Speech Quality of Wideband VoIP under Packet Loss

Alexander Raake, Marcel Wältermann, Nicolas Côté, Sebastian Möller

Quality & Usability Lab, Deutsche Telekom Laboratories, TU Berlin
alexander.raake@telekom.de

Abstract: In this paper, we propose a model that predicts the quality impairment of wideband (50-7000 Hz) speech transmitted using Voice over Internet Protocol (VoIP) under packet loss. Our approach is an extension of the so-called E-model, a parametric tool recommended by the International Telecommunication Union (ITU-T) for the planning of telephone networks [3]. The present study follows an earlier extension of the E-model's quality rating scale from narrowband (300-3400 Hz) to wideband [23, 4], and the specification of so-called Wideband Equipment Impairment Factors Ie,wb that quantify the impairment due to wideband speech coding under error-free conditions [18]. Based on an extensive series of listening tests, we present some new considerations on how wideband speech quality is impacted by uniform packet loss. These considerations are well in line with earlier work on narrowband speech quality under VoIP packet loss [21, 22]. Following the proposed method, we present new wideband equipment impairment factors for a number of codecs and a set of so-called Packet Loss Robustness Factors Bpl for the a number of speech codecs.

1 Introduction

Modern telecommunication systems employing Voice over Internet Protocol (Voice over IP) allow the narrow telephone transmission bandwidth of 300-3400 Hz to be extended to wideband (WB). In case of a clean, otherwise undistorted channel, the speech quality of a WB system is higher than that of a narrowband (NB) system. However, WB speech quality may be degraded by different factors, such as non-linear distortion due to the applied low-bitrate WB codecs, or IP packet losses that are translated into a time-varying perceptive impairment by the employed decoder and packet loss concealment technique [22].

Since the perceived speech quality is internal to a user of a given system, the most valid way for assessing it is to carry out tests with human subjects [17]. Since such tests are time-consuming and expensive, and often even practically impossible, several instrumental models have been developed that (within limits) predict speech quality as perceived by human subjects (see e.g. [23] for an overview). These models can be differentiated according to the application aimed at (e.g. network planning, codec development, monitoring), the network components or configuration under test (e.g. entire connection mouth-to-ear, codecs, user-interfaces), the model input (signals, parameters), and the required information on the reference before transmission across the test system (No Reference – NR, Reduced Reference – RR, Full Reference, FR).

The present work introduces extensions of the E-model, which is a parameter-based, no-reference network planning model initially designed for traditional, narrowband, circuit-switched telephone networks operated with handset telephones ([3]. The quality measure provided by the E-Model is the Transmission Rating Factor *R*, which is expressed on the so-called *R*-scale.

The E-model relies on the assumption that different types of degradations are additive in terms of the perceptual impairment they cause. This is reflected by its basic formula:

$$R = R_0 - Is - Id - Ie,eff \qquad (1)$$

Here, R is the Transmission Rating, expressed on the model's quality scale that ranges from 0 to $R_{0,max}$. For NB speech, the bandwidth the model was developed for initially, $R_{0,max} = 100$. In previous work, we have extended this maximum range to WB, with $R_{0,max} = 129$ [23, 18]. With this extension, NB and WB speech quality can be expressed on the same scale. In Equation (1), R_0 reflects the base-quality that is related to the basic signal-to-noise-ratio; Is is the simultaneous impairment factor, which expresses the quality impairment due to degradations such as signal-correlated noise; Id is the delayed impairment factor, which accounts for the degradation due to pure delay and echo; Ie,eff is the effective equipment impairment factor which accounts for the quality impairment due to speech coding (Ie) and eventual packet loss in VoIP-type systems ($Ipl = Ie,eff - Ie$).

The WB-version of the E-model is currently under development. In the meantime, the scale extension with $R_{0,max} = 129$, and several equipment impairment factors for error-free WB and NB codecs (Ie_{WB}) have been standardized [4]. Note that WB equipment impairment factor values for NB codecs can easily be obtained as $Ie_{WB} = Ie_{NB} + 35.8$ (see [18] for details). In this context, Ie_{NB} are the equipment impairment factor values for NB codecs recommended in [5].

