Time Varying Distortions and Traditional Telephony: Speech Quality under Combined Network Distortions

Alexander Raake
Institut für Kommunikationsakustik, Ruhr-Univeristät Bochum, raake@ika.ruhr-uni-bochum.de

Introduction

A user of today's telecommunication systems is not necessarily right when he believes that the classical PSTN/ISDN[i] are applied for his connections, even if he uses wire-line handset telephones. Instead of PSTN/ISDN, Voice over Internet Protocol (VoIP) may be the technology for parts of the network he uses. In VoIP-systems, degradations occur that are perceptually different from the stationary ones known from PSTN/ISDN-systems. These degradations are time-varying by nature, such as delay variation (jitter) or packet-loss.

One of the most typical network scenarios to be encountered for VoIP will be of the structure PSTN – VoIP – PSTN. A user of such a network will be exposed to degradations resulting from both transmission types, which will impact his satisfaction with the telephone service. Consequently, a provider of such a service wants to design his network so that it meets specific quality criteria. For this purpose, he may rely on predictions delivered by quality models (for an overview of models cf. Möller and Raake[1]). The so-called E-model is the model currently recommended by the International Telecommunication Union (ITU-T) for network planning[2].

In the current version of the model, time-varying distortions are covered in a very limited way. So far, only random packet-loss has been included in the formulae. Moreover, the E-model's assumption of additivity of impairments on a perceptual scale has to be verified for the case of packet-loss combined with other types of degradations. The latter is the purpose of the present paper. A series of four conversation tests is presented studying speech quality for combinations of different impairments with random packet-loss. The results are used in order verify how random packet-loss is handled by the model, and to discuss the assumption of impairment additivity for combined degradations.

E-Model

The E-model is based on a parametric description of the transmission path[2]. Its fundamental assumption is that different degradations can be transformed onto a psychological scale as *impairment factors*. On this scale, the degradations are assumed to be additive, yielding a 'Transmission Rating Factor' *R* for the particular transmission condition ($R \in [0, 100]$, 100 for highest quality, eq.1).

$$R = Ro - Is - Id - Ie,eff + A \qquad \text{eq. 1}$$

Ro, the 'Basic Signal-to-Noise Ratio', is calculated from the send loudness rating *SLR* (weighted attenuation) and the level sum of all noise sources present in the connection (e.g. circuit noise *Nc* and wideband line noise *Nfor*). *Is* is the 'Simultaneous Impairment Factor' for degradations simultaneous to the transmitted speech signal, such as signal-correlated noise. The 'Delayed Impairment Factor' *Id* accounts for the degradations delayed to the speech signal, such as echo (e.g. talker echo, with echo attenuation *TELR* and delay *T*) or absolute delay *Ta*. *A*, the 'Advantage Factor', is an additional factor quantifying the effect of user expectation.

Ie,eff is the 'Effective Equipment Impairment Factor' for specific speech processing equipment such as speech codecs. Since recently[3], *Ie,eff* also includes the effect of random packet-loss, based on the equation

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

The 'Equipment Impairment Factor' *Ie* is a codec-specific value derived from auditory tests. *Bpl* is the 'Packet-Loss Robustness Factor', which is codec-specific, too. For recommended values of *Ie* and *Bpl* for different codecs cf. ITU-T Rec. G.113 [4].

Formulae[2] exist for the conversion of the 'Transmission Rating Factor' to Mean Opinion Score (MOS), as obtained in subjective quality tests[7], and vice versa. Although validity of the additivity assumption has been investigated for some combinations of impairments[5], no auditory experiments have so far been reported on the combination of packet-loss with other types of degradations. Verification of predictions for this case is necessary, as the 'Effective Equipment Impairment Factor' *Ie,eff* does not depend on any input parameters joint with other impairment factors.

Test Set-Up

An online-tool for simulation of the majority of degradations typical for PSTN/ISDN/VoIP-networks was used, which was developed at IKA based on the parametric network description underlying the E-model[2] (for details cf. Rehmann et. al.[6]). The system set-up reflects the usage scenario PSTN – VoIP – PSTN, with handset-telephones as user interfaces. For the introduction of packet-loss, the NISTNet network simulation was used (www.nist.gov, 2001).

