Predicting speech quality under packet loss: Extension of the E-model

Alexander Raake

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

Introduction

In today's voice communication systems, packet-based transmission increasingly replaces the synchronous circuit-switched network technology. Voice over Internet Protocol (VoIP) is typically applied in order to use the existing network infrastructure for voice communications. Packet-based transmission is characterized by time-varying impairments like packet loss which are perceptually different from the degradations known from traditional circuit-switched telephone networks.

In order to know which level of quality a network designer can achieve with the network he is planning, so-called network planning models are typically used. They deliver unidimensional quality indices based on instrumentally measurable network parameters available already in the planning phase. The most prominent example of such models is the E-model [1]. It was developed to predict speech transmission quality in case of impairments such as noise, the attenuation of the line, the sidetone levels, the impairment caused by the applied speech codecs, and conversational impairments such as echo or transmission delay. Since its revision in 2002, the E-model also covers the effect of random packet loss. Random loss, however, is not very typical for real IP networks. Instead, actual loss distributions can more accurately be modelled using an n-state Markov model, with defined transition probabilities between the states. Such "bursty" loss distributions are not yet covered by the E-model.

In this paper, a classification of the loss behaviour of a network is provided. The random loss approach currently recommended in the E-model will briefly be described. Finally, an extension of the model towards bursty (i.e. more correlated) packet loss will be presented and compared to the results obtained in listening tests.

Parametric Description of Packet Loss

For quality prediction, the loss distributions encountered in real networks can be classified as follows:

Microscopic loss behaviour: Consecutively lost coding frames affect some packet loss concealment algorithms (PLC) more than for others. A reason for consecutive frame loss in VoIP can be a correlation between packet loss events: The loss of a particular packet may be more likely if previous packets were lost.

Macroscopic loss behaviour: Gros and Chateau studied how listeners judge the quality of a speech sample that is degraded by periods of random packet loss of different loss rates [2]. They presented their subjects with long speech samples (up to 190s) impaired by different loss profiles. The subjects were asked for an instantaneous judgement using a slider method, as well as for a judgement of overall quality at the end of each sample. Gros and Chateau showed that the time average over the instantaneous judgement is a good predictor for overall quality.

A given *macroscopic* loss profile can be viewed as the concatenation of different sections of a particular *microscopic* loss behaviour. Based on this classification, the following scheme for modelling the impairment due to packet loss can be conceived (cf. also [3]):

- A generalized packet loss profile reflecting an expected loss behaviour is described by a metric distinguishing between *microscopic* and *macroscopic* loss behaviour.
- An estimate of instantaneous quality is calculated for the passages of stationary *microscopic* loss behaviour using an appropriate model (see below).
- A general prediction of speech quality under packet loss is calculated by time averaging over the prediction of instantaneous quality. The averaging over segments reflects the *macroscopic* burst behaviour.

Packet Loss in the E-model

The basic formula of the E-model implies the additivity of different classes of impairments on a perceptual impairment scale. Therefore, the instrumentally measurable input parameters of the model are transformed onto the "Transmission Rating Scale" or "R-scale", cf. Equation (1). The R-scale ranges from 0 to 100, with 100 reflecting the best possible quality. A prediction on the R-scale can be converted to a mean opinion score (MOS, 5-point ACR-scale [4]), according to a formula provided in Annex I of [1].

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

Ro is the transmission rating factor due to the basic signal-to-noise ratio, including all noise sources on the line and in the rooms at send and receive side. Is is the simultaneous impairment factor summarizing all impairments that are simultaneous to the transmitted speech signal, like an excessive speech level or signal-correlated noise. Id stands for the impairments delayed to the speech signal, such as echo or transmission delay. The impairment due to low bit-rate coding was included in the model as an add-on impairment factor, namely the *equipment* impairment factor Ie. Following the same rationale, random packet loss is included using a formula based on two new model input parameters, namely the packet loss percentage Ppl and the robustness of the applied codec and PLC (Bpl: "Packet Loss Robustness Factor"). The combined impairment due to low bit-rate coding and packet loss is handled in the effective equipment impairment factor Ie,eff [1].

In the light of the classification into *microscopic* and *macroscopic* types of loss behaviour, the random loss approach can be viewed as the basis for predicting instantaneous quality during phases of "constant" microscopic loss behaviour. An overall estimate can be derived by time-averaging over these periods (cf. [2,3,5]). The random loss approach does not describe microscopic loss behaviours showing a correlation between losses, and hence are characterized by higher numbers of consecutively lost packets, leading to a reduced performance of the codec–PLC algorithms.