In the present work, we have conducted an extensive series of listening tests to investigate the speech quality of a variety of NB and WB connections. One key aspect of the test was to extend the partly established WB E-model towards the handling of packet loss, starting from first related considerations in [18]. The remainder of the paper is structured as follows: Section 2 describes the listening tests, and in Section 3 we give a first analysis of the results. In Section 3.1, we present new equipment impairment factors for error-free codecs, and Section 4 presents the impairment model for WB packet loss.

2 Listening Test Set-up

Two listening tests were conducted (referred to as Test 1 and 2 in the following).

2.1 Test 1

In Test 1, different conditions of WB and NB speech codecs were assessed, namely:

1. in single and tandem operation,
2. under IP packet loss,
3. in the presence of background noise at send side.

In total, 114 test conditions plus 11 reference conditions were tested, using source recordings from four speakers (two female, two male). The conditions included WB-codecs such as clean PCM, the AMR-WB (ITU-T Rec. G.722.2), the G.722, and the G.729.1 [11, 8, 14], and NB-codecs such as the G.711, the G.729A, and the G.726 [7, 13, 12]. In addition to the single operation mode, both WB/WB and NB/WB codec tandems were tested. Most codecs in single operation were also tested with additional background noise at send side. Here, two types of noise were used: Cafeteria noise and car noise, each at two different levels.

For the majority of tested WB codecs, a number of conditions involved uniform packet loss, with loss-rates from the set $0, 1, 2, 4, 8\%$. The ITU processing tools were used for sample generation. As opposed to classical tests with samples from different speakers being assessed during one test session, we have used one set of listening sessions per each of the four speakers. In all other respects the tests were conducted according to [20]. As sentence material, shortened versions of the German EUROM sentence material were used [1]. 38 sentences were selected from the available 40. Each sentence from each speaker was processed with the 125 conditions, yielding a total of $38 \cdot 4 \cdot 125 = 19000$ files.

For each listener, the playlists (one per speaker and per listener) were created by – for each condition – randomly selecting one of the available 38 sentences. For the purpose of finding an appropriate E-model extension for WB codecs under packet loss, this procedure has the advantage that it yields a much better sample of the space of possible uniform packet loss patterns; in contrast, the more typical fixed combination of one specific sentence with a given speaker-conditions-pair yields a more arbitrary coverage of that space. Hence, for each loss condition, $4 \cdot 38 = 152$ (speakers · sentences) loss patterns were tested in total.

The speech files were presented diotically using Sennheiser HD 25 headphones. Open headphones could not be used, since 6 subjects participated in the tests at the same time (but listened to different playlists). The test was administered using 6 separate laptop computers each equipped with RME HDSP Cardbus Cards and RME HDSP Mulitface II Soundcards. Each subject could listen to each of the sound files only once, and gave ratings using a mouse and a slider-based version of the 5-point ACR-scale [20] using a test GUI (both Test 1 and Test 2). The diotic presentation-level was 73 dB SPL. 120 paid subjects took part in Test 1 (appr. 50% female, 50% male; age 17 to 80 years).

In the analysis of the results, only 100 subjects were retained: The initial goal of the test was to collect both quality ratings on WB-relevant channels, but also to provide insight into the relation between quality perception and user group aspects (IT experience, age, etc., see [24]). The audiometric screening conducted in this context indicated that the user group consisting of the 20 oldest users needed to be excluded to yield more fine-grained results.

2.2 Test 2

In Test 2, the goal was to assess the quality of different WB and NB codecs under packet loss. Due to a sub-optimal usage of the processing tools for the G.722.2 under packet loss in Test 1, particular emphasis was here laid on obtaining data that would be usable for the loss impairment model to be derived for the G.722.2 codec. As opposed to Test 1, only 30 subjects (app. 50% female, 50% male) participated in Test 2, with now fixed pairings between the conditions and the sentences for each speaker. Here, the same speaker-condition-sentence-combination was employed for all listeners, hence testing only 4 traces per loss rate (one trace per sentence). For Test 2, we have also added several reference conditions for NB speech under packet loss, yielding 50 test conditions and 18 reference conditions in total. Each subject carried out 2 sessions with two sub-sessions each, listening to one speaker per each of the four sub-sessions. The technical set-up during the test was identical to that of Test 1 (see Section 2.1). Several error-free conditions were identical for both Test 1 and 2, and were used to linearly adjust the Test 2 results to the Test 1 results (see Section 3).

