INTERSPEECH 2019
September 15–19, 2019, Graz, Austria

Extending the E-Model Towards Super-wideband and Fullband Speech Communication Scenarios

Sebastian Möller [1,2], *Gabriel Mittag* [1], *Thilo Michael* [1], *Vincent Barriac* [3], *Hitoshi Aoki* [4]

[1] Quality and Usability Lab, Technische Universität Berlin, Germany
[2] Deutsches Forschungszentrum für Künstliche Intelligenz (DFKI), Berlin, Germany
[3] Orange, Lannion, France
[4] Nippon Telegraph and Telephone Corporation, Japan

firstname.lastname@[tu-berlin.de|orange.com|hco.ntt.co.jp]

Abstract

In order to plan speech communication services regarding the quality experienced by their users, parametric models have been used since a long time. These models predict the overall quality experienced by a communication partner on the basis of parameters describing the elements of the transmission channel and the terminal equipment. The mostly used model is the E-model which is standardized in ITU-T Rec. G.107 for narrowband and in ITU-T Rec. G.107.1 for wideband scenarios. However, with the advent of super-wideband and fullband transmission, the E-model needs to be extended. In this paper, we propose a first version of an extended E-model which addresses both super-wideband and fullband scenarios, and which predicts the effects of speech codecs, packet loss, and delay as the most important degradations to be expected in such scenarios. Predictions are compared to the results of listening-only and conversational tests as well as to signal-based predictions, showing a reasonable prediction accuracy.

Index Terms: speech quality, parametric quality prediction, speech coding, packet loss, delay

1. Introduction

The prediction of speech quality is a requirement for efficiently planning speech communication services since the old days of telephony. In contrast to signal-based predictions which require speech signals to be measured either from simulations or in operating networks (such as the Perceptual Evaluation of Speech Quality, PESQ [1], or the Perceptual Objective Speech Quality Assessment, POLQA [2]), service planning is commonly done based on assumptions about service components, as signals are not readily available. These components are described in terms of qualitative or quantitative parameters, such as the type of codec used, the level of circuit noise or background noise, the attenuation and delay of a talker echo, etc. The parameters are then merged to form a prediction of the overall quality a user of the service would experience in a conversational situation. The mostly used such model is the E-model which is the only model recommended for such purposes by the International Telecommunication Union, ITU-T, in ITU-T Rec. G.107 [3].

The E-model was developed on the basis of older proprietary models to cover the effects of attenuation, circuit and ambient noise, non-optimum sidetone (i.e. back-coupling of the own voice in the terminal), talker and listener echo, pure delay, as well as digital coding at different bitrates in narrowband (300-3400 Hz, NB) analogue/digital telephony. It was later extended towards the effects of packet loss and packet discard occurring in packet-based transmission. A further step was the establishment of a wideband (100-7000 Hz, WB) version of the model, which is standardized as ITU-T Rec. G.107.1 since 2015 [4]. Still, the introduction of packet-based transmission in the backbone network gives rise to a number of scenarios (Voice-over IP, Voice-over-LTE) which do not require a limitation of the audio bandwidth, thus enabling super-wideband (50-14000 Hz, SWB) or fullband (0-20000 Hz, FB) transmission. For such scenarios, the current versions of the E-model are not applicable.

It is the aim of the present paper to propose an extension of the E-model towards SWB and FB services, and to initially check its performance on a limited set of data. The extension of the model is based on ongoing work in Study Group 12 of the ITU-T, which however has not yet led to a new recommended version. This work has led to the definition of a new maximum quality level for a future model, as well as proposals for the prediction of coding degradations with and without packet loss. In addition, we introduce a first attempt to model the effects of pure delay. We illustrate the performance of the proposed model on the basis of 5 listening-only and 2 conversation tests which were available to us, as well as by comparing to the estimations from POLQA.

The remainder of the paper is structured as follows: In Chapter 2, we review the algorithmic principles of the current NB and WB E-model. In Chapter 3, we introduce the proposed extension for SWB and FB scenarios. In Chapter 4, we describe a set of databases for analyzing the performance of our model. The new model predictions are compared to the subjective data in Chapter 5, and some perspectives for future work are given in Chapter 6.