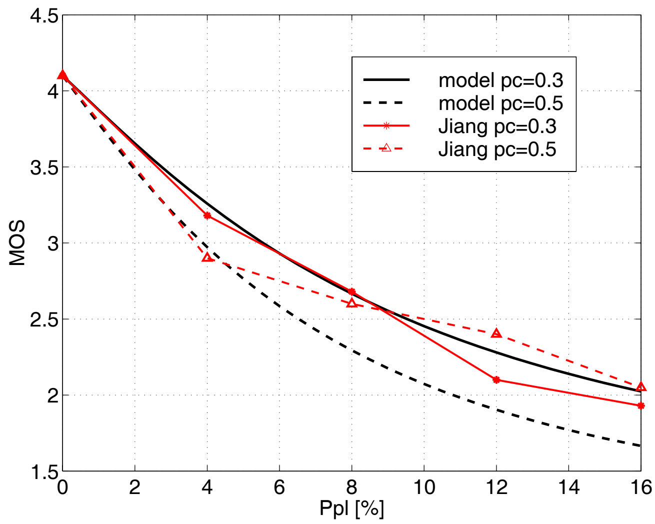

Figure 1: Listening test results obtained for 2-state random loss by Jiang and Schulzrinne [6], and model predictions using Equation (2). *pc* stands for the conditional loss probability, with $pc = 1\text{-}q$.

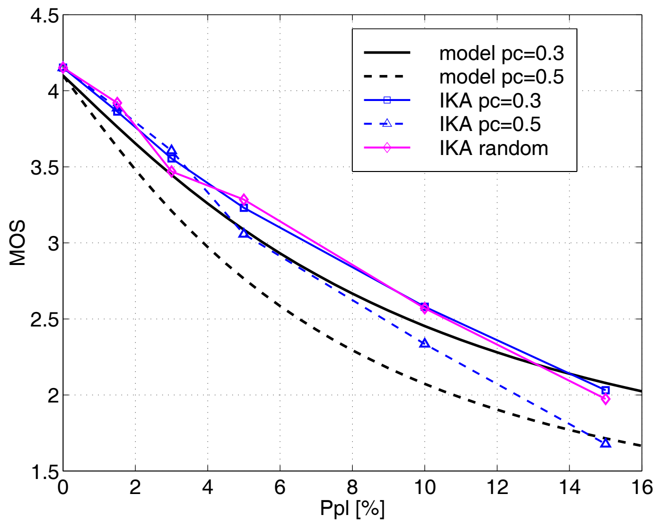

Figure 2: Listening test results obtained at IKA for different settings of 2-state Markov loss, and model predictions derived using Equation (2). *pc* stands for the conditional loss probability, with $pc = 1\text{-}q$.

Figures 1 and 2 illustrate the behaviour of speech quality in case of loss distributions obtained using a 2-state Markov loss model. The 2-state model has a good state ("0", found) and a bad state ("1", lost), and is characterized by the transition probabilities p from found to lost (0 to 1) and q from lost to found (1 to 0). Different degrees of correlation between loss events can be achieved, depending on these probabilities. Figure 1 shows listening test results reported by Jiang and Schulzrinne [6], and Figure 2 listening test results obtained in our lab (both on the MOS-scale, cf. [5]). To predict quality in case of this type of correlated losses, the E-model formula for random loss [1] can be extended, cf. Equation (2).

$$Ie,eff = Ie + (95 - Ie) \cdot \frac{Ppl}{Ppl + Bpl / Br} \qquad (2)$$

$$Br = \frac{\overline{\#\,\text{consecutively lost packets}_{\text{2-state}}}}{\overline{\#\,\text{consecutively lost packets}_{\text{random}}}} \qquad (3)$$

Here, the "Burst Ratio" *Br* is used as an additional parameter [7], see Equation (3). In case of 2-state Markovian loss, *Br* can be expressed as $Br = 1/(p+q)$, and the overall packet loss percentage *Ppl* as $Ppl = Br{\cdot}p{\cdot}100$.

With this approach, both the low quality ratings at low overall loss percentages found by Jiang and Schulzrinne, and the low ratings found in our tests for higher loss percentages can be captured, so that predictions with this formula lie on the safe side, i.e. are accurate or slightly more pessimistic than actual user ratings. For random packet loss, *Br* equals one, and the formula converges to the random loss formula used in the current E-model version [1].

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).

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

[3] Clark, A. Modeling the Effects of Burst Packet Loss and Recency on Subjective Voice Quality. In: Internet Telephony Workshop (IPtel 2001), USA-New York, 2001.

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

[5] ITU-T Delayed Contribution D.179. E-Model: Average Quality vs. Quality as a Function of Average Packet Loss. Source: Germany (A. Raake). International Telecommunication Union, CH-Geneva, 2003.

[6] Jiang, W., Schulzrinne, H. Comparison and Optimization of Packet Loss Repair Methods on VoIP Perceived Quality under Bursty Packet Loss. In: Proc. NOSSDAV'02, USA-Miami Beach, 2002.

[7] ITU-T Delayed Contribution D.020. The Burst Ratio: A Measure of Bursty Packet Loss. Source: Lucent (J. McGowan). International Telecommunication Union, CH-Geneva, 2001.