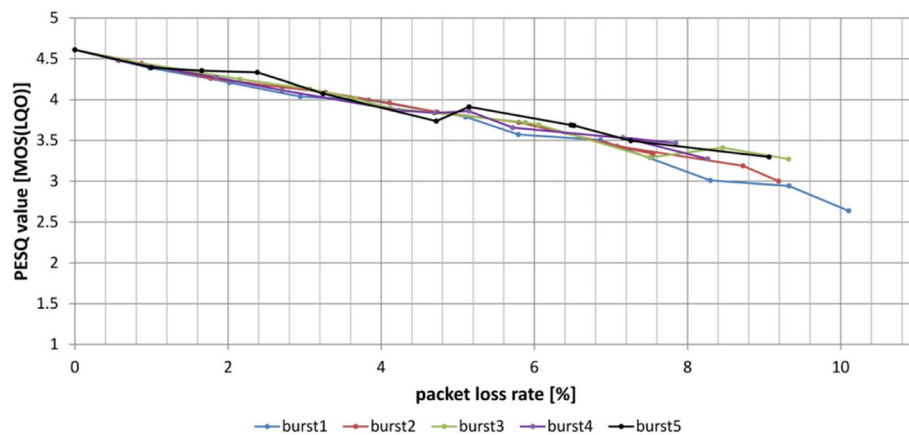


Fig. 4 PESQ values as a function of packet loss rate and burst size with a 3 frame pro RTP packet for the codec MP3 (80 kbps)

a 1 frame pro RTP packet. It becomes apparent that both increasing burst size and increasing frame numbers have a positive influence on the speech quality. Furthermore, it is noticeable that with a 1 frame pro RTP packet and a burst size of 1, the quality is considerably poorer than in other cases.

In addition, it can be seen that PESQ values for the same product of the burst size and the number of frames per RTP packet are comparable. The product will be referred to here as burst frame product (BFP). This paper contains the first printed mention of this variable. It is very worthwhile using the new parameter BFP during the evaluation of the PESQ values, as the work described in this paper has done.

4 The new, parametrised MP3 Model for determining QoS in voice streaming

One of the most influential impairment parameters or voice streaming (VoIP, VToIP) over the IP transport platform is packet loss. End-to-end delay is redressed in terminal equipment through the use of echo cancellers. Minor jitter and minimum out-of-order packet deliveries

can also be levelled out by implementing a jitter buffer in the terminal equipment.

The PESQ curves as a function of packet loss rate display a distinctly linear character (see Figs. 3 and 4). Other parameters that affect PESQ values include number of frames per RTP packet, encoding rate and burst size (average number of consecutively lost packets in the event of random loss). The size of the jitter buffers in the terminal equipment has a significant influence on quality of service as well. As was pointed out above, minor jitter and some out-of-order packet sequences can be smoothed out in the jitter buffer. If these impairments are too great in relation to the size of the jitter buffer, however, additional losses will occur, and the model will register them. All of these factors must be considered when creating a new, parametrised model for determining QoS in VoIP. Figure 5 shows the parametrised MP3 Model for determining the quality of audio streams with the MP3-encoded data in the IP environment.

In practice, it is assumed that when the QoS is determined, the packet streams from an RTP session are collected by a protocol analyser and then passed on to a suitable evaluation tool. The new MP3 Model is such a tool. It works on the following principles: all network

