How do users perceive speech quality under combined "stationary" and non-stationary degradations?

Alexander Raake
IKA, Ruhr-Universität, D-44780 Bochum, Germany, Email: alexander.raake@ruhr-uni-bochum.de

Introduction

Users of today's telephone networks are faced with perceptually new types of degradations. This is due to the packet-based network technology increasingly applied, such as Voice over IP (VoIP). Since a given connection may be routed across different types of networks, users may be faced with combined "stationary" impairments like noise or echo, and non-stationary impairments like packet loss.

This paper addresses the question, how users perceive speech quality in case of such combined impairments. The question is not only a fundamental one. It is also motivated by the model currently recommended by the ITU-T (International Telecommunication Union) for network planning, the E-model [1]: In this model, it is assumed that different types of impairments can be grouped into certain classes and be transformed onto a perceptual impairment scale; on this scale, the resulting "impairment factors" are assumed to be *additive* (e.g. [2]). Based on conversation test results, the paper discusses how users perceive speech quality in case of combined impairments, and whether the perceived overall impairment actually corresponds to the sum of the perceived individual impairments.

Conversation tests

Two series of conversation tests were conducted at IKA. The first series of four tests investigated the quality perception under *random* packet loss combined with different levels of line noise, of transmission delay and of talker echo. The tests are discussed in more detail in [3]. The second series of three tests, which is reported here, was based on the same concept, however using *burst* instead of random packet loss. Random packet loss can be described using one parameter, the loss percentage *Ppl*. Random packet loss is comparable to a "stationary" impairment, as the variation of quality can be considered as a short-term variation (cf. concept of *microscopic* loss behaviour described in [2]). In contrast, the *burst* loss was created with a 3-state Markov model under the following constraints (cf. Figure 1):

- Four different overall loss rates $\in \{0\%, 3\%, 5\%, 15\%\}$.
- Mean distance between loss bursts (gap length): 18s.
- Mean duration of loss bursts (burst length): 18s.
- Within bursts: On average 2 packets lost consecutively.

With these constraints the four independent variables of the Markov model could be defined. The burst and gap lengths are chosen to yield stable quality impressions during the burst periods and during the pauses between bursts. The corresponding time constants are adopted from [4]. Text-files were created off-line indicating the packets to be lost. The files were used with the software "NetDisturb" (ZTI, Lannion, France), which allows packet loss to be inserted in a given Voice over IP connection. For the simulation of the different network conditions, the online system described in [5] was used with the G.729A codec (ITU-T Rec. G.729A).

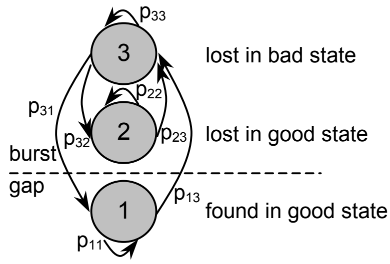

Figure 1: 3-state Markov model for introducing burst packet loss. The parameters p_{ij} quantify the probabilities for a transition between states *i* and *j*.

The mean ratings obtained in the conversation tests on the MOS-scale (mean opinion score: 5-point ACR-scale, [6]) were transformed onto the E-model *R*-scale using a formula provided in Appendix I of [1]. In order to make the interpretation of the results independent of the E-model predictions (which apply only to *random* loss), the transformed ratings obtained for combined impairments are compared to "expected ratings": These were calculated under the assumption of impairment additivity, using the transformed ratings obtained for individual impairments, according to Equation (1):

$$Rc_{G.729A}(Ppl,X) = R_{G.729A}(X) - \left[R_{G.729A} - R_{G.729A}(Ppl)\right] \quad (1)$$

$Rc_{G.729A}(Ppl,X)$ is the *expected* rating as a function of packet loss *Ppl* and the additional impairment level *X*. $R_{G.729A}(X)$ is the mean rating obtained for the G.729A-connection impaired by the impairment level *X* (no packet loss). $R_{G.729A}$ is the mean rating obtained for the G.729A-connection without further impairments. $R_{G.729A}(Ppl)$ is the mean rating for the connection under the packet loss percentage *Ppl*.

As can be seen from Figure 2, noise only had a minor effect on the ratings in the test on packet loss and noise floor *Nfor*, especially for higher loss rates. For lower loss rates, some masking of the packet loss artefacts due to higher noise levels can be observed, as the slope of the mean ratings plotted over the loss rate is lower. Obviously, an additivity of the packet loss impairment and of the impairment due to noise does not hold. This finding is similar to the observations described in [3]. However, burst loss seems to dominate the quality perception, as the effect of noise is considerably lower than for random packet loss.

From the results for packet loss and talker echo, a similar effect can be observed: Burst packet loss is the dominating impairment type. The impact of talker echo on quality is lower than expected (e.g., the E-model predictions, which are based on extensive auditory tests, imply a much more prominent effect of echo). The lower importance of talker echo can be explained with the attention of the subjects: An ANOVA was carried out using the echo attenuation *TELR* (talker echo loudness rating), the packet loss percentage *Ppl*, and the subject ID-N°. *n* as fixed factors. All three factors prove to be significant, as well as the interactions *TELR*Ppl*, *TELR*n* and *Ppl*n*. Obviously, it depends on the individual subject, which weight a particular impairment type gains.

