3rd International Workshop on Perceptual Quality of Systems (PQS 2010)
6-8 September 2010, Bautzen, Germany

# Extending the E-Model to Better Capture Terminal Effects

*Sebastian Möller*[1], *Frank Kettler*[2], *Hans-Wilhelm Gierlich*[2], *Nicolas Côté*[3], *Alexander Raake*[4], *Marcel Wältermann*[1]

[1] Quality and Usability Lab, Deutsche Telekom Laboratories, TU Berlin, Germany
[2] HEAD Acoustics GmbH, Herzogenrath, Germany
[3] Ecole Nationale d'Ingénieurs de Brest, France
[4] Assessment of IP-based Applications Lab, Deutsche Telekom Laboratories, TU Berlin, Germany
{sebastian.moeller, alexander.raake, marcel.waeltermann}@telekom.de,
{frank.kettler, h.w.gierlich}@head-acoustics.de, nicote@free.fr

## Abstract

In this paper, we addres s the param etric prediction of speech quality for planning telecom munication networks. More specifically, we propose an extension to the E-model recommended by the Internati onal Telecommunication Union in order to better capture the effects of signal processing equipment integrated in the terminals. We propose to use two additional parameters for desc ribing the effects of noise reduction, namely the signal- to-noise ratio improvement during speech and the total noise level reduction. Furthermore, we propose to introduce an additional equipm ent im pairment factor for describing the speech degradations resulting from imperfect noise reduction. Th e proposed extensions are analyzed on the basis of data collected with seven modern handsets, and directions for future work are discussed.

**Index Terms**: speech quality prediction, E-m odel, term inals noise reduction

## 1. Introduction

In order to design an optimum se rvice for their users, planners of telephone networks need to know about the perceptual quality of their future networks prior to im plementation. For this purpose, transmission pl anning models have been developed which estimate the overall quality of the speech link from parameters describing a ll relevant characteristics of the involved network and term inal components. The most popular model is the E-mode l developed within ETSI [1][2] and s tandardized in its m ost recent version by the Telecommunication Standardization Sector of the International Telecommunication Union (ITU-T) in ITU-T Rec. G.107 [3].

The E-model estimates the overall quality mouth-to-ear in a conversational situation on the basis of so-called "impairment factors". S uch factors quantify the degradations which are expected simultaneously to the s peech s ignal (*Is*), delayed with res pect to the speech signal (*Id*), or resulting from non-linear and/or time-variant characteristics *(Ie,eff*) on the s o-called " transmission rating s cale". On this s cale, impairment factors are s ubtracted from the bas ic s ignal-to-noise ratio *Ro* which is determined by power summation of the equivalent weighted noise power levels of all noise sources (circuit nois e, nois e floor of the receiver's s ubscriber line, background noises at send and recei ve side). Thus, the overall transmission rating *R* can be calculated as

$$R = Ro - Is - Id - Ie,eff + A \quad (1)$$

The s o-called " expectation factor" *A* represents the quality advantage sometimes observed in particular situations such as mobile telephony or calls to hard-to-reach areas; it compensates for half of the particular degradation expected in this scenario[1].

Whereas the current version of the E-m odel accurately predicts most perceptual effects associated with the network, it is less precise when it comes to the effects of m odern terminal equipment, such as handset s involving noise reduction and echo cancellation, or even hands -free term inals and headsets. Networks planners often unde restimate the influence of terminals on the overall quality of com munication. It is typically disregarded that speech quality is m ainly determined by the acous tic interfaces and the implemented signal processing. This severely limits the usefulness of the E-model.

In the current E-m odel vers ion, term inal effects are captured as follows:

- The effect of background noise at the send side is captured in the basic signal-to-noise ratio *Ro*. In the calculation of *Ro*, the acoustic background noise on the send side ( *Ps*, measured in dB(A)) is attenuated by the send-side sensitivity of the handset (Send Loudness Rating, *SLR*) and the s o-called *D*-factor which takes into account the sensitivity differences of the m icrophone for direct (speech) vs. indirect (background) sound, cf. Eq. (2):

$$Nos = Ps - SLR - Ds - 100 + 0.004(Ps - OLR - Ds - 14)^2 \quad (2)$$

  The resulting value *Nos* is the electric equivalent of the room noise at send side; it is power-summed with the other noise sources in the calculation of *Ro*. Whereas the exact source of this formul a is unknown, it shows that noise reduction present in many handset and hands-free type terminals is not taken into account by the model.
- The effects of echo cancellers are not explicitly captured by the current model. The model only takes into account the residual echo in its calculation of *Id*, by means of the talker echo impairment factor *Idte* which is calculated from the echo delay *T* (in ms), the Talker Echo Loudness Rating *TELR* (in dB), and the masking effects of the