The tests were carried out as short conversation tests (SCTs)[5]. In tests of this type, two interlocutors are asked to carry out short conversation tasks of 2 – 3 minutes duration over a simulated telephone connection (e.g. a train ticket reservation). After each conversation, the subjects were requested to rate the overall quality of the connection on the 5-point MOS-scale[7] typically used in telecommunications. An additional degradation scale was used in the tests, but for clarity, only the results obtained with the MOS-scale are described here. In each test series, 14 different conditions were used, two of which were reference conditions. One of these references was a clean connection with G.711 (logarithmic PCM, A-law, ITU-T Rec. G.711). Note that no other type of reference was presented, in order to avoid the introduction of an additional quality dimension. For the remaining 12 conditions of each test, the VoIP-typical codec G.729A (CS-ACELP, ITU-T Rec. G.729 Annex A) was used, with a packet-size of 20ms, and VAD disabled. Four different rates of random packet-loss were chosen:

- $Ppl \in$ [0%, 3%, 5%, 15%].

The additional degradations in the tests were:

- SCT1 (Noise): $Nfor \in$ [-64 dBmp, -50 dBmp, -40 dBmp]

[i] PSTN: Public Switched Telephone Network; ISDN: Integrated Services Digital Network

- SCT2 (Delay): *Ta* ∈ [200 ms, 400 ms, 600 ms]
- SCT3 (Echo, *T* = 100 ms): *TELR* ∈ [20 dB, 35 dB, 50 dB]

SCT1 was run in two versions for different attenuations of the speech paths (SCT1.1: *SLR* = 8 dB; SCT1.2: *SLR* = 13 dB), in order to enable a better comparability between tests: SCT2 was carried out with a high *SLR* of 13 dB, to avoid audible echoes due to coupling in the handset at receive-side and long delay, and SCT3 with *SLR* = 8 dB. 22 subjects participated in each test-series.

Results and Discussion

The results show that in general the E-model (eq. 2) delivers valid predictions for speech quality under random packet-loss, also for a conversational situation. When line noise is the additional degradation (SCT1.1 and 1.2), two observations can be made:

- Lower attenuation of the speech path (*SLR* = 8 dB, SCT1.1) leads to more negative test results than higher attenuation (*SLR* = 13 dB, SCT1.2). Obviously, the degradation due to packet-loss or low-bitrate coding alone is rated more critically for higher speech levels. Distortions are more audible in this case. These results are in contradiction to the E-model predictions.
- For increasing noise, small differences between packet-loss rates are no longer well differentiated by the subjects. The perceptual degradation due to packet-loss seems to be partly masked by the noise.

In case of additional delay (SCT2), the chosen test method led to results almost independent of delay (Fig. 1).

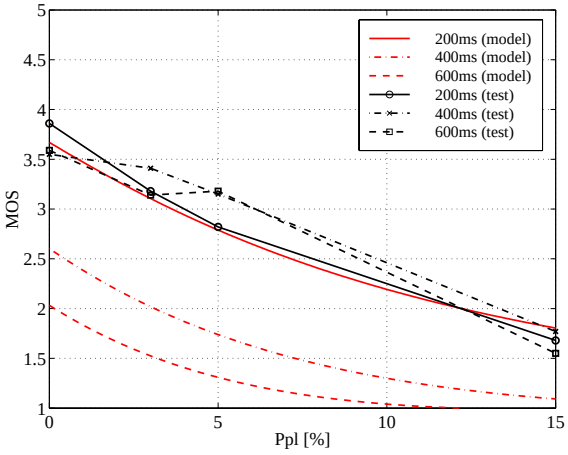

Figure 1: Results for packet-loss and delay (SCT2, lines without markers) and E-model predictions (lines with markers).