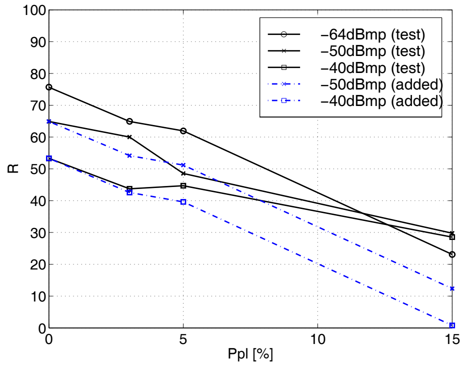

Figure 2: Conversation test results on the *R*-scale for burst packet loss and noise plotted over packet loss percentage *Ppl* [%]. The noise floor *Nfor* serves as curve parameter. Continuous lines: Actual test results. Slash-dotted lines: Expected rating behaviour obtained based on the ratings for the individual impairments (cf. text for more details).

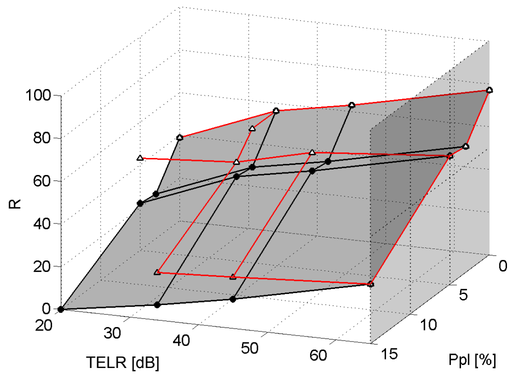

Figure 3: Conversation test results on the *R*-scale for burst packet loss and talker echo plotted over packet loss percentage *Ppl* and echo attenuation *TELR* (1-way echo delay *T*=100ms). Upper grid: Actual test results. Grey surface: Expected rating behaviour obtained based on the ratings for the individual impairments (cf. text for more details).

As depicted in Figure 4, delay did not show any major effect on speech quality, in contrast to how it is predicted by the E-model. The validity of the E-model predictions for delay is currently under discussion at the ITU-T. The comparison to the expected results $Rc_{G.729A}(Ppl,Ta)$ is not shown here, as it does not provide any additional information, due to the little impact of delay in this test. The tests on packet loss and delay were not carried out with the same conversation scenarios as the ones used in [3]. Instead, more interactive scenarios were developed, in which the subjects had to exchange certain data as quickly as possible. Since each pair of subjects new each other, and they were instructed to call the other by the first name, the conversation discipline was low enough in order not to limit the desired interactivity. Obviously, almost the entire attention of the subjects was drawn to the additional packet loss impairment presented in the test.

The observed prominence of the burst packet loss impairment can be explained with the severe degradation introduced in case of an overall rate of 15%: As gap and burst durations were chosen equally long, a correlated loss of 30% during bursts was used, which affected intelligibility considerably (from the recorded conversations it could be revealed that the subjects had to ask back several times during these severe bursts). Obviously, if intelligibility is the quality dimension mainly affected, other types of impairments only play a minor role.

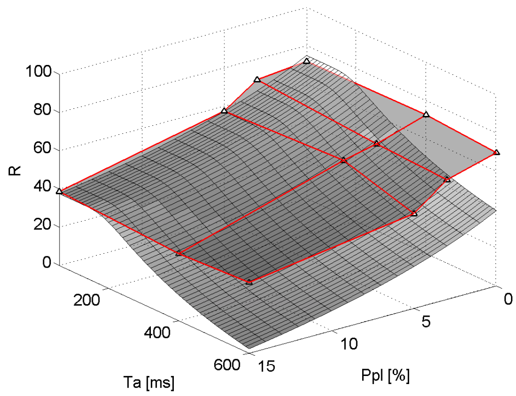

Figure 4: Conversation test results on the *R*-scale for burst packet loss and transmission delay plotted over packet loss percentage *Ppl* and delay *Ta*. Grid: Actual test results. Grey-shaded surface: E-model predictions.

Acknowledgement

The work was carried out at the Institute of Communication Acoustics (PD. U. Jekosch, Prof. R. Martin). It was partly funded by the EC in the framework of the IST project INSPIRE (IST-2001-32746, www.inspire-project.org). The author wants to thank Florian Hammer (Telecommunication Research Centre Vienna, Austria: ftw.) for his cooperation in running the conversation tests and many fruitful discussions.

References

[1] ITU-T Rec. G.107. The E-Model, a Computational Model for Use in Transmission Planning. International Telecommunication Union, CH-Geneva, 2003.

[2] Raake, A. Predicting speech quality under packet loss: Extension of the E-model. In: Proceedings CFA/DAGA 2004, Dtsch. Ges. Akust., D-Oldenburg, 2004.

[3] Raake, A. Time Varying Distortions and Traditional Telephony: Speech Quality under Combined Network Distortions. In: Fortschr. Akust. DAGA 2003, Dtsch. Ges. A-kust., D-Oldenburg, 2003.

[4] Gros, L., Chateau, N. Instantaneous and Overall Judgements for Time-Varying Speech Quality: Assessments and Relationships. Acta Acustica **87** (3), pp. 367-77, 2001.

[5] Rehmann, S., Raake, A., Möller, S. Parametric Simulation of Impairments Caused by Telephone and Voice over IP Network Transmission. In: Proc. 3rd EAA Convention, ES-Sevilla, 2002.

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