Parameter-based Modelling of Speech Quality in Telephone Networks

Alexander Raake

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

Introduction

Monitoring of speech quality in telephone networks can be carried out by measuring different network parameters, which are then related to speech quality using quality models. Examples for such parameter-based models are the so-called SUBMOD model, the Call Clarity Index (CCI), which is based on the SUBMOD approach, and the E-model, which is currently recommended by the ITU-T as a network planning tool. In this paper, these two models and their application to network monitoring are described. As an example, instrumental measurements of a commercially available monitoring system are presented, which were carried out using an on-line telephone line simulation tool. The quality estimates as determined by the monitoring system are compared to judgements as obtained from listeners in auditory tests. An outlook shows, how monitoring with parameter-based modelling approaches can be improved to be applicable to modern network scenarios.

1. Parametric Models

Monitoring can be carried out offline or online: (1) Specific test calls are set up and measurement signals such as noises or speech are transmitted across the network. From the comparison of the output- and input-signals, quality-relevant network parameters can be obtained. This offline measurement is called “intrusive”. (2) At a specific point of the network, a measurement signal is acquired during network operation. From this signal, network- or conversation-parameters relevant to quality can be derived. This online measurement is called “non-intrusive”. Monitoring devices based on non-intrusive measurements are referred to as INMDs (in-service non intrusive measurement devices, cf. ITU-T Rec. P.562, 2000). For a more detailed description, e.g. of the accessible parameters in both measurement cases, cf. (Möller and Raake, 2002).

For both intrusive and non-intrusive measurements models exist, which deliver estimates of the quality as perceived by a user of the network. The network parameters measured either intrusively or non-intrusively serve as input data to these parameter-based models. The two most relevant parametric speech quality models are the SUBMOD or CATNAP model, and the so-called E-model.

SUBMOD

The SUBMOD model, developed by British Telecom, was first described by (Richards, 1974). In a later, enhanced version it was renamed CATNAP (ITU-T Suppl. 3 to P-Series Rec., 1993). For parameters measured using INMDs, a specific model version has been defined, which is called Call Clarity Index (CCI; ITU-T Rec. P.562, 2000).

The SUBMOD model is based on spectral input parameters and an algorithm similar to that used in telephony for calculating loudness. The principle is the one suggested by (Fletcher and Munson, 1937), which relies on critical bands as obtained from masking experiments. Here, the threshold of audibility of a continuous spectrum sound such as speech is described as the pure tone threshold reduced by the critical bandwidth in dB units (Figure 1, curve c). Following this approach, the perceived loudness λ can be quantified using equation (1):

$$\lambda = const \int Q(Z) \cdot B'(f)\, df \quad (1).$$

The Sensation Level Z is the difference between the spectral density of the sound reaching the listener’s ear (which can be obtained from an average speech spectrum when the sensitivity of the transmission path is known, curve b) and the threshold of audibility of a continuous spectrum sound (here speech) masked by noise (curve d). *Q(Z)* is the Loudness Growth Function, and *B’(f)* a frequency weighting function.

Applying a similar integral-function, the so-called Listening Opinion Index (*LOI*) is calculated by the SUBMOD model, equation (2).

$$LOI = A_{LOI} \cdot D_{LOI} \int_0^{\infty} P(Z)\, B'_{LE}(f)\, df \qquad (2)$$

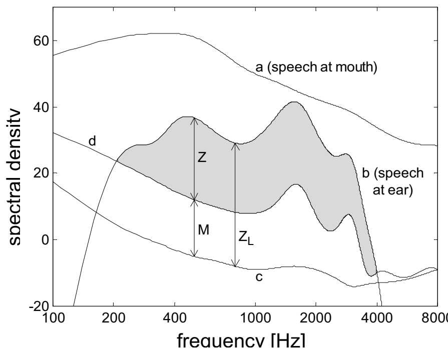

Figure 1: *Spectral densities underlying the loudness calculation in telephony (Richards, 1974). a: Speech spectrum at the speaker‘s mouth; b: Speech spectrum at the listener‘s ear; c: Hearing threshold for continuous spectrum sounds; d: Hearing threshold masked by noise.*

Here, *P(Z)* is the listening-opinion growth function, ranging from 0 to 1 (as there is no infinite growth of listening opinion). B'_{LE} is the frequency weighting function for listening effort. The loudness dependent factor A_{LOI} is a function of the grey-shaded area between curves b and d and ranges from 0 to 1. The noise-factor D_{LOI} is related to the section of curve d limiting this area (the audibility threshold in noise), and ranges from 0 to 1.