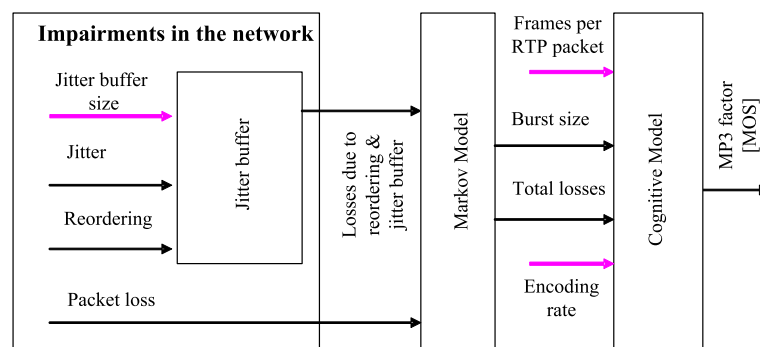


Fig. 5 Block diagram for the MP3 Model

impairments are collected and processed in the first block of the diagram. The effects of jitter and out-of-order packet delivery are converted into losses, not forgetting that some errors can be offset with the aid of the jitter buffer. The values attained in this block and the packet losses from the network are passed on to the second block where total losses and burst size are determined. For the MP3 Model, the Markov property “memorylessness”, which is widely used in analyses of networks, has been assumed. Going off this assumption, it is possible to determine the likelihood of a packet being lost, depending on whether the packet immediately before it was received or lost. The ensuing recalculated parameters are passed on to the third and final block. Further inputs for the third block include information about the number of frames in one RTP packet and which encoding rate is being used. These data are gained from measuring the RTP streams. The last block, called “cognitive model” calculates and outputs the MP3 factor as a value on the MOS(LQO) scale [13].

The cognitive model aims to replicate how humans process input parameters when evaluating the resulting speech quality. This is emulated here with the use of a suitable formula. The mathematical dependencies needed to do this are stored in the block in the form of a table. Its contents are calculated through the following steps:

4.1 Step 1

Use the tool from Fig. 2 to determine the PESQ curves as a function of packet loss rate, burst size and number of frames per RTP packet. The curves serve as a basis for further calculations. This approach replaces analyses in special studios in which groups of people are needed to voice a subjective opinion, and consequently saves time and money when it comes to developing parametrised QoS models.

4.2 Step 2

Plot the PESQ values against the product of the number of frames per RTP packet and the burst size (“burst frame product”) for a packet loss rate of 1, 2, 3, 4, 5, 6, 7 and 10%. Approximate these curves using Eq. 1. The approximation method “least squares” was used.

$$\text{MP3-Factor} = m \cdot \text{BFP} + b \quad (1)$$

The coefficient “ m ” is responsible for the gradient of the curve while the coefficient “ b ” equals the value at a fictitious burst frame product of “0”.

4.3 Step 3

Continue to use the values of “ m ” and “ b ” calculated in Step 2. Plot the coefficients “ m ” and “ b ” as functions of the packet loss rate. Approximate these curves using Eqs. 2 and 3 leading to Eq. 4. The approximation method “least squares” was used here, too.

$$m = a \cdot \text{pl}^2 + b \cdot \text{pl} + c \quad (2)$$

$$b = d \cdot \text{pl} + e \quad (3)$$

$$\text{MP3-Factor} = (a \cdot \text{pl}^2 + b \cdot \text{pl} + c) \cdot \text{BFP} + d \cdot \text{pl} + e \quad (4)$$

All five coefficients a , b , c , d and e of Eq. 4 are dependent on the encoding rate used.

These steps will now be illustrated for the MP3 codec (see Chapter 3) using an encoding rate of 80 kbps as an example.

Figure 6 shows the approximations for various packet loss rate achieved using Eq. 1. For reasons of clarity, only five curves have been included in the graph.

For an increasing level of packet loss rate, the coefficient “ m ” increases while the coefficient “ b ” decreases.