3 Test Results

The MOS-data were transformed onto the WB E-model R-scale [0, 129], following a similar procedure as in [18], i.e.:

1. The Test 2 MOS-data were linearly transformed to the Test 1 data using the clean codec conditions employed in both tests. The respective linear relation is shown in the left panel of Figure 1.

2. MOS-data [1,5] were transformed to E-Model's NB R-scale [0,100] using the transformation given in ITU-T Rec. G.107 (2008).

3. The R,nb-values were linearly transformed to obtain R,wb-values, using: $R,wb = 1.29 \cdot R,nb$.

4. From the R,wb-values, preliminary wideband equipment impairment factor values Ie,wb were calculated using $Ie,wb = 129 - R,wb$.

5. The Ie,wb-values were linearly normalization following (within limits) the approach described in [16].

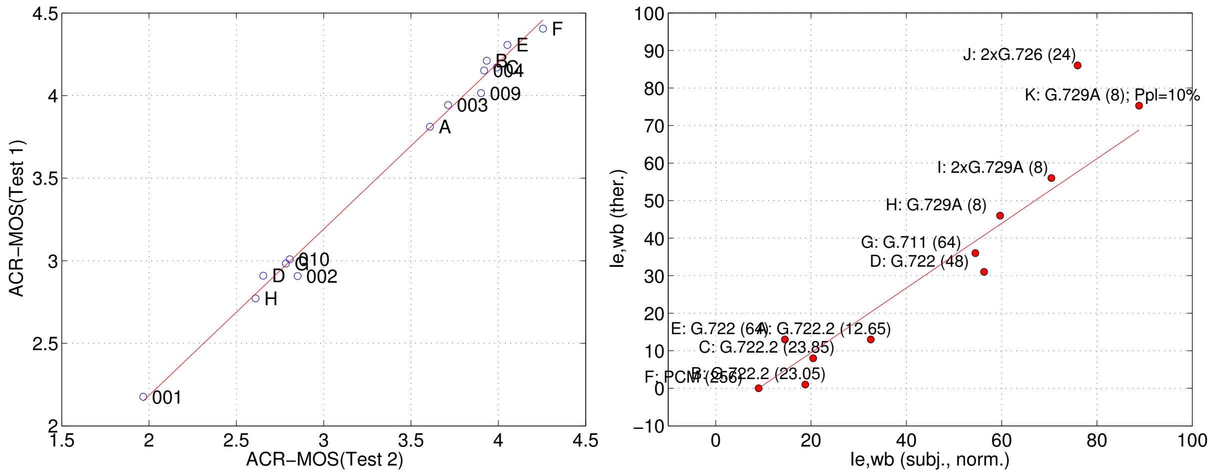

Figure 1 - Left panel: Correlation between the Test 1 and Test 2 results, used for transforming the Test 2 results into the Test 1 rating space. Right panel: Transformation to adjust the test results for conditions of known impairment to the values recommended in [5, 6]; the corresponding fitting-function is then applied to all test results to yield normalized values, following a method adopted from [16].

For Test 1, the results of step 5 are depicted in Figure 1, where the transformed test results are plotted in terms of preliminary WB equipment impairment factors $Ie,wb(subj.,norm)$ against the corresponding expected $Ie,wb(theor.)$ taken from [6]. As can be seen from the figure, a line appears to be a good approximation of the relation between the two data sets. However, it can also be observed that there are some rank-order differences between the recommended values and the test data points derived in steps 1. and 2., as will be discussed in some more detail below. Note that the G.726*G.726 tandem has been omitted from the transformation, since the large deviation from the general trend cast doubt on the additivity property or the recommended Ie-values in this case.

The plot in Figure 2 shows a comparison of the test results for single-coding conditions. The deviations of the transformed results from the R-values expected based on [4, 6] are highlighted: The blue boxes indicate the mean test results; the stacked boxes serve to indicate conditions yielding lower quality ratings than expected (pink), and conditions with higher quality than expected (purple).