2. NB and WB E-model

The main output of the E-model is a so-called transmission rating *R* which describes the overall quality experienced by a communication partner engaged in a conversation over a telephone channel which shows the characteristics as detailed by the parameter values. Thus, in contrast to most signal-based models like PESQ or POLQA, the estimation of the E-model addresses a conversational situation.

The basic assumption of the E-model is that degradations which are in principle independent of each other can be expressed as so-called "impairment factors" on the "transmission rating scale". The transmission rating *R* can then be calculated by subtracting the individual impairment factors from a maximum transmission rating *Ro* which is given by the

http://dx.doi.org/10.21437/Interspeech.2019-1340

effective signal-to-noise ratio of the connection, summing up all (ambient and circuit) noise sources[1]:

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

where Is is the impairment factor describing degradations occurring simultaneously to the transmitted speech signal (such as quantizing noise or sidetone), and Id is the impairment factor for degradations which occur delayed with respect to the transmitted speech signal (such as talker or listener echo, or the degradation of the conversation due to pure delay). *Ie,eff* is an impairment factor which has been introduced to describe the impairments caused by codecs under the effects of packet loss or discard. Strictly speaking, these effects also occur simultaneously to the speech signal, but they have been separated from circuit and ambient noise (*Ro*), quantizing noise and sidetone effects (*Is*), as they were considered independent.

For NB scenarios, the transmission rating varies between 100 (optimum NB quality) and 0 (worst quality). For WB, the maximum transmission rating is 129 (optimum WB quality), and the minimum still 0. In both cases, R can be transformed to an expected rating of users on a 5-point overall conversational quality scale as defined in [5], MOS_{CQE}, using an S-shaped function defined in [3]:

For $Rx < 0$: $MOS_{CQE} = 1$

For $0 < Rx < 100$: $MOS_{CQE} = 1 + 0.035\, Rx + Rx\,(Rx - 60) \cdot (100 - Rx) \cdot 7 \cdot 10^{-6}$ (2)

For $Rx > 100$: $MOS_{CQE} = 4.5$

with $Rx = R$ for NB, and $Rx = {}^{R}/_{1.29}$ for WB. Eq. (2) considers that in a typical subjective experiment, the maximum average rating commonly being observed is around 4.5 on the MOS scale ranging from 1 to 5, irrespective of the bandwidth of the presented speech stimuli. This fact will be addressed again in Chapter 3 regarding the consequences for SWB and FB.

Each of the values *Ro*, *Is*, *Id* and *Ie,eff* is calculated from parameters describing the communication system (network and terminal elements, including the surrounding noise situation). The exact parameters and the resulting formulae of the NB and WB E-model can be found in [3] and [4]. Here, only two impairment factors will be detailed, as they also form a basis for the extension towards SWB and FB. The first is the impairment factor *Id* which contains 3 contributions, *Id,te* for talker echo, *Id,le* for listener echo, and *Idd* for pure delay:

$$Id = Id,te + Id,le + Idd \quad (3)$$

with the contribution *Idd* calculated from the absolute delay between sender and receiver, *Ta*:
for $Ta \leq 100$ ms:

$$Idd = 0 \quad (4a)$$

for $Ta > 100$ ms:

$$Idd = 25\left\{(1+X^6)^{\frac{1}{6}} - 3\left(1+\left[\frac{X}{3}\right]^6\right)^{\frac{1}{6}} + 2\right\} \quad (4b)$$

and:

$$X = \frac{\log\frac{Ta}{100}}{\log(2)} \quad (5)$$

The second is the packet-loss dependent effective equipment impairment factor *Ie,eff*. It is derived using the codec-specific value for the equipment impairment factor at zero packet-loss *Ie,* the packet-loss probability *Ppl*, and the packet-loss robustness factor *Bpl*:

$$Ie,eff = Ie + (95 - Ie) \cdot \frac{Ppl}{Ppl+Bpl} \quad (6)$$

Values for *Ie* and *Bpl* are listed in Appendix I of [6].