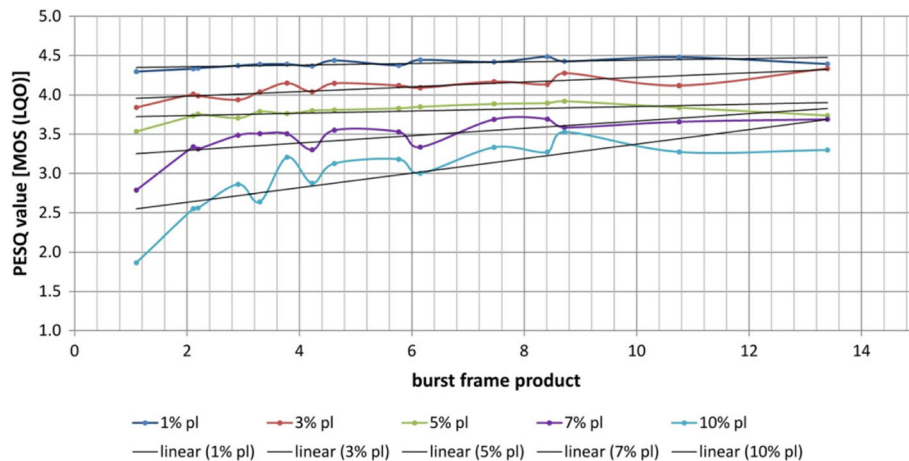


Fig. 6 Approximation of the PESQ values for different levels of packet loss rate and an encoding rate of 80 kbps

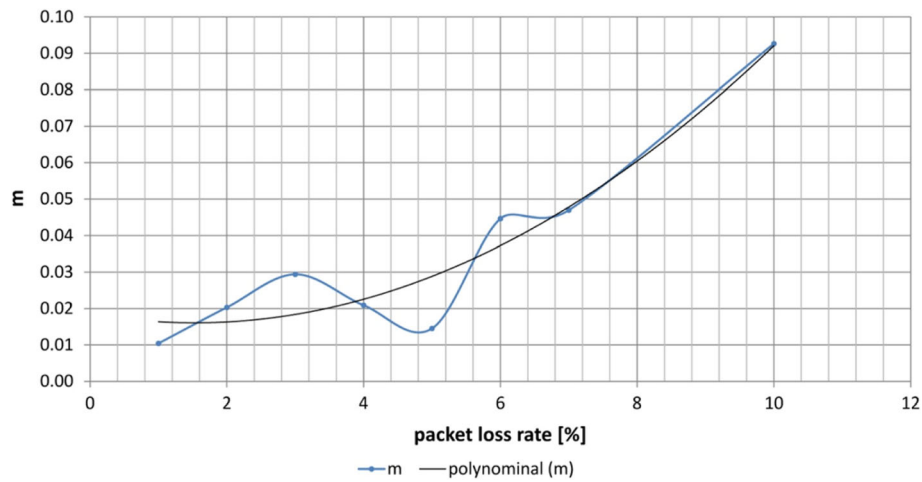


Fig. 7 Approximation of the parameter “*m*” for an encoding rate of 80 kbps

The approximations of the parameter “*m*” and “*b*” are illustrated in Figs. 7 and 8, respectively.

These results lead to the following equations for the MP3 Model with an encoding rate of 80 kbps:

$$\text{MP3-Factor} = (0.0011 \cdot \text{pl}^2 - 0.0032 \cdot \text{pl} + 0.186) \cdot \text{BFP} - 0.2036 + 4.5837 \quad (5)$$

Further equations were developed for the MP3 Model in Eq. 5 for other typical encoding rates of the MP3 codec, i.e. 64 and 96 kbps. This extends the possible areas of application of the QoS Model that was constructed here. The following describes a comparison study in which the suitability of the new QoS Model for real practical use was put to test.

5 Comparison study

The comparison study was performed within a numerical environment using the tool from Fig. 2. The following settings were assumed:

- Nondeterministic (binominal distribution) distributed packet loss rate from 0 to 10%, nondeterministic (exponential distribution) distributed burst size “1” and 1 frame per RTP packet.
- Nondeterministic (binominal distribution) distributed packet loss rate from 0 to 10%, nondeterministic (exponential distribution) distributed burst size “1” and 3 frames per RTP packet.
- Nondeterministic (binominal distribution) distributed packet loss rate from 0 to 10%, nondeterministic (exponential distribution)