A first observation to be made is that the test confirms the quality advantage of WB over NB of more than 35 points on the 129-point WB R-scale; the E-model prediction for a clean G.711 channel is $R_{NB} = 93.2$. An unexpected result is the reversed quality rank-ordering of the G.722 at 64 kbit/s and the G.722.2 at 23.05 kbit/s (compared to [2, 6]; see bars 4 and 9). For the

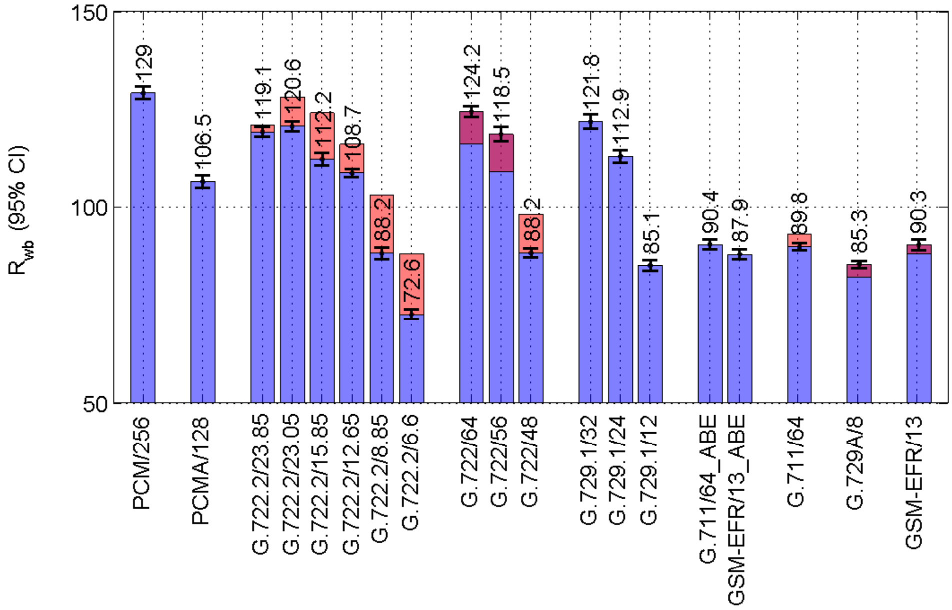

Figure 2 - Transformed and normalized results for error-free codecs. The blue boxes and respective errorbars show the test-results. The stacked pink and purple boxes indicate the deviation of the test results from the expected quality according to [6] and [4] (pink/errorbars below top of box: Test results lower than expected; purple/errorbars on top of box: test results higher than expected. Note: ABE indicates two NB conditions with artificial bandwidth extension.

G.722.2, the quality-decrease with decreasing bitrate is stronger than expected (bars 9 and 10). In contrast, the G.722 is rated better than expected, at least at the two higher bitrates.

Informal listening to the processed samples and a comparison with other source material processed using the same channel conditions reveal an influence of the room-acoustics at the recording site: An increasing amount of (especially early) room reflections audible in the signal seems to lower the quality-impact of nonlinear coding distortions. A similar observation has been described for audio codec distortions perceived at receive side in reverberant conditions in [25]. However, in the tests underlying [6], the speech material was recorded in a moderately reverberant environment, i.e. reverberant at send-side; this can be revealed by an informal listening to the source test data (which are taken from [15]). According to the respective test descriptions summarized in [18, 2], this source sentence material was used in most of the tests underlying the equipment impairment factor values for WB-codecs now listed in [6]).

When the source speech material is en- and decoded, some of the coding artifacts appear to be less audible when the input speech is moderately reverberant. In contrast, in our tests non-reverberant recordings have been used. Another difference between our test and the majority of previous Ie,wb-derivations [18, 2] is the employed diotic headphone-listening; in most of the previous WB codec tests summarized in [18, 2], monotic headphone-listening was used. A test series conducted by Nagle et al. [19] already indicated the increased sensitivity towards quality degradations when listening diotically to coded speech (as compared to monotic listening). In their results, however, a systematic rank-order change between codecs when comparing the diotic and monitic test was not observed, probably partly due to the more classicalnnumber of test subjects in their case (32 versus the 100 subjects considered in our test). For diotic listening, the G.722 appears to be less critically perceived than in case of monotic presentation, which may be explained by

1. The preference for wider bandwidths in case of diotic listening: A lower cut-off frequency

may be perceived as more plausible in a diotic listening context, whereas a 200-7000 Hz bandpass is preferred over a 50-7000 Hz bandpass in case of monotic listening (see [23]).