3. Extension Towards SWB and FB

In a subjective experiment, the use of the MOS scale is largely affected by the stimulus set, the test participant group, the language, and alike [7]. Regarding the stimulus bandwidth, it has been shown in several experiments that MOS ratings for NB stimuli differ between tests where only NB stimuli are presented, and tests where both NB *and* WB are presented. An explanation is that test participants try to make optimum use of the available scale range regarding the offered stimuli. On the other hand, there is also experimental evidence that judgements for WB samples collected in a purely WB context do not differ significantly from those collected in a mixed NB/WB context, see [8] and [9].

For the extension of the E-model towards SWB, this implies that the transmission rating scale should show a higher R value for SWB situations in comparison to the WB (maximum R = 129) and NB (maximum R = 100), whereas the corresponding MOS ratings would still remain in the range [1;4.5]. Thus, the relationship between MOS and R needs to be adjusted towards a new maximum value for R. This can be done with the help of subjective mixed-bandwidth experiments in which WB scenarios, for which *Ie* values are already defined or for which other subjective judgments from a purely WB test are available, are judged together with SWB scenarios. The subjective judgements of the WB scenarios in relationship to the SWB scenarios then indicate by how much the transmission rating scale has to be extended for SWB.

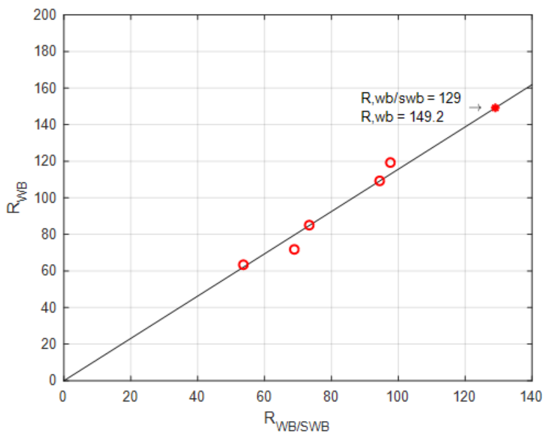

Figure 1: *Exemplary mixed-bandwidth test results. See paragraph below for a description of the variables.*

In Study Group 12 of the ITU-T, 7 subjective mixed-bandwidth experiments that contain a direct SWB system setting, as well as a number of NB or WB codecs for which the *Ie* values for the NB and WB E-models are known or for which corresponding subjective NB and/or WB tests were available, have been identified [10][11]. The subjective MOS values of

[1] In line with the notations of the ITU Recommendations, we do not use subscripts in this and the following equations.

the mixed-band conditions have been transformed to the R scale using Eq. (2). $R_{\text{WB/SWB}}$ then presents the R values of the NB/WB/SWB mixed bandwidth database compressed onto the WB R-scale which is limited to 129, whereas R_{WB} presents their uncompressed R values. A simple linear regression was used to map the compressed $R_{\text{WB/SWB}}$ values (derived from the subjective MOS) onto R_{WB} (derived from the defined *Ie* values). The linear regression was calculated without intercept term and therefore the estimation model goes through the origin. The R_{WB} value which corresponds to $R_{\text{WB/SWB}}$ = 129 then corresponds to the maximum value *Rmax* of the R scale in the SWB case. Fig. 1 shows an example of such a linear regression. Averaging over 7 databases and discussing about the limitations of individual databases [12], it was concluded that a value of R = 148 is appropriate as a new upper maximum of the R scale in SWB. The value seems also more realistic than the value of 179 proposed in [13].

Regarding FB, a test carried out by NTT [14] suggests that there was no statistically significant difference between the ratings for the clean (i.e. otherwise undegraded) SWB and the clean FB scenario. It can thus be concluded that the maximum R value of the transmission rating scale in both SWB and FB conversation scenarios is 148. The corresponding factor *Rx* of Eq. (2) linking the transmission rating scale to the MOS scale in SWB and FB tests then needs to be set to $Rx = {}^{R}/_{1.48}$.