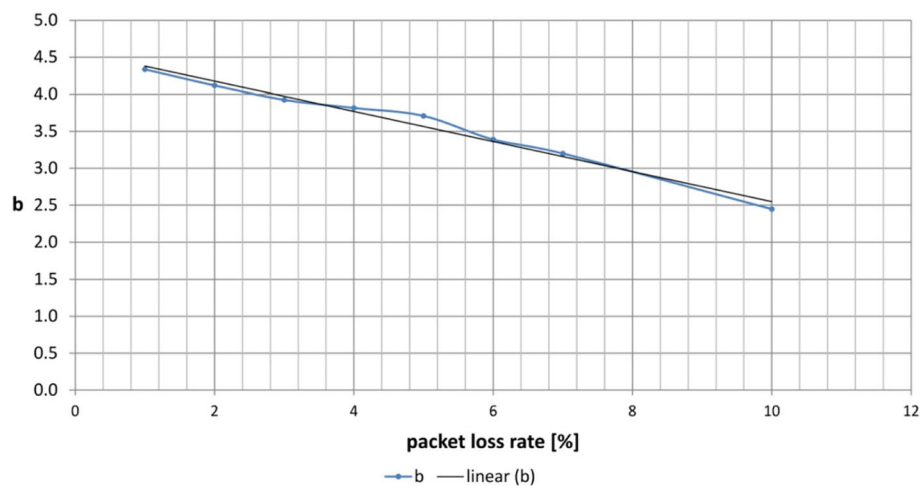


Fig. 8 Approximation of the parameter “*b*” for an encoding rate of 80 kbps

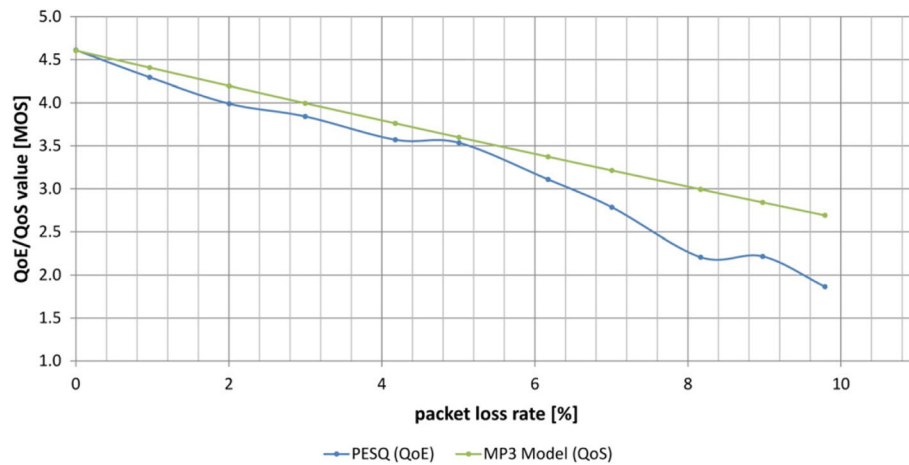


Fig. 9 QoE/QoS values as a function of packet loss rate using PESQ and the MP3 Model as measurement methods for the MP3 codec (80 kbps), a burst size of “1” and a frame number per RTP packet of “1”

distributed burst size “3” and 1 frame per RTP packet.

- Nondeterministic (binominal distribution) distributed packet loss rate from 0 to 10%, nondeterministic (exponential distribution) distributed burst size “3” and 3 frames per RTP packet.
- MP3 (MPEG-2 Audio Layer III) audio codec with an encoding rate of 80 kbps.
- Thirty-one measurements for each determined performance value. This number of tests ensures a confidence interval of less than 10% of the estimated average (with a probability of error of 5%).
- PESQ and MP3 Model as QoE/QoS measurement methods.

The results of the comparison study are illustrated in Figs. 9, 10, 11 and 12.