2. The noise introduced by the G.722 being perceived as less disturbing for diotic listening.

In total, all of the above-mentioned effects (room-acoustics at send-side and diotic vs. monotic listening mode) combined may explain the unexpected rank-ordering of the G.722 and G.722.2 at their best bitrates. The effect is currently being investigated in more detail at our lab.

3.1 Wideband Equipment Impairment Factors Ie, wb for Diotic Listening

The test results were transformed into the *R*-scale and then Impairment-factor-domain following steps 1-5 described above. This leads to a first set of wideband equipment impairment factors for diotic listening for the employed WB codecs: These results underline the differences found

Table 1 - WB Equipment Impairment Factors for WB codecs under diotic listening, and corresponding values from [6]. Notes: () The proposal is valid for diotic listening. () At 12 kbit/s the G.729.1 behaves like an NB codec. In the NB-context and with diotic listening, the corresponding Ie is of $Ie = 44 - 35.8 = 8.2$.

Codec	**Bitrate [kbit/s]**	Ie, wb	
		Monotic, standardized ([6])	proposed
G.722	64	13	5
	56	20	10
	48	31	41
G.722.2	23.85	8	10
	23.05	1	8
	15.85	7	17
	12.65	13	20
	8.85	26	41
	6.6	41	56
G.722.1	32	-	7
	24	-	16
	12()	-	44

between our tests and the tests compiled for [6]. The proposed values will be used in the following for the case of IP transmission errors in coded speech.

4 Impairment Factors for WB Speech Codecs under Packet Loss

In our listening tests Test 1 and Test 2, we have employed packet loss rates from the set $\{0, 1, 2, 4, 8\}\%$. We will now use the results obtained for the G.722 at 64 kbit/s for both the high- and low-quality packet loss concealment algorithms [9, 10], for the G.722.2 at 12.65 kbit/s, 23.05 kbit/s and 23.85 kbit/s, and the G.729.1 at 12 kbit/s, 24 kbit/s and 32 kbit/s to derive a framework for the Wideband Effective Equipment Impairment Factor.

From the test results shown in the plots in Figure 3, a first general observation can be made: As can be seen when comparing the data from Test 1 (left) to those obtained in Test 2 (right), the

Test 2 results show a much larger variation from loss rate to loss rate. This smoothing-effect for Test 1 can be explained as follows: In Test 1, for each subject and condition, the sample to be used per speaker was randomly selected from a set of 38 available ones. In contrast, in Test 2 only a limited number of test files were employed per condition (30 subjects, 4 sentences/loss-patterns per condition, reflecting a more common practice in such packet loss tests). The Test 1 results can be considered as a sample over a large variety of error-pattern-to-speech-pattern mappings. This reflects the application of a parametric model for network planning very well, where predictions for average situations are being sought. For the case of e.g. signal-, i.e. file-based models, this approach is less practical, since in this case the aim is to dispose of the link between individual files and quality ratings. However, since the G.722.2 was properly tested only in Test 2, we will rely on Test 2 for the G.722.2. For the two other tested codecs, we will use the Test 1 results. The general form of the curves shown in Figure 3 is highly similar

