Monitoring Speech Quality in Telephone Networks: An Overview

Sebastian Möller

Institut für Kommunikationsakustik (IKA), Ruhr-Universität Bochum, D-44780 Bochum
moeller@ika.ruhr-uni-bochum.de

Summary

Monitoring speech quality in telephone networks is a two-step procedure. First, all data which are relevant for speech quality have to be measured during network operation. Then, models can be defined which calculate quality indices on the basis of the measurements. This talk will give an overview of monitoring approaches related to speech communication quality, which is one component of the quality of a service (QoS). Intrusively or non-intrusively measured data forms an input to different types of quality prediction models. Such models are based on the perceptively based comparison between input and output signals of the transmission line (signal-based models, e.g. PESQ, TOSQA, etc.), or on empirically determined relations between network planning parameters and user quality judgements (network planning models, e.g. the E-model or SUBMOD). They produce an estimate for certain quality dimensions which depend on the type of input parameters taken into account by the measurements. Various combinations between input parameters and prediction models are discussed, and limitations of the available methods are pointed out.

1. Introduction

Speech communication quality in telephone networks has been a topic of interest for several decades. Network operators who aim at providing an adequate quality to their users have to respect guidelines for the individual parameters which are related to quality aspects, such as loudness ratings of the connection and the terminals, echo attenuation and delay, or quantizing distortion. Auditory tests under controlled laboratory conditions open the possibility to investigate the users' experience of the overall quality of the connection, or of individual quality dimensions.

Due to the liberalized telecommunication market, interconnected networks become more and more frequent, showing characteristics which differ (both physically and perceptually) from the ones of traditional PSTN/ISDN networks. Examples are mobile networks with time-varying connection characteristics and acoustically different terminals, or voice over internet protocol (VoIP) connections impaired by packet loss and variable delays. In such interconnected networks, it is not sufficient to regard the perceptive effects of individual parameters on perceived quality. Rather, a combination of degradations is responsible for the overall (integral) quality.

In order to determine the overall quality of such interconnected networks, quality prediction models have been developed. They perform a transformation of instrumentally measurable network and user interface characteristics on quality judgments which users of the connection are expected to give. Two types of models have to be distinguished:

Signal-based models perform a comparison of input and output signals of a part of the transmission channel, and transform the observed differences or similarities – using knowledge about the human auditory system – into estimated user judgments. They have mainly been developed for predicting the effects of non-waveform codecs on speech quality in a listening-only situation, but newer versions also partly cover the degradations resulting from ambient noise and transmission errors. Signal-based models can be used e.g. for defining and optimizing new codecs, in cases where it is not possible to carry out auditory tests due to cost and time considerations. Their predictions normally reflect the listening-only situation, and are influenced by the choice of tests which have been used for the model definition. Examples are the PESQ model recommended by the ITU-T (ITU-T Rec. P.862, 2001), or the TOSQA model (Berger, 1998).

A different type are *parametric models* normally used for network planning. They start from planning values (which can also be measured instrumentally), e.g. loudness ratings, delay times, or noise levels, and perform a transformation on a quality scale which can be related to auditory user judgments. Such models predict the impact of perceptively diverse degradations (loudness, circuit and ambient noise, echo, pure delay) on quality, and refer also to the conversational situation. Examples are the E-model (ITU-T Rec. G.107) which relies on scalar parameters, or the SUBMOD model which makes use of frequency characteristics (ITU-T Suppl. 3 to P-Series Rec., 1993). Because the input parameters are planning values, parametric models can be used during the planning phase, before an actual network has been set up. On the other hand, this fact may lead – if the parameters are estimated incorrectly – to a certain degree of impreciseness, because the implemented connection may differ to a certain extent from the one expected during planning.

In existing networks, it is possible to directly measure input parameters, and to base quality predictions on instrumental measurement results. A new class of models has recently been developed for this purpose. These models are still strongly related to the model types discussed before. They can be used for automatically monitoring speech quality in telephone networks. In this way, network operators can detect and localize sources of degradations, and take appropriate measures in order to ensure good network performance over time. In the remainder of this paper, an overview is given of the measurement techniques which are at the basis of these models (Section 2), and the combination possibilities between measurements and prediction models (Section 3). An outlook discusses some limitations of the current approaches, and points at open research topics.

2. Measurement of Input Signals and Parameters

Two possibilities exist for instrumentally measuring network characteristics. One way is to set up test connections in the network, and to measure the applied input signal and the output signal of a part or of the whole transmission chain. This approach is called *intrusive*, because it requires purposely set up test connections. Thus, measurements are not based on real-life data. Both signals are recorded and transmitted to an evaluation unit. In this way, a number of parameters can be identified, namely

- the attenuation/frequency distortion of the channel (from this, loudness ratings can be calculated)
- the level of stationary or impulsive noise, including its frequency characteristics and the frequency of impulses
- the amount of signal-correlated noise
- the talker echo path loss and delay (provided that both directions of the connection are available)
- the listener echo attenuation and delay
- the pure one-way delay
- speech clipping and interruptions/fading of the channel

These parameters can e.g. be used as an input to the parametric models, see below.