Figures 9, 10, 11 and 12 show that the quality of service deteriorates linearly as packet loss rate increase. This is the case for both QoE/QoS measuring methods used here. Furthermore, the curves fall less steeply as burst size increases. The upshot of this is it is far better for the service if fewer, larger bundles of packets are lost than lots and lots of smaller bundles. The reason for this lies in the properties of the human sense of hearing. The PESQ and the MP3 Model curves proceed very close to each other, meaning that the numerical comparison study has proved that the MP3 Model is quite suitable for practical use.

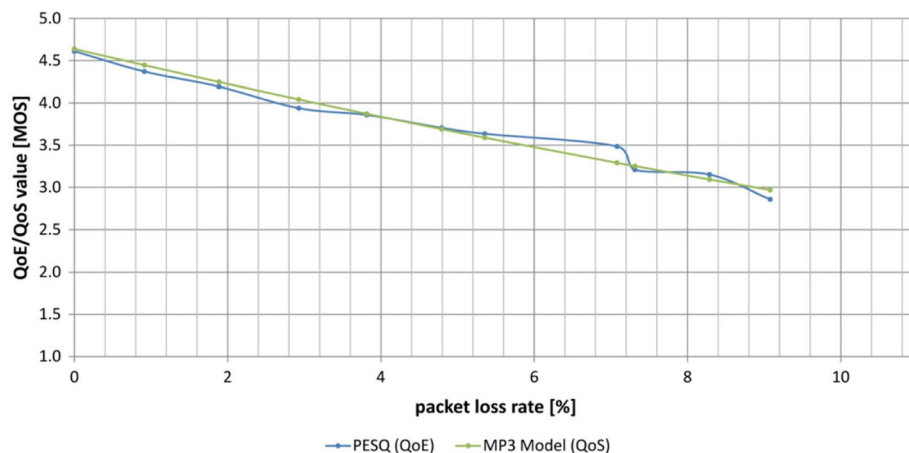


Fig. 10 QoE/QoS values as a function of packet loss rate using PESQ and the MP3 Model as measurement methods for the MP3 codec (80 kbps), a burst size of “3” and a frame number per RTP packet of “1”

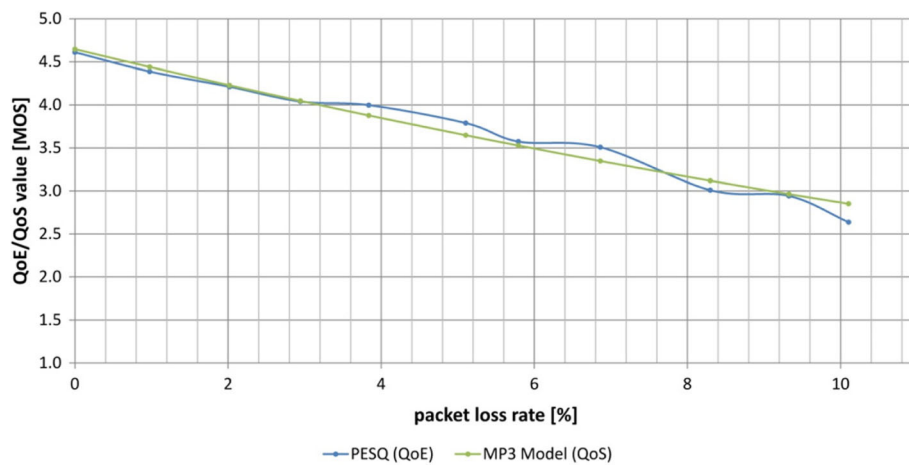


Fig. 11 QoE/QoS values as a function of packet loss rate using PESQ and the MP3 Model as measurement methods for the MP3 codec (80 kbps), a burst size of "1" and a frame number per RTP packet of "3"

6 Conclusions

In the course of the work documented in this paper, a new, simple model was defined for determining the QoS of audio data coded according to the MP3 coder and transported in an IP environment. Its suitability for practical use was tested in a comparison study. This so-called MP3 Model is one of a group of parameter-based measuring techniques; it works efficiently and delivers results that come very close to their corresponding perceptual evaluation of speech quality (PESQ) curves, which is, of course, of great practical benefit.