Starting from this new maximum value, the basic formulae for the FB E-model can now be established. The formula corresponding to Eq. (1) (with index *FB*) now reads:

$$R = Ro,FB - Is,FB - Id,FB - Ie,eff,FB \quad (8)$$

So far, no subjective experiments addressing the effect of ambient or circuit noise, of quantizing distortions, or of non-optimum sidetone in SWB or FB are available to us. However, it can be assumed that the most dominant degradations to be expected in fully-digital SWB and FB scenarios are the effects of coding, packet loss, and delay. We will thus concentrate for this first extension of the E-model on these degradations, set *R,FB* to its maximum $Ro,FB = 148$, and set $Is,FB = 0$.

For describing the effects of pure delay, we will make use of Eq. (3), (4a), (4b) and (5), however neglecting the effects of talker and listener echo, as no subjective test data covering those degradations is available to us; thus, we set $Id,te = 0$ and $Id,le = 0$.

The effective equipment impairment factor *Ie,eff,FB* is derived using the codec-specific value for the equipment impairment factor at zero packet-loss *Ie,FB*, the packet loss probability *Ppl*, and the packet-loss robustness factor *Bpl*:

$$Ie,eff,FB = Ie,FB + (132 - Ie,FB) \cdot \frac{Ppl}{Ppl+Bpl} \quad (9)$$

Compared to Eq. (6), the value of 95 has been increased to the higher value of 132; this adjustment is discussed in Chapter 5.

4. Databases

In order to analyze the performance of the extended E-model and to identify parameter values *Ie* and *Bpl* for SWB codecs, 5 subjective listening-only test databases and 2 conversation test databases were kindly provided by ITU-T SG12 members. The details of the databases can be summarized as follows.

Orange (O1 and O2): These listening-only databases contain speech samples in French under clean and packet loss conditions. They were rated on the 5-point overall quality (MOS) scale by 24 test participants and the speech files were presented diotically with the frequency response limited to SWB. In total, there are 20 conditions (10% WB, 90% SWB) for O1, and 54 conditions (39% WB, 61% SWB) for O2.

Rohde & Schwarz (R&S): The listening-only database was kindly provided by Rohde & Schwarz and contains speech samples in German under clean and packet loss conditions [15]. The speech files were presented diotically and rated for MOS by 24 test participants. In total, there are 52 conditions (23% NB, 37% WB, 19% SWB, 21% FB) with each four different sentences. The direct FB condition received approximately the same rating as the direct SWB condition.

Qualcomm (QC): The listening-only database contains speech samples in American English; the results are published in the P.863 Implementer's Guide for EVS [16]. There are 49 conditions with varying bitrates available (17% NB, 38% WB, 45% SWB), with 24 different sentences for each condition that were rated by 32 test participants regarding MOS.

NTT (NTT): The database contains speech samples in Japanese. There are 37 conditions with different codecs and signal-correlated noise levels (27% NB, 46% WB, 27% SWB), with sentences from two male and 2 female speakers, each rated by 24 test participants regarding MOS [17].

TUB 1: This conversational database [18] followed the Short Conversation Test (SCT) scenarios defined in [19] which reflect everyday conversational situations and neither lead to very relaxed nor to very interactive conversational structures. The test was carried out with a SWB conversational system which allowed the online manipulation of *Ta* in 8 steps between 0 and 2500 ms. Stereo headsets were used as talking and listening devices, as to minimize potential acoustical echo. 24 naïve participants took part in the experiment and rated the overall quality of the conversation on the 5-point MOS scale.

TUB 2: The second conversation database contains an experiment carried out by Uhrig and Michael at the University of Western Sydney in Australia [20], in the English language. Participants were located in separate soundproofed rooms and communicated through monaural headsets. The speech signal was encoded with linear PCM in fullband, and end-to-end delay levels of *Ta* = 0, 800 and 1600 ms were introduced. 20 participants took part in the test. They each carried out 13 SCT conversations and rated the overall quality after each call.