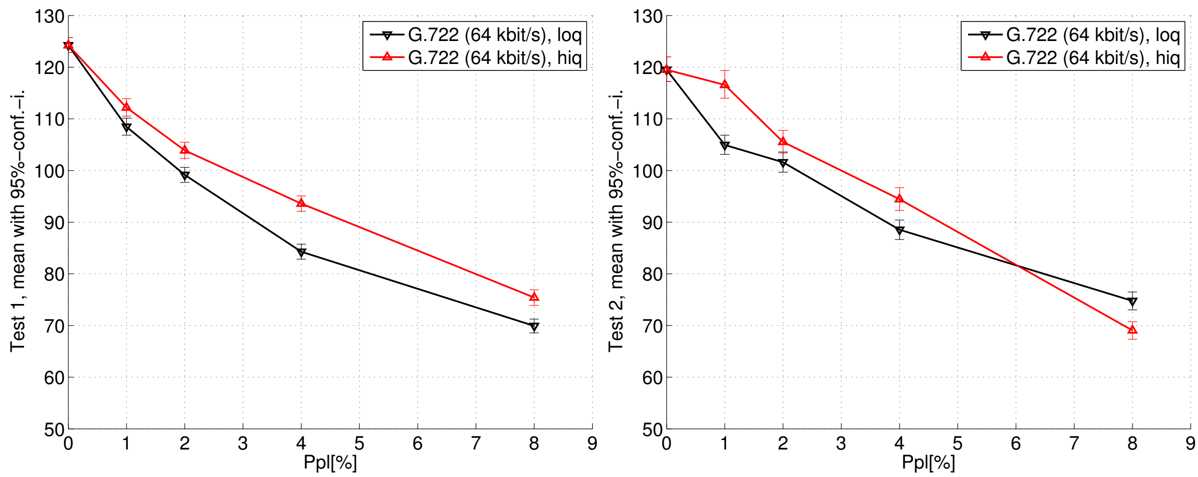

Figure 3 - Comparison of Test 1 (left) and Test 2 (right) test results for the G.722 (64 kbit/s) under packet loss. It can clearly be seen that the overall shape of the mean quality ratings over packet loss rate is much smoother for Test 1, where due to the employed approach a large number of traces were assessed.

to those of the curves delivered by the E-model for NB codecs. For NB, we had derived the following dependency of Ie,eff on the (uniform) packet loss percentage Ppl [23, 22, 3]:

$$Ie,eff = Ie + (95 - Ie) \cdot \frac{Ppl}{Ppl + Bpl}. \quad (2)$$

Here, Bpl is the codec/PLC-dependent Packet Loss Robustness Factor typically determined via curve-fitting of auditory test results.

We have carried out a curve-fitting of the WB test results presented in this paper taking the same approach as in Equation (2) for NB, using:

$$Ie,wb,eff = Ie,wb + (95 - Ie,x) \cdot \frac{Ppl}{Ppl + Bpl}, \quad (3)$$

with

$$Ie,x = \begin{cases} Ie, & \text{if NB codec} \\ Ie,wb, & \text{if WB codec} \end{cases} \quad (4)$$

Curve-fitting of the results for each codec at a given bitrate was done separately, yielding individual Packet Loss Robustness Factors Bpl for each codec-bitrate-combination, see Table 2.

Table 2 - Packet Loss Robustness Factors *Bpl* for WB codecs under diotic listening.

Codec	G.722		G.722.2			G.729.1	
Bitrate [kbit/s]	64		23.85	23.05	12.65	32	24
PLC	App. III	App. IV	G.722.2			G.729.1	
Ie,wb	5	5	10	8	20	7	16
Bpl	7.1	5.1	4.9	4.6	4.3	6.1	7.3

Comparing the predicted quality ratings with the transformed test ratings reveals a very high correlation of 0.991 and a Root Mean Squared Error of the predictions of $RMSE = 2.48$ (on the extended R-scale 0-129). Note that in [18], we had initially proposed a somewhat different formula for the impairment of WB speech quality under packet loss. However, the approach according to Equation (3) is considered to be more appropriate for a further extension of the E-model to WB-speech under packet loss, since:

- It is fully downward-compatible to NB, because it employs the same equation as the NB E-Model, Equation (2).
- It may be the basis for further extensions: For example, it is compatible with the approach of the bandwidth impairment factor Ibw introduced in [23, 26] to cover the quality impairment due to linear distortion, e.g. introduced by the employed terminal equipment; consequently, the approach can be used for both NB and WB, as setting $Ie,x = Ie,wb - Ibw$ enables a first separation of a codec's bandwitdh-related contribution to the overall impairment from that due to non-linear codec distortion.

Due to these considerations and the high agreement with the subjective tests, Equation (3) has now been adopted by ITU-T Study Group 12 for an update of [4]. Figure 4 shows the respective model performance.