[1] As the number of mobile telephones has outperformed the number of fixed phones in many industrial countries, the applicability of the expectation factor has been put into question; therefore, we disregard this factor in the following analyses, setting $A = 0$.


10.21437/PQS.2010-11

noise (via *Ro*). Effects of level-switching devices are not captured at all.

- The effects of codecs included in the terminal are captured by the effective Equipment Impairment Factor *Ie,eff*; this can be calculated from the results of listening-only tests or from measurements with full-reference models like PESQ [4], following the methodologies of ITU-T Rec. P.833 [5] and P.834 [6].
- The processing delay on the transmission path is captured by the impairment factor for too long delay, *Idd*, which is part of *Id* and calculated from the absolute delay *Ta* (in ms).

Obviously, the effects of imperfect noise reduction and echo cancellation are not yet adequately captured by the model.

In this paper, we make a first approach to better predict the effects linked to noise-reduction algorithms integrated in modern terminals. The approach is presented in Section 2. In Section 3, we perform a first validation of the approach on the basis of measurements obtained from seven state-of-the-art handsets. We critically discuss the results in Section 4 and provide an overview of future work which is necessary to fully capture terminal effects in the E-model in Section 5.

## 2. Approach

As a step forward, we propose the following 5-step intermediate solution:

1. The residual noise resulting from imperfect background noise reduction may occur either during speech intervals or during pauses; parameters describing these two situations are defined in ITU-T Rec. G.160 [7], namely *SNRI* (the SNR improvement during speech in dB) and *TNLR* (the total noise level reduction in dB). We propose to use a weighting of half and half (corresponding to roughly 50% speech activity) for the speech and silence parts and change Eq. (2) as follows:

$$Nos = Ps - SLR - Ds - 0.5(SNRI - TNLR) - 100 + 0.004(Ps - OLR - Ds - 14)^2 \quad (3)$$

   This amendment is meant to capture the effect of residual noise of the noise reduction mechanism.
2. The effects of speech degradation from imperfect noise reduction can be captured by estimating the speech quality associated with the terminal with the objective model of ETSI EG 202 396-3 [8]. This model provides an estimation of the S-MOS associated with the degraded speech signal alone; the S-MOS can then be translated to an additional equipment impairment factor *Ie,nr* reflecting the noise reduction equipment, following the procedure of ITU-T Rec. P.834 [6]. Ideally, the test would be made with the normalization procedure and the reference conditions given in that recommendation; however, as this will frequently not be possible, we expect that the raw *Ie,nr* values will also be meaningful.
3. The effects of residual echo are taken into account in the standard way of the E-model, i.e. via the talker-echo impairment factor *Idte*; the attenuation of the residual echo path has to be used for the calculation of *TELR* at this stage.
4. The effects of speech degradation from imperfect echo cancellation can be estimated via an additional equipment impairment factor *Ie,ec* which is calculated with the help of the procedure of ITU-T Rec. P.834 [6], using the instrumental model of ITU-T Rec. P.862 [4]. Both *Ie,nr* and *Ie,ec* are added to the effective equipment impairment factor *Ie,eff* before calculating the overall transmission rating *R*.
5. The effects of delay are captured in the usual way via *Idd*.

With these modifications, we hope to cover the most important effects related to noise reduction and echo cancellation integrated in modern terminals. Linear distortions related to hands-free terminals and the acoustic situation in the sending room mainly result in a coloration of the speech signal; these can e.g. be described in terms of a bandwidth impairment factor, as it has been proposed in [9][10]. No approaches have yet been made to consider the effects of imperfect voice activity detection or comfort-noise insertion.