From the Listening Opinion Index, the so-called Listening Effort Score Y_{LE} is calculated. These two quality measures are valid only for the listening-only situation. In order to capture also impairments relevant for the conversational situation, a Conversation Opinion Score Y_C is determined. It additionally copes for the vocal levels of the two interlocutors and talker echo. Both Y_{LE} and Y_C range from 0 to 4, and can be transformed into a Mean-Opinion Score (MOS) by simply adding 1.

E-Model

The E-model currently is the only recommended model for network planning (ITU-T Rec. G.107, 2000). It has been developed by a group within ETSI (Johannesson, 1997).

The E-model is based on a parametric description of the transmission path, as depicted in Figure 2. Examples for such parameters are loudness ratings of the main speech transmission path (send loudness-rating *SLR*, receive loudness rating *RLR*, overall loudness rating *OLR*), talker echo loudness rating and delay (*TELR* and *T* respectively), sidetone loudness ratings (*STMR* and *LSTR*), and noise sources (circuit noise *Nc*; room noise at send and receive side, *Ps* and *Pr*).

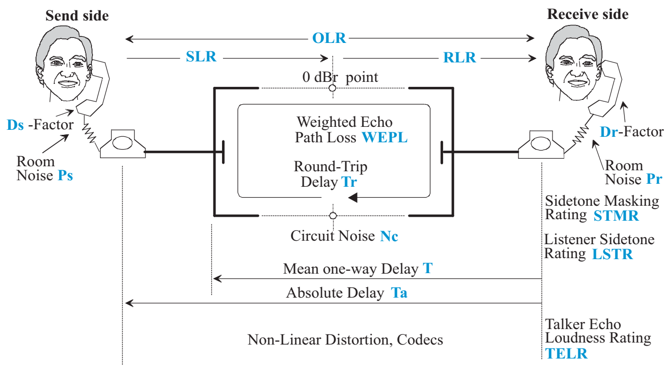

Figure 2: *Parametric representation of the transmission path for a PSTN/ ISDN telephone network, as it underlies the E-model (ITU-T Rec.G.107, 2000).*

The fundamental assumption of the E-model is that the “[...] evaluation of psychological factors on a psychological scale is additive”, which means that all impairments present in a specific connection are transformed onto an appropriate psychological scale and are added up to form an integral transmission rating factor. This is reflected by the basic formula of the E-model, equation (3):

$$R = Ro - Is - Id - Ie + A\,. \quad (3)$$

R is the Transmission Rating Factor, ranging from 0 to 100, 100 for the best possible quality. It results from the transmission quality due to the Basic Signal-to-Noise Ratio R_0 degraded by additional impairments on the line: Is is the Simultaneous Impairment Factor for degradations simultaneous to the transmitted speech signal, such as non-optimum loudness or signal-correlated noise; Id is the Delayed Impairment Factor accounting for the degradations delayed to the speech signal, such as echoes or pure delay; Ie is the Equipment Impairment Factor for specific speech processing equipment such as speech codecs. A, the Advantage Factor, is an additional factor quantifying the effect of user expectation.

In case of network monitoring, the network parameters forming the different impairment factors have to be calculated from intrusive or non-intrusive measurements. Parameters which are not accessible with the chosen measurement approach have to be replaced with estimates, according to the network type being monitored (e.g., in most GSM-networks the usage of the GSM-EFR codec can be assumed and a corresponding Equipment Impairment Factor of 5 be used).

2. Application: Intrusive Monitoring and E-model

As an application example, the usage of the E-model in combination with an intrusive monitoring approach will be described in the following. The aim of this study was to evaluate both the measurement accuracy and quality prediction performance of a commercially available intrusive monitoring tool for GSM-networks.

In order to dispose of an environment for controlled setting of those transmission parameters the monitoring device was aimed at, an online simulation-tool was used that had been set up at IKA in previous projects (cf. Möller, 2000). The tool was modified so as to also include interruptions of controlled length and distribution.

In the first set of measurements, the simulation settings were verified using a noise measurement device and a network analyser. In the

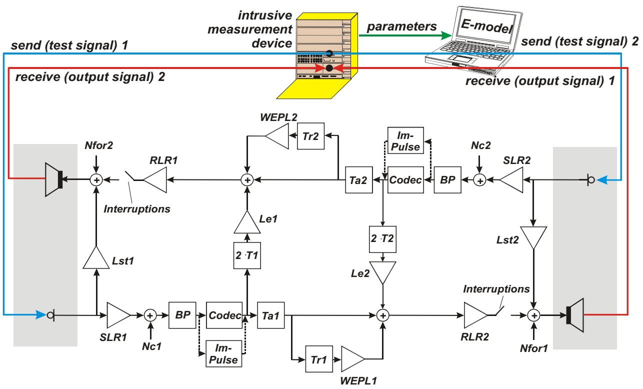

Figure 3: *Line simulation tool connected to intrusive monitoring device.*

second step, the online-simulation was connected to the intrusive monitoring device, as shown in Figure 3.