The ITU-T Standard G.1070 [16] (it, too, is a parametrised model) is often used to evaluate the QoS of the VToIP service. The G.1070 Model does, however, have its limitations: it uses the familiar E Model according to ITU-T G.107 [8] to evaluate the audio streams. But the

E Model does not support the new, innovative MP3 codec. This is where the MP3 Model, which was developed within the scope of this paper, comes in. It can be used to great effect in practice and is a sensible enhancement of the G.1070 Model.

As Table 1 shows, the MP3 codec also has modes in which the so-called Super-Wideband Signals (48 kHz sampling rate) are used. So, it would make sense to expand the functionality of the available MP3 Model with such features. One way to do this would be to use the procedure presented in this paper (cf. Chapter 4) for deriving the corresponding equations for the cognitive model. Also, it must be pointed out that for such functionality enhancement, the PESQ algorithm in the tool in Fig. 1 will have to be replaced by the software package POLQA [17]. The authors are about to work on that.

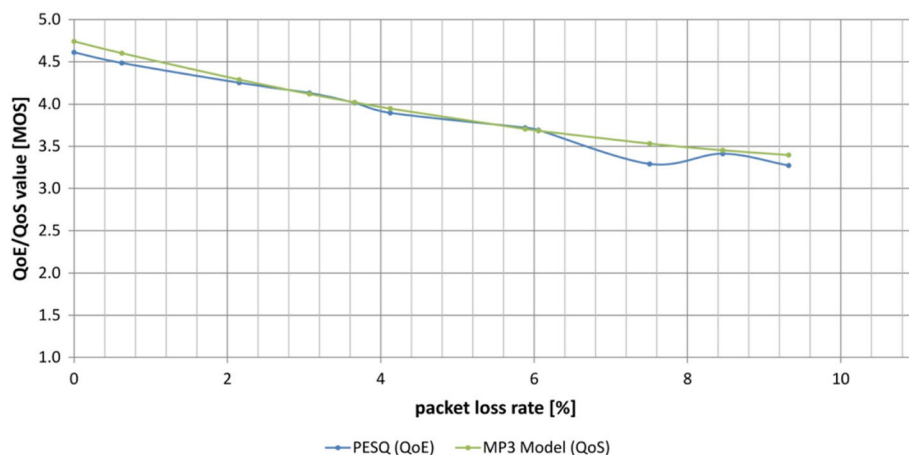


Fig. 12 QoE/QoS values as a function of packet loss rate using PESQ and the MP3 Model as measurement methods for the MP3 codec (80 kbps), a burst size of "3" and a frame number per RTP packet of "3"

The approach as described in the paper can also be applied in a very similar way to audio (music) streaming on IP platforms. However, for such cases, the PEAQ algorithm [6] has to be applied as measurement method for QoS. A more generalised approach of the topic provides an interesting aspect and opportunities for further scientific studies and enhances the importance of this paper.

Other interesting lines of work would be the following: the implementation of an alternative method of encapsulating audio data according to the IETF Specification RFC 3119 [18]; the development of corresponding QoS transformation equations and their practical implementation. By doing that, the flexibility of the MP3 Model could be increased even further. In practice, it is possible to glean information on the type of encapsulation used from the received data stream (by using the appropriate measuring instruments) and through evaluation of the measurement data. Further work in this direction is planned by the authors as well.

Authors' contributions

Tadeusz Uhl 40%; Stefan Paulsen 30%; Krzysztof Nowicki 30%.

Competing interests

The authors declare that they have no competing interests.

Author details

¹Flensburg University of Applied Sciences, Kanzlei St. 91-93, 24943 Flensburg, Germany. ²Maritime University of Szczecin, H. Pobożnego St. 11, 70-507 Szczecin, Poland. ³KGU-Consulting GmbH Flensburg, Lise-Meitner-St. 8, 24941 Flensburg, Germany. ⁴Gdańsk University of Technology, G. Narutowicza St. 11-12, 80-233 Gdańsk, Poland.