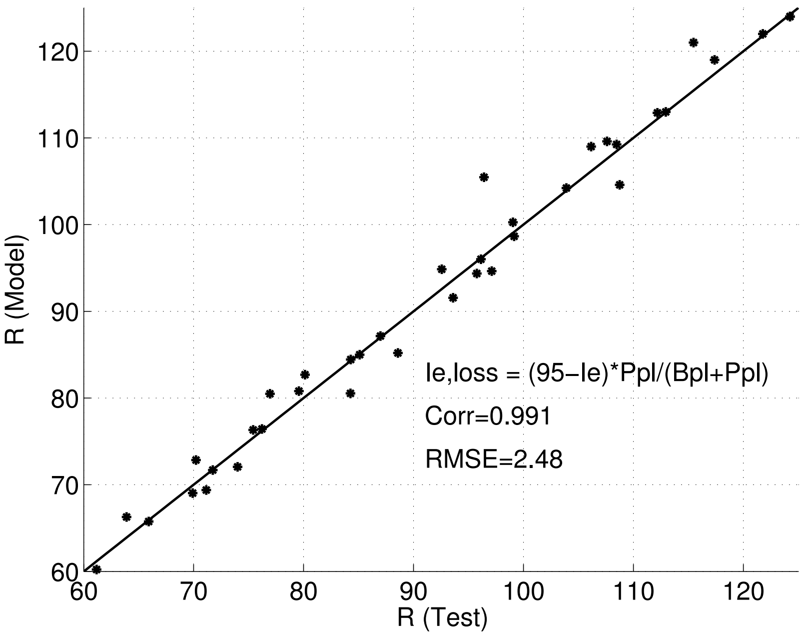

Figure 4 - Correlation between model predictions and test results for Equation (3).

5 Conclusions

An extensive listening test has been conducted on speech quality under wideband and narrowband speech coding and VoIP packet loss. It was shown that the test results deviate from previous results from the literature. Several reasons for these deviations were discussed. A diotic versus monotic presentation method as well as the room-acoustic recording conditions of the source speech material were identified as possible causes for the observed results: The perceptual dimensions of the quality impairment introduced by a given codec appear to be interacting with the source recording conditions and presentation method. It has further been shown that the effect of uniform packet loss on wideband speech quality can be modeled using the same approach as used for NB speech for the E-model. The resulting model equation enables a unified handling of both narrowband and wideband speech codecs under packet loss. It allows future extensions to easily be made, since it implicitly comprises the separation between the linear distortion introduced by e.g. a narrowband codec, and the codec's non-linear distortion that interacts with the impairment due to packet loss. This aspect is of relevance for the usage of the WB E-model for NB connections: The model assumes an additivity of its underlying impairment factors, so that the bandwidth-contribution to the overall impairment of narrowband codec tandems is considered twice. Separating the linear- and non-linear-distortion component as proposed in [26] is in line with the packet loss handling as it is proposed here. Our current and future work consists in developing the model parts relative to talker echo and delay, with the aim of proposing a first full WB-E-Model.

References

[1] GIBBON, D.: *EUROM.1 German Speech Database*. ESPRIT project 2589 report (SAM, Multi–Lingual Speech Input/Output Assessment, Methodology and Standardization), Universität Bielefeld, D–Bielefeld, 1992.

[2] ITU–T DELAYED CONTRIBUTION D.151: *Towards a wideband E-model: R-scale extension and Impairment factors for wideband speech codecs*. Source: Federal Republic of Germany and Japan; authors: Möller, S., Raake, A., Kitawaki, N., Takahashi, A., Wältermann, M. International Telecommunication Union, CH–Geneva, June 2006.

[3] ITU–T REC. G.107: *The E–Model, a Computational Model for Use in Transmission Planning*. International Telecommunication Union, CH–Geneva, 2009.

[4] ITU–T REC. G.107 APPENDIX II: *Provisional Impairment Factor Framework for Wideband Speech Transmission*. International Telecommunication Union, CH–Geneva, June 2006.

[5] ITU–T REC. G.113 APPENDIX I: *Provisional Planning Values for the Equipment Impairment Factor Ie and Packet-Loss Robustness Factor Bpl*. International Telecommunication Union, CH–Geneva, May 2002.

[6] ITU–T REC. G.113 APPENDIX IV: *Provisional Planning Values for the Wideband Equipment Impairment Factor Ie,wb*. International Telecommunication Union, CH–Geneva, June 2006.

[7] ITU–T REC. G.711: *Pulse Code Modulation (PCM) of Voice Frequencies*. International Telecommunication Union, CH–Geneva, November 1988.