From the comparison of the input and output signals, the tool delivers measurements of different network parameters, as listed in Table 1, column #1.

Parameter	Measurement		E-model	
	Method	Main problem	Method	Update?
Loudness rating (*OLR*)	✓	-	✓	-
Noise (*N*)	✓	-	✓	-
Echo loss (*EL*)	✓	-	✓	-
One-way Delay (*T*)	✓	-	✓	-
Interruptions	(✓)	Length Categorization	Not yet included	cf. Equation (4)
Impulsive Noise	(✓)	Separation from Stationary noise	Not yet included	No update available

Table 1: *Performance of intrusive measurement of network parameters and quality predictions.*

Columns #2 and #3 of Table 1 indicate the accuracy of the parameter measurements carried out for individual simulation settings, and show what major problems occurred. It has to be noted that for the measurement of interruptions and of impulsive noise currently no recommended procedures exist, and that the solutions applied in the monitoring device were proprietary ones. The accuracy of categorizing interruptions depend on the moment the interruptions occur, as in the case of real life conversations between human interlocutors. Here, an interruption is only perceived when it is situated within a speech passage, or effects the background noise transmission. The categorization was especially difficult for long interruptions, which are – however – quite uncommon in real GSM-connections (in this case, the connection is usually lost). For loud noise levels, the device was sometimes unable to differentiate impulsive noise from stationary noise.

The parameters measured in this way serve as input-parameters to the E-model version implemented in the device. Neither interruptions nor impulsive noise are included in the current version of the E-model; for these parameters, proprietary solutions were used in the device. For interruptions, this proprietary solution is based on a suggestion by Johannesson (ITU-T SG 12 D.71, 1995) for speech clipping:

$$I_{e,eff} = I_e + 95 \cdot \frac{Cl}{Cl + 95/7.5}\,. \quad (5)$$

In this case, Ie in equation (3) has to be replaced by the effective equipment impairment factor Ie,eff; Cl corresponds to the percentage of clipped speech.

In the third step of the study reported here, the quality predictions delivered by the tool were evaluated, using theoretical E-model predictions based on simulation settings on the one hand, and auditory listening and conversation tests with these settings on the other hand. For all tested parameters currently included in the E-model, the quality predictions were accurate. For interruptions, the proprietary solution worked reliably, as long as the parameters were measured correctly. The predictions for impulsive noise could not satisfactorily be evaluated, as these are neither included in the E-model, nor can this impairment type be simulated with sufficient accuracy using the current version of our simulation tool.

3. Conclusion and Outlook

The algorithms of the two main parametric models used for speech quality monitoring, namely the SUBMOD model and the E-model, have been outlined. As an example, the performance of the E-model together with an intrusive measurement approach as implemented in a commercially available monitoring device was described. It was shown that the parameters of wireline (PSTN/ISDN) networks such as loudness and echo can reliably be measured using the standard procedures implemented in the device. When particular characteristics of GSM-networks such as impulsive noise and interruptions are aimed at, both instrumental measurement as well as the E-model currently show limitations. Especially in the light of modern systems involving terminal equipment other than handset-telephones and transmission techniques potentially effected by time-varying distortions, such as GSM, UMTS or VoIP, new measurement procedures have to be developed and the quality models (i.e. the E-model) have to be significantly enhanced.

Acknowledgement

The present work has been performed at IKA, Ruhr-University Bochum (Prof. J. Blauert, PD U. Jekosch).

Literature

Fletcher, H. and Munson, W. (1937). Relation between Loudness and Masking. J. Acoust. Soc. Am., 9(1), 1-10.

ITU-T Rec. G.107 (2000). The E-Model, a Computational Model for Use in Transmission Planning. International Telecommunication Union, CH-Geneva.

ITU-T Rec. P.562 (2000). Analysis and interpretation of INMD voice-services measurements. International Telecommunication Union, CH-Geneva.

ITU-T Suppl. 3 to P-Series Rec. (1993). Models for Predicting Transmission Quality from Objective Measurements. International Telecommunication Union, CH-Geneva.

ITU-T SG12 Delayed Contribution D.71 (1995). Impairment factors for speech clipping. Source: Ellementel, Sweden. International Telecommunication Union, CH-Geneva.

Johannesson, N.O. (1997). The ETSI Computation Model: A Tool for Transmission Planning of Telephone Networks. IEEE Comm. Magazine, Jan., 70-79.

Möller, S. (2000). Assessment and Prediction of Telephone Speech Quality. Kluwer Academic Publishers, Boston-USA.

Möller, S. and Raake, A. (2002). Telephone Speech Quality Prediction: Towards Network Planning and Monitoring Models for Modern Network Scenarios. Accepted in Speech Communication.