The tests were not highly interactive, and the subjects were not pointed out to the delay introduced in the test. Consequently, their ratings are dominated by the more obvious impairment of packet-loss. Although the findings are similar to other results reported in the literature[8], it can be assumed that for more interactive tasks the E-model leads to valid predictions.

The test results for SCT3 (packet-loss and talker echo) are considerably more positive than the corresponding E-model predictions. This is ascribed to the relatively high number of low quality connections presented in the test. Hence, the results were linearly transformed to match the E-model predictions for the connections with G.711 (reference) and with *TELR* = 20 dB, *Ppl* = 0%. As can be seen from Fig. 2, the E-model predictions are generally in good agreement with the transformed test data.

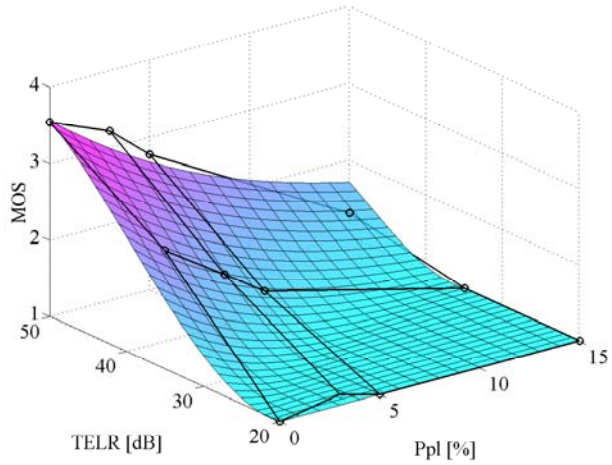

Figure 2: Linearly transformed SCT3-results for packet-loss and echo (grid) and E-model predictions (surface).

However, a small interdependence between the two degradation types can be observed: Quality under combined packet-loss and talker echo is higher than predicted by the model. Moreover, the effect of packet-loss is slightly reduced by additional echo, and vice versa. This could either be due to an interaction on the signal level (e.g. a reduction of echo by packet-loss on the echo signal), or to an interaction on the perceptual level (e.g. due to the attention of the subjects caused by a difference in quality dimensions).

In future tests, three aspects should be addressed in more detail:

- Quality under combinations of bursty packet-loss with other degradation types.
- The effect of speech level on quality in case of non-linear degradations and packet-loss.
- The effect of delay in combination with packet-loss for highly interactive tasks.

Acknowledgement

The present work has been performed at IKA, Ruhr-University Bochum (Prof. J. Blauert, PD U. Jekosch). It was partly funded by the EU (IST-project INSPIRE, www.inspire-project.org).

[1] Möller, S. and Raake, A., "Telephone Speech Quality Prediction: Towards Network Planning and Monitoring Models for Modern Network Scenarios", *Speech Communication*, 38: 47-75, 2002.

[2] ITU-T Rec. G.107, *The E-Model, a Computational Model for Use in Transmission Planning*. International Telecommunication Union, CH-Geneva, 2002.

[3] ITU-T Delayed Contribution D.044, *Modelling Impairment due to Packet-Loss for Application in the E-Model*, Source: Deutsche Telekom AG, Germany (A. Raake), International Telecommunication Union, CH-Geneva, 2001.

[4] 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, 2002.

[5] Möller, S., *Assessment and Prediction of Speech Quality in Telecommunications*, Kluwer Academic Publishers, USA-Boston, 2000.

[6] Rehmann, S., Raake, A., and Möller, S., "Parametric Simulation of Impairments Caused by Telephone and Voice over IP Network Transmission", *Proceedings EAA 2002 – Forum Acusticum*, ES-Sevilla, 2002.

[7] ITU-T Rec. P.800, *Methods for Subjective Determination of Transmission Quality*, International Telecommunication Union, CH-Geneva, 1996.

[8] Karis, D., "Evaluating Transmission Quality in Mobile Telecommunication Systems Using Conversation Tests", *Proc. of the Human Factors Soc. 35th Ann. Meeting*, 217-221, USA-San Francisco CA, 1991.