Received: 4 September 2016 Accepted: 20 December 2016

Published online: 19 January 2017

References

1. Definition of the NGN. (ITU-T, 2004). <http://www.itu.int/rec/T-REC-Y.2001>. Accessed Nov 2016.
2. DIRECTIVE 2009/136/EC Of the European Parliament and of the Council amending Directive 2002/22/EC on universal service and users' rights relating to electronic communications networks and services, Directive 2002/58/EC concerning the processing of personal data and the protection of privacy in the electronic communications sector and Regulation (EC) No 2006/2004 on cooperation between national authorities responsible for the enforcement of consumer protection laws (hereinafter: Citizens' Rights Directive). CELEX number: 32009L0136, November 10th 2009.
3. DIRECTIVE 2009/140/EC of the European Parliament and of the Council amending Directives 2002/21/EC on a common regulatory framework for electronic communications networks and services, 2002/19/EC on access to, and interconnection of, electronic communications networks and associated facilities, and 2002/20/EC on the authorization of electronic communications networks and services (hereinafter: Better Regulation Directive). CELEX number: 32009L0140, November 25th 2009.
4. A Raake, *Speech Quality of VoIP* (John Wiley & Sons, Chichester, 2006).
5. Standard ITU-T P.862. (ITU-T, 2001). <http://www.itu.int/rec/T-REC-P.862/en>. Accessed Nov 2016.
6. Standard ITU-R BS.1387. (ITU-R, 2001). <http://www.itu.int/rec/R-REC-BS.1387/en>. Accessed Nov 2016.
7. F Benedetto, G Giunta, C Belardinelli, Audio watermarking of MP3 music signals for quality assessment in multimedia communications. *The Mediterranean Journal of Electronics and Communications* 5(3), 106–113 (2009).
8. ITU-T Recommendation G.107. <http://www.itu.int/rec/T-REC-G.107/en>. Accessed Nov 2016.
9. S Paulsen, T Uhl, *Adjustments for QoS of VoIP in the E-Model* (Paper presented at the World Telecommunications Congress, Vienna, 2010).
10. E(IP) model, *German Patent 102010044727*, 2010.
11. IETF Standard, *REC 5219*, 2008. <https://tools.ietf.org/html/rfc5219>. Accessed Nov 2016.
12. S Paulsen, T Uhl, Numerisches Tool zur Untersuchung der QoS bei VoIP (in German). Paper presented at the 7th Workshops MMBnet2013, Hamburg, 5-6 September 2013.
13. Standard ITU-T P.564 (ITU-T, 2007). <http://www.itu.int/rec/T-REC-P.564/en>. Accessed Nov 2016.
14. RFC 2250: RTP Payload Format for MPEG1/MPEG2 Video. (The Internet Society, 1998). <http://www.ietf.org/rfc/rfc2250.txt>. Accessed Nov 2016.
15. Silence insertion method in accordance with ITU-T G.113 (ITU-T, 2007). <http://www.itu.int/rec/T-REC-G.113/en>. Accessed Nov 2016.
16. Standard ITU-T G.1070. (ITU-T, 2012). <http://www.itu.int/rec/T-REC-G.1070/en>. Accessed Nov 2016.
17. Standard ITU-T P.863. (ITU-T, 2011). <http://www.itu.int/rec/T-REC-P.863/en>. Accessed Nov 2016.
18. RFC 3119: RTP Payload Format for MPEG1/MPEG2 Video. (The Internet Society, (2001). <http://www.ietf.org/rfc/rfc3119.txt>. Accessed Nov 2016.
19. MPEG, *Audio Layer I/II/III Frame Header*, 2015. http://mpegedit.org/mpegedit/mpeg_format/MP3Format.html. Accessed Nov 2016.
20. MP3 audio encoding. (2015). http://www.mp3encoding.de/index_9.html. Accessed Nov 2016.

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Immediate publication on acceptance
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► springeropen.com