In the following paragraphs, we will present analyses relating to the first two points of our approach, namely the consideration of background noise and its reduction in the terminal. These analyses will show the usefulness, but also the limitations of the proposed approach with respect to background noise transmission and reduction. The effects of imperfect echo cancellation will be addressed in follow-up analyses.

## 3. Experimental analysis

For testing points 1 and 2 of our approach, auditory and instrumental data from seven state-of-the-art handsets including noise reduction algorithms were available to us. These data consist of noisy speech files which have been judged upon by listeners, and which have been analyzed in order to extract the parameters needed for the E-model calculations, in the way described below.

### 3.1. Speech and noise data

Noisy speech recordings from the handsets have been generated using a Head And Torso Simulator (HATS) HEAD acoustics Type HMS II.3 in accordance with ITU-T Rec. P.58 [11], following the method described in ITU-T Rec. P.310 [12] and using an artificial ear of Type 3.4 in accordance with ITU-T Rec. P.57 [13]. Speech signals consisted of 8 sentences of 4 speakers (2 male, 2 female, 2 sentences each, native English) and were reproduced by the HATS in order to obtain an active speech level of -1.7 dBPa; this level is 3 dB higher than the usual level in quiet, and was selected to roughly simulate the Lombard effect, i.e. the effect that the talker speaks up in a noisy environment. Background noise of five different types was generated in the recording room in a diffuse way, using a four loudspeaker arrangement plus subwoofer as described in ETSI EG 202 396-1 [14] grouped around the device. The noise files were taken from the ETSI EG 202 396-1 database [14]. The noise level was determined at both ears of the HATS prior to mounting the handset, with the following results:

- Stationary *car* noise recorded in a luxurious car driving at a constant speed of 130 km/h: 70.4 dB(A) at left ear, 69.8 dB(A) at right ear.
- *Road* noise: 74.9 dB(A) at left ear, 74.0 dB(A) at right ear.
- *Cafeteria* noise similar to speech babble: 64.0 dB(A) at left ear, 62.3 dB(A) at right ear.

- Noise recorded at the pavement of a *crossroad* section: 69.1 dB(A) at left ear, 69.6 dB(A) at right ear.
- Noise recorded at a business *office*: 56.6 dB(A) at left ear, 57.8 dB(A) at right ear.

As it can be seen both, noise types and levels are quite diverse and reflect the multitude of noises to be encountered in real-life mobile telephony situations in an appropriate way.

### 3.2. Auditory test

The noisy speech files have been judged upon by 24 listeners (average age 29 years, no reported hearing impairment) in a quiet room according to ITU-T Rec. P.835 [15]. The selection of test participants and the listening environment mainly followed ITU-T Rec. P.800 [16]. The speech samples were IRS receive filtered [17] and played back for the listeners via headphones (diotic representation). The listening level was adjusted to -21 $dB_{Pa}$ ASL corresponding to 73 $dB_{SPL}$ ASL.

The judgments of the listeners have been averaged in order to obtain a rating for the speech quality (S-MOS), a rating for the noise quality (N-MOS), and a judgment for the overall quality (G-MOS). Although the test set-up deviates from the pure P.800 setting (which was not developed for background noise scenarios), the G-MOS value is the one which comes closest to the overall quality rating of ITU-T Rec. P.800 and should consequently be predicted by the E-model.

### 3.3. Determination of input parameters

From the recorded noisy speech files and the experimental settings, a number of input parameters to the E-model have been determined. These include:

- The level of background noise at the send side, *Ps*, in dB(A): This was calculated as the arithmetic mean value of the values observed for the left and the right ears of the HATS, see Section 3.1.
- The Send Loudness Rating *SLR*, in dB: This was measured for each handset according to ITU-T Rec. P.79 [18].
- The Receive Loudness Rating *RLR*, in dB: This value was set in order to best reflect the characteristics of the auditory test: In a pre-test, participants were allowed to adjust the level of noisy speech files in order to obtain optimum results; this lead to the 79 $dB_{SPL}$ ASL for diotic representation as mentioned above. The pre-test will have resulted in an "optimum" speech level with respect to the E-model. The E-model assumes that this optimum is reached with an Overall Loudness Rating *OLR* of 10 dB [2]; thus, we set *RLR* = 10 dB – *SLR* and used the values of *SLR* which have been measured for each handset individually.
- The Signal-to-Noise Ratio Improvement *SNRI*, in dB: This was determined from the noisy speech files according to the procedure described in ITU-T Rec. G.160 [7]. We thus obtain an *SNRI* value for each handset and noise condition. This value only reflects the improvement by the noise reduction algorithm. The reference signal is the noisy speech recorded at the terminal's microphone position.
- The Total Noise Level Reduction *TNLR*, in dB: As *SNRI*, this was determined according to ITU-T Rec. G.160 [7] for each handset and background noise condition. As for *SNRI*, only the effect of the terminal's noise reduction algorithm is taken into account.
- The Equipment Impairment Factor *Ie*: Handsets 1 and 5 were 2G handsets and used the GSM-FR codec for which *Ie* = 20 is defined in ITU-T Rec. G.113 [19]; the other handsets were 3G handsets and used the AMR-NB codec at 12.2 kbit/s which is equivalent to the GSM-EFR codec; for the latter, *Ie* = 5 is defined in ITU-T Rec. G.113 and was used for the calculations.
- The difference of the sensitivities for the direct (speech) sound and the diffuse (background noise) sound, the so-called *Ds* factor of the sending handset: For handsets including noise reduction algorithms, the sensitivity – when measured from the point-of-view of the resulting electrical signal – depends on the noise reduction algorithm, and consequently on the type and level of the noise. The E-model, however, has mainly been optimized for a default setting of this factor, namely *Ds* = 3. In order to address this issue, three different values for *Ds* have been used for each handset: (1) the default value *Ds* = 3; (2) a value measured with stationary pink noise at *Ps* = 70 dB(A) which is similar in character to the car noise of Section 3.1, termed *Ds,pink* in the following; and (3) a value measured with the cafeteria-type noise described in Section 3.1, termed *Ds,café* in the following. Whereas the first setting mostly reflects the E-model optimization procedure, the second and third value reflect the capacities of the noise reduction algorithm integrated into the handset for stationary and non-stationary noises, respectively.

The other input parameters of the E-model were mostly set to their default values given in Table 1 of ITU-T Rec. G.107 [3], namely:

- Delay times *T* = *Ta* = *Tr* = 0 ms
- Talker Echo Loudness Rating of the talker echo *TELR* = 65 dB
- Weighted Echo Path Loss of the listener echo *WEPL* = 110 dB
- Sidetone Masking Rating for the speech sound *STMR* = 15 dB
- Listener Sidetone Rating for the background noise sound *LSTR* = *STMR* + *Ds*, using the values of *Ds* described above
- Circuit Noise *Nc* = -70 dBm0p
- Noise floor of the receiving subscriber line *Nfor* = -64 dBm0p
- D-factor of the receiving terminal *Dr* = 3
- Room noise at the receiving end *Pr* = 35 dB(A)
- Quantizing noise in terms of quantizing distortion units *qdu* = 1.
- Packet loss rate *Ppl* = 0 and packet-loss robustness factor *Bpl* = 1 corresponding to no packet loss
- Expectation factor *A* = 0

These values were used for the calculations cited hereafter.

### 3.4. Predictions of residual background noise effects

In order to test step 1 of our approach described in Section 2, we compare four different "versions" of E-model predictions to the auditory G-MOS obtained from our test participants. The four versions are as follows:

- The original E-model, as given in ITU-T Rec. G.107 [3], without any modifications. This version is considered as a baseline for the verification of the approach.
- The modified version of the E-model, replacing Eq. (2) by Eq. (3), and using the default setting of the parameter *Ds* = 3.