[8] ITU–T REC. G.722: *7 kHz Audio-Coding Within 64 kbit/s*. International Telecommunication Union, CH–Geneva, November 1988.

[9] ITU–T REC. G.722, APPENDIX III: *A high quality packet loss concealment algorithm for G.722*. International Telecommunication Union, CH–Geneva, 2006.

[10] ITU–T Rec. G.722, Appendix IV: *A low-complexity algorithm for packet loss concealment with G.722*. International Telecommunication Union, CH–Geneva, 2007.

[11] ITU–T Rec. G.722.2: *Wideband Coding of Speech at Around 16 kbit/s Using Adaptive Multi-Rate Wideband (AMR-WB)*. International Telecommunication Union, CH–Geneva, January 2002.

[12] ITU–T Rec. G.726: *40, 32, 24, 16 kbit/s Adaptive Differential Pulse Code Modulation (ADPCM)*. International Telecommunication Union, CH–Geneva, December 1990.

[13] ITU–T Rec. G.729 Annex A: *Reduced Complexity 8 kbit/s CS-ACELP Speech Codec*. International Telecommunication Union, CH–Geneva, November 1996.

[14] ITU–T Rec. G.729.1: *G.729 based Embedded Variable Bit-rate Coder: An 8-32 kbit/s Scalable Wideband Coder Bitstream Interoperable with G.729*. International Telecommunication Union, CH–Geneva, May 2006.

[15] ITU–T Rec. P.501 Amendment 1: *Test signals for use in telephonometry; New Annexes A and B*. International Telecommunication Union, CH–Geneva, 2004.

[16] ITU–T Rec. P.833.1: *Methodology for the Derivation of Equipment Impairment Factors from Subjective Listening-only Tests for Wideband Speech Codecs*. International Telecommunication Union, CH–Geneva, May 2008.

[17] Jekosch, U.: *Voice and Speech Quality Perception — Assessment and Evaluation*. Signals and Communication Technology. Springer, D–Berlin, 2005.

[18] Möller, S., A. Raake, N. Kitawaki, A. Takahashi and M. W. altermann: *Impairment Factor Framework for Wideband Speech Codecs*. IEEE Trans. Audio Speech and Language, 14(6):1969–1976, 2006.

[19] Nagle, A., C. Quinquis, A. Sollaud and A. Battistello: *Quality Impact of Diotic Versus Monaural Hearing on Processed Speech*. In: Proc. 123rd Audio Engineering Society (AES) Convention, Oct. 5-8, USA–New York, 2007.

[20] P.800, I.-T. R.: *Methods for Subjective Determination of Transmission Quality*. International Telecommunication Union, CH–Geneva, June 1996.

[21] Raake, A.: *Predicting Speech Quality under Random Packet Loss: Individual Impairment and Additivity with other Network Impairments*. Acta Acustica united with Acustica, 90(6):1061–1083, 2004.

[22] Raake, A.: *Short- and Long-Term Packet Loss Behavior: Towards Speech Quality Prediction for Arbitrary Loss Distributions*. IEEE Trans. Audio Speech and Language, 14(6):1957–1968, 2006.

[23] Raake, A.: *Speech Quality of VoIP – Assessment and Prediction*. John Wiley & Sons Ltd, Chichester, West Sussex, UK, 2006.

[24] Raake, A., S. Spors, H.-J. Maempel, T. Marszalek, S. Ciba and N. Côté: *Speech Quality of Wide- and Narrowband Codecs: Object- and Subject-oriented View*. In: Fortschr. Akust., 34. Jahrestagung für Akustik (DAGA 2008), D-Dresden, pp. Dtsch. Ges. Akust. (DEGA), D–Berlin., 2008.

[25] Schobben, D. and S. van de Par: *The effect of room acoustics on MP3 audio quality evaluation*. In: Proc. 117th Audio Engineering Society (AES) Convention, Oct. 28-31, USA–San Francisco, 2004.

[26] Wältermann, M. and A. Raake: *Towards a new E-Model Impairment Factor for Linear Distortion of Narrowband and Wideband Speech Transmission*. In: Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP 2008), March 30 - April 4, USA–Las Vegas:4817– 4820, 2008.