The second way for measuring network characteristics is to perform real-life recordings of actual speech signals, and analyze them. This is typically done at a specific point of the network (potentially for both paths of the connection), without interrupting or disturbing the connection. Therefore, the approach is called *in-service, non-intrusive measurement*, and devices for such measurements (called INMDs) are available for both wirebound and mobile networks. The measurements are in principle more limited with respect to the characteristics which can be assessed, because they have to rely on the available signals, covering only a part of the connection (namely the part up to the measurement device). They have, however, the advantage of being performed under realistic conditions, and allow parameters to be accessed which are only observable in this situation (e.g. real speech and noise levels, percentage double-talk, etc.). A large amount of data can be made available in this way, and it is mandatory to process them automatically. Parameters which can be measured include

- the active speech level, and the speech activity factor
- the levels of stationary and non-stationary noise, resulting both from the connection and from the background at the talker's side
- the talker echo path loss and delay
- the percentage of single- and double-talk
- eventual temporary loss of the transmission
- estimated front-end clipping and amplitude clipping

ITU-T Rec. P.561 (1996) provides detailed information on the measurement of these parameters.

3. Monitoring Approaches

Monitoring models can make use of either type of measurement, and combine them with the available quality modeling approaches in order to provide quality estimates for the connection under consideration. The situation is depicted in Fig. 1. At least 4 different combinations are possible.

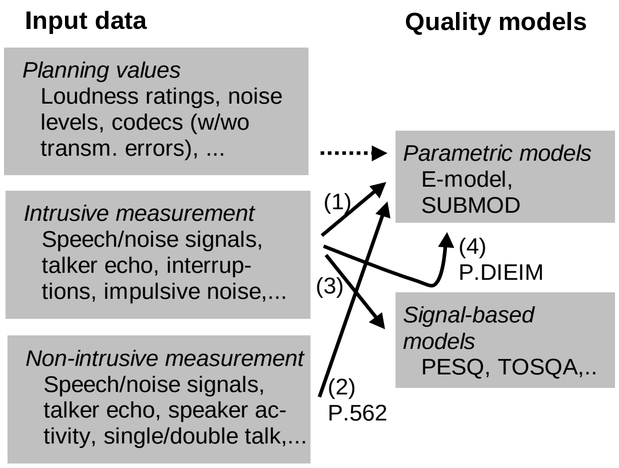

Figure 1: Approaches for quality monitoring.

One possibility (combination 1) is to use intrusive measurements in order to determine the input parameters of the parametric (network planning) models. As most of the models' input parameters are defined to be measured in an intrusive way, this combination is obvious. However, some parameters are used as an input which cannot be directly measured, e.g. the type of codec in the network. In addition, intrusive measurements are often limited to the electrical path, because the acoustic user interfaces (of real users) are not accessible to the network operators. In these cases, expected or planning values have to be used instead, indicated by the dashed line in Fig. 1.

A different approach (2) starts from non-intrusive measurements, which can be mapped to the parametric models. ITU-T Rec. P.562 (2000) defines this mapping for two such models, namely the so-called Call Clarity Index, CCI (which is closely related to the SUBMOD model) and the E-model. Whereas the CCI approach directly makes use of the measured speech and noise spectra, these inputs have to be reduced to one-dimensional scalar values for the E-model.

A third method is to use intrusively measured input and output signals of a part of the transmission channel as an input to signal-based models. This approach is obvious, but it carries the risk that degradations are to be predicted which have not been taken into account for the model development. The approach has recently also been applied to the whole transmission chain including the user interfaces, thus for acoustic-to-acoustic measurements.

Following this approach, another possibility (4) is to estimate input parameters for the network planning models via signal-based models. E.g., the degradations due to low bit-rate codecs can be well predicted with these models, and then be combined with other impairments (e.g. conversational ones) via the E-model. A methodology with this aim is currently under discussion within the ITU-T (ITU-T Draft Rec. P.DIEIM, 2002).

4. Comparison and Outlook

Due to the large number of potential measurement methods and quality prediction models, it is difficult to judge which approach is suited best for validly and reliably predicting quality in real-life networks. Only very few investigations are reported in this respect, and they are often limited to justify specific types of models.

The choice will partly depend on the possibilities given for input signal or parameter measurement. In addition, there are degradations which may be better predicted by one or the other approach. E.g., the perceptive effects of non-linear codecs are predicted in a relatively accurate way by signal-based models, whereas network planning models like the E-model take them into account only in a very simplified manner (the SUBMOD model not at all). On the other hand, the E-model covers the widest range of input parameters, and gives predictions also for the conversational situation. New signal-based models have recently been reported to predict part of these effects as well, but they still have to be validated.

Several limitations exist for the types of degradations which can be measured and predicted by the models. E.g., via instrumental measurements, it is still difficult to identify the codec(s) used in the connection. User terminals other than handsets are not covered by the existing quality prediction models. These are, however, factors which carry a major influence on the perceived quality. The validity of quality monitoring approaches will thus largely depend on how the current limitations can be overcome.

Acknowledgement

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

References

Berger, J. (1998). *Instrumentelle Verfahren zur Qualitätsschätzung - Modelle auditiver Tests,* Dissertation, Christian-Albrechts-Universität, D-Kiel.

ITU-T Draft Rec. P.DIEIM (2002). *Methodology for the Derivation of Equipment Impairment Factors from Instrumental Models*, Contribution to ITU-T SG 12 Meeting (Author: S. Möller), 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 Rec. G.107 (2000). *The E-Model, a Computational Model for Use in Transmission Planning,* International Telecommunication Union, CH-Geneva.

ITU-T Rec. P.561 (1996). *In-Service Non-Intrusive Measurement Device - Voice Service Measurements*, 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 Rec. P.862 (2001). *Perceptual Evaluation of Speech Quality (PESQ), an Objective Method for End-to-End Speech Quality Assessment of Narrowband Telephone Networks and Speech Codecs*, International Telecommunication Union, CH-Geneva.