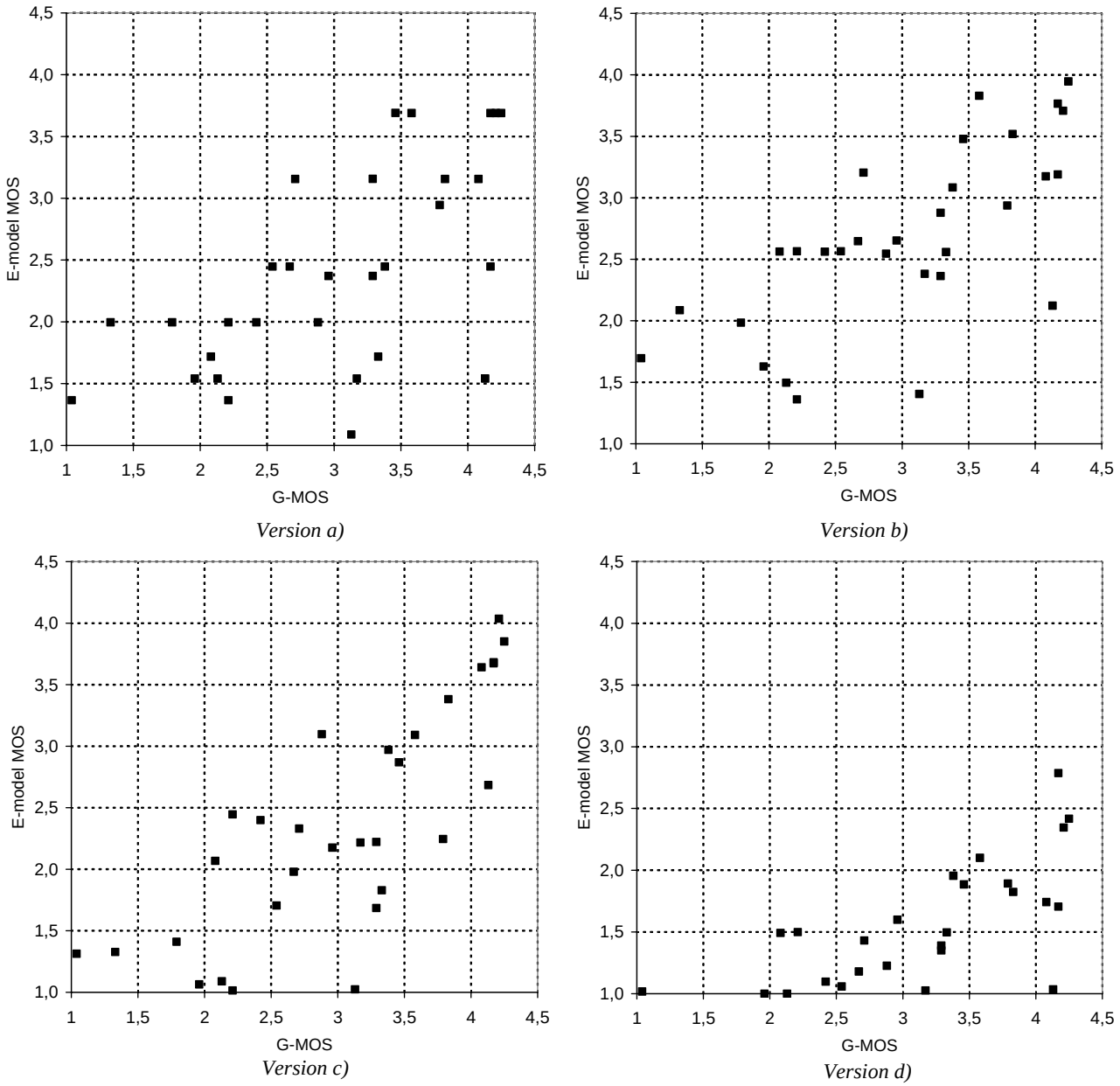


Figure 1: *Comparison of auditory* G-MOS *scores with E-model predictions.*

- The modified version of the E-model as in b), but using the *Ds,pink* value derived from the measurement with a stationary background noise.
- The modified version of the E-model as in b), but using the *Ds,café* value derived from the measurement with a non-stationary background noise.

Figure 1 shows scatter plots of the auditory G-MOS scores compared to the estimated E-model scores for each of the E-model versions. The corresponding Pearson correlation coefficients *r* and Root Mean Squared Errors *RMSE* are listed in Table 1.

The original version of the E-model shows a relatively poor correlation to the auditory test data, and a considerable prediction error. Although most of the scatter plots are below the diagonal line (indicating that the E-model is too pessimistic in its estimations), there are also a number of points where the model is too optimistic. This is a significant shortcoming for a network planning model which should generally aim for a worst-case estimation if accurate predictions cannot be reached.

Table 1*: Correlations and prediction errors for different E-model versions*

| E-model version | Result | |
|---|---|---|
| | *r* | *RMSE* |
| a) original | 0.62 | 0.95 |
| b) improved with *Ds* | 0.70 | 0.72 |
| c) improved with *Ds,pink* | 0.78 | 0.88 |
| d) improved with *Ds,cafe* | 0.71 | 1.65 |

The predictions of the modified version b) of the model are considerably better, both in terms of correlation and prediction error. Figure 1 shows that in particular the too negative estimations are corrected by the modification of Eq. (3). However, there are even more too optimistic estimations in this version compared to version a).

Replacing the default value $Ds = 3$ by the value measured using stationary pink noise significantly improves the correlation to the auditory G-MOS values. The prediction error increases compared to version b), but it is still better than the one of the original version a). The scatter plot in Figure 1 shows that there are almost no too optimistic estimations of the model; nearly all points are below the diagonal line.

Version d) of the model shows that the estimations are very sensitive to the way how the *Ds* factor is measured. Using non-stationary cafeteria noise, the values of *Ds,cafe* are significantly lower than the ones measured using stationary pink noise, *Ds,pink*. As a consequence, the E-model estimations get significantly more pessimistic. This can be clearly seen in the lower right panel of Figure 1, where the data points lay close to the x-axis. Although the correlation is higher than for version b), the estimations are far too pessimistic, as it is indicated by the large prediction error in Table 1.

### 3.5. Predictions of speech degradations resulting from imperfect noise reduction

The modifications made so far address the impact of residual background noise after the noise reduction algorithm implemented in the terminal. However, an imperfect noise reduction algorithm might also degrade the speech signal. For example, an improperly adjusted spectral subtraction algorithm might subtract too much energy from the speech spectrum, thus resulting in lower intelligibility and speech quality.

Unfortunately, the speech files with no background noise have not been assessed in the auditory test. However, we have processed the speech files recorded through the respective handsets in quiet by a different speech quality algorithm which is expected to reflect terminal-side speech quality degradations, the Telecommunications Objective Speech Quality Assessment (TOSQA) developed by Berger [20] in its 2001 version [21]. From the raw TOSQA MOS estimation (T-MOS), we derived an estimation of the maximum *R* value which can be expected in quiet, using the relationship between MOS and *R* given in the E-model [3]. Then, we used the speech-quality-related judgment of the auditory test (S-MOS) to derive the *R* value of the degraded speech signal, again using the E-model relationship. The noise-reduction impairment factor *Inr* is then calculated as the difference between these two *R* values in the following way:

$$Inr = min\left(R(\text{T-MOS}) - R(\text{S-MOS}), 0\right) \quad (4)$$

The ceiling at zero avoids that small differences between the auditory S-MOS score and the instrumental T-MOS score impact the estimated *Inr* values.

The noise-reduction impairment factor *Inr* was subtracted from the *R* value for all four versions of the E-model calculations described in Section 3.4. The resulting overall *R* value has been transformed back to the MOS area, and correlations and prediction errors to the subjective G-MOS have been calculated. These results are summarized in Table 2.

Compared to the original version of the model (Table 1, version a), the correlation between auditory G-MOS and E-model estimations improves considerably; however, also the prediction error is higher. With respect to the improved E-model versions (b to d), the inclusion also leads to increased correlations in all cases, however again at the expense of an increased prediction error.

Table 2: *Correlations and prediction errors for different E-model versions including the noise-reduction impairment factor* Inr

| E-model version | Result | |
|---|---|---|
| | *r* | *RMSE* |
| a) original | 0.72 | 1.12 |
| b) improved with *Ds* | 0.82 | 0.85 |
| c) improved with *Ds,pink* | 0.86 | 1.01 |
| d) improved with *Ds,cafe* | 0.73 | 1.71 |

## 4. Discussion

In order to better capture terminal effects in the E-model, we proposed modifications for noise reduction and echo cancellation algorithms in Section 2. Based on noisy speech data which has been recorded with seven different handsets and five different types and level of background noise, an auditory test has been performed in the frame of ETSI activities which was available to us. We used the speech data in order to analyze the improvements proposed in steps 1 and 2 of Section 2, namely the consideration of residual background noise by means of the parameters *SNRI* and *TNLR* in Eq. (3), and the consideration of degradations caused by the noise reduction algorithm on the speech signal itself by means of an additional impairment factor *Inr* given in Eq. (4). The echo cancellation effects could not be analyzed with the available data.

Different ways have been followed in order to derive the necessary E-model input parameters. As far as possible, measured values have been taken, as these reflect best the situation captured in the auditory test which is considered as the reference here. Where this was not possible, we used the default values of the E-model which are given in Table 1 of ITU-T Rec. G.107 [3]. In particular, we considered three different values for *Ds*: The default value of the model, a value measured using stationary pink noise, and a value measured using un-stationary cafeteria noise.

The results show that the prediction accuracy improves considerably by using the parameters *SNRI* and *TNLR* in Eq. (3). In all cases, the correlation to the auditory G-MOS values increases, however in one case (using *Ds,café*) at the expense of an increased prediction error. Subtracting the *Inr* parameter from the resulting *R* value to also reflect the speech-signal-related degradations caused by the noise reduction algorithm increased the correlation again, however here the prediction error also increases. Overall, the combination of both E-model amendments (steps 1 and 2) and a measurement of *Ds* using stationary pink noise lead to a relatively high correlation and a moderate prediction error for both analyses. A drawback of this method is that *Ds* contains both the SNR gain/loss provided by the terminal and the one provided by the noise canceller which could not be deactivated during the tests.

It has to be emphasized that the noise-reduction related impairment factor *Inr* has been extracted from the results of another instrumental model, namely TOSQA, without any normalization procedure. It would have been preferable to use auditorily-derived values instead, and to perform some normalization (e.g. the one described in ITU-T Rec. P.833 [5]); however, the quiet conditions were not part of the auditory test, so this could not be performed in our analysis.

## 5. Future work

We proposed 5 steps in order to better capture terminal effects in the E-model. Steps 1 and 2 relate to the consideration of

residual background noise and sp eech degradations resulting from imperfect noise reduction. The proposed modifications following these steps lead in all cas es to a higher correlation with auditory test results, however sometimes at the expens e of a higher prediction error. This effect has to be further analyzed with additional test data.

The modifications proposed for echo cancellers could not be analy zed with the data which was available to us. We expect that s peech degradations due to im perfect echo cancellation will also be predictable through an additional impairment factor *Iec*. It rem ains to be seen how this impairment factor can be dete rmined: Either using auditory test results, or using one of th e available instrumental models. Besides PESQ [4] and TOSQA [21], also the new upcoming ITU-T standard P. OLQA [22] which is expected to replace PESQ is an interesting candidate , as it is expected to better reflect s peech-processing-related degradations than PESQ does. As soon as the ITU-T standa rd gets available, we would like to carry out new analyses using this model.

Step 5 of the proposed amendmen ts relates to the effect of delay introduced by s peech proces sing integrated into the terminal. The prediction of delay by the current version of the E-model is a subject of con tinuous dispute in ITU-T Study Group 12. It is assumed that ne w subjective test methods are needed in order to capture the delay-related degradations in a more realistic way.

Another aspect of the term inal is the sound quality which might suffer due to the acoustic characteristics of the terminal and the recording room . S uch effects m ight either be addressed by the directness-related impairment factor described in [10], or by the new POLQA model.

So far, we did not y et verify the approach for hands-free terminals. We also did not consider the term inal at the receive side, which is also expected to capture receive-side noise and speech and s hapes the incom ing speech signal via the integrated echo canceller. Thes e as pects need to be studied further before coming up with a complete version of the E-model capturing all term inal-related degradations in a reliable way. The described investigations are meant as a first step into this direction.

Finally the proposed approach could possibly be improved by reducing *Ds* to its acoustical component and considering the improvement achieved by the NR algorithm in the phone purely in the new *SNRI*/*TLNR* term, see Eq. (3). Furthermore the dynamic structure of the background noise signal could be considered in the calculation of *Nos* and *Ie,nr* because it will lead to differences in the perceived quality.

## 6. Acknowledgements

The authors are grateful for the support from their colleagues in generating this speech data, and in particular to Jan Reimes from HEAD acoustics for his help in the analy sis and the determination of the E-model input parameter values.

## 7. References

[1] Johannesson, N. O., “The ETSI Computation Model: a Tool for Transmission Planning of Telephone Networks”, IEEE Communications Magazine, 35(1):70–79, 1997.