5. Analysis and Comparison

In order to analyze the extended E-model regarding the effects of coding and packet loss, *Ie* values have to be calculated for SWB and FB codecs. The 5 listening-only databases can be used for this purpose. The calculation is done in 3 steps:

1. As the MOS_{CQE} values of Eq. (2) are limited to the range [1;4.5], the MOS values of the subjective experiment first need to be limited to that maximum value: $MOS_{norm,i} = \frac{MOS_i - 1}{MOS_{max} - 1} 3.5 + 1$ (10)
2. The normalized MOS values can now be transformed to R using Eq. (2) and setting $Rx = {}^{R}/_{1.48}$.
3. *Ie* values can now be calculated on the R scale by calculating the difference between the R values for the coded condition and the R value of the clean (otherwise not degraded) SWB channel condition.

This procedure was applied to calculate the *Ie* values for the EVS codec which was part of 5 databases, with different bitrates, see Tab. 1. The results were then discussed in ITU-T Study Group 12, and in order to come up with stable values for the EVS codec, it was decided to average the results, giving a ½ weighting to both Orange databases as they covered only a smaller range of degradations [21]. The resulting weighted average is given in the column "Average" or Tab. 1.

In order to check the derived *Ie* values for plausibility, a comparison will be done with predictions from the POLQA model. For this purpose, the 32 SWB test signals from the ITU-T P.501 database [22] were coded with the 9 different EVS SWB bitrate modes for the instrumental derivation. Then, POLQA scores were computed for each file, resulting in 32 POLQA MOS scores per EVS bitrate. The POLQA MOS scores have then been transformed to *Ie* values, using the procedure outlined above. The one but last column of Tab. 1 shows that the resulting *Ie* values are slightly higher, but comparable in their mostly descending order to the ones derived from the subjective experiments. It is however still unclear where the rather large differences for 32 and 48 kbit/s stem from. More details can be found in [23].

Table 1: *Ie and Bpl values for EVS codecs calculated from subjective tests and POLQA predictions.*

Bit-rate	O1	O2	RS	QC	NTT	Average	P.OLQA	Bpl
9.6	22.4	17.5	28.7	19.4		22.7	34.3	13
13.2	16.4	11.1	21.2	15.4	18.0	17.1	24.8	11.7
16.4	13.9	6.0	12.3	7.6	13.3	10.8	16.3	10.3
24.4	5.1			7.6	11.4	7.2	8.7	11.4
32.0		7.6		10.5	7.4	8.7	17.2	9.3
48.0				11.7	8.7	10.2	2.2	9.6

The P.501 speech files were also degraded with random packet loss, using the EVS codec at 16.4 kbit/s, and the resulting POLQA scores were calculated. The POLQA MOS scores were then transformed to *Ie* values, and using the average value of Tab.1, a corresponding value for *Bpl* was calculated by fitting the data points in a least-squares sense, and varying the constant term in Eq. 6. An optimum fit was found for a constant term of 132. This value was then used as a new constant in Eq. (13), and compared to the POLQA scores. The results in Fig. 2 show a very good fit for this part of the extended E-model.

In order to assess the extended E-model predictions for pure delay, the 2 conversational test databases are used. As the FB E-model predicts a maximum MOS of 4.5 for the degradation-free case, and nearly exactly this value has also been observed in the TUB 2 database, the MOS ratings of TUB 1 have been adjusted by shifting the maximum rating for the delay-free condition to 4.5. The resulting comparison to the FB E-model predictions is shown in Fig. 3.

Fig. 3 shows that applying Eq. (3), (4a), (4b) and (5) from the WB E-model to the SWB and FB test results provides a reasonable, but slightly pessimistic estimation for the range *Ta* = 0…1700 ms. For higher values of delay above 1700 ms, the subjective ratings degrade further, whereas the E-model predictions seem to reach a saturation. A better fit for that part of the curve can be reached by increasing the value of the pre-factor 25 in Formula (4b). However, we think that such high delay values are rather unrealistic, and that the FB E-model should rather concentrate on the range of delay values *Ta* below 1600ms to provide reasonable predictions in that range.

6. Conclusions and Future Work

We have presented a first version of a parametric planning model which addresses SWB and FB speech communication scenarios. The model is based on the NB and WB E-model, and extends it with a new maximum transmission rating, new values for *Ie* and *Bpl* describing SWB codecs, as well as an adjusted