[2] ETSI ETR 250, “Speech Com munication Quality from Mouth to E ar for 3, 1 kHz Handset T elephony acr oss Networ ks”, European Telecom munications Standards Institute, Sophia Antipolis, 1996.

[3] ITU–T Rec. G.107, “The E-Model, a Computational Model for Use in T ransmission Planning”, I nternational Telecommunication Union, Geneva, 2009.

[4] ITU–T Rec. P.862, “ Perceptual Evaluation of Speech Quality (PESQ): An Objective Method for End-to-end Speech Quality Assessment of Narrow-band Te lephone Networks and Speech Codecs”, International T elecommunication Union, Geneva, 2001.

[5] ITU–T Rec. P.833, “Methodology for Derivation of Equipment Impairment Factor s Fr om Subjective L istening-only Tests”, International Telecommunication Union, Geneva, 2001.

[6] ITU–T Rec. P.834, “M ethodology for the Der ivation of Equipment I mpairment Factor s fr om Instrumental Models”, International Telecommunication Union, Geneva, 2002.

[7] ITU-T Rec. G. 160, “Voice Enhancement Devices”, International Telecommunication Union, Geneva, 2008.

[8] ETSI EG 202 396-3, “ Speech Qu ality Perform ance in the Presence of Backgr ound Noise – Par t 3: Background Noise Transmission – Objective T est M ethods”, European Telecommunications Standards Institute, Sophia Antipolis, 2007.

[9] Raake, A., “Speech Quality of VoIP – Assessm ent and Prediction”, Wiley, UK-Chichester, West Sussex, 2006.

[10] Wältermann, M., Raake, A., “ Towards a New E-Mo del Impairment Factor for L inear Distor tion of Nar rowband and Wideband Speech Transm ission”, in: Proc. IEEE Int. Conf . on Acoustics, Speech, and Signa l Processing (ICASSP 2008), 30 March - 4 April, Las Vegas NV, 4817-4820, 2008.

[11] ITU-T Rec. P. 58, “Head and T orso Simulator for Telephonometry”, I nternational T elecommunication Union, Geneva, 1996.

[12] ITU-T Rec. P.310, “Transmission Characteristics for Telephone Band ( 300-3400 Hz) Digital Telephones”, I nternational Telecommunication Union, Geneva, 2009.

[13] ITU-T Rec. P.57, “ Artificial Ears”, International Telecommunication Union, Geneva, 2009.

[14] ETSI EG 202 396-1, “ Speech and Multim edia Transm ission Quality (STQ), Speech Quality Pe rformance in the Presence of Background Noise, Par t 1: Backgr ound Noise Sim ulation Technique and Backgr ound Noise Database”, E TSI, Sophia Antipolis, March 2009.

[15] ITU-T Rec. P. 835, “Subjective T est M ethodology for Evaluating Speech Communication Systems That Include Noise Suppression Algor ithm”, I nternational T elecommunication Union, Geneva, 2003.

[16] ITU-T Rec. P.800, “M ethods for Subjective Deter mination of Transmission Quality”, International Telecom munication Union, Geneva, 1996.

[17] ITU-T Rec. P. 830, “Subjective Per formance Assessment of Telephone-band and W ideband Dig ital Codecs”, I nternational Telecommunication Union, Geneva, 1996.

[18] ITU-T Rec. P.79, “Calculation of L oudness Ratings for Telephone Sets”, I nternational T elecommunication Union, Geneva, 2007.

[19] ITU-T Rec. G.113, “ Transmission Impairments Due to Speech Processing”, I nternational T elecommunication Union, Geneva, 2007.

[20] Berger, J., „ Instrumentelle Verf ahren zur Sprachqualitätsschätzung – Modelle auditiver Tests”, Shaker, Aachen, 1998.

[21] ITU–T Contrib. COM 12–19, “ Results of Objective Speech Quality Assessm ent of Wideband Speech Using the Advanced TOSQA–2001”, International T elecommunication Union, Geneva, 2000.

[22] ITU-T T D 90Rev4 ( WP2/12), “Requir ement Specification for P.Objective Listening Quality Assessm ent (P.OLQA)”, Rapporteur for Question 9/12, I TU-T SG12 M eeting, 22- 30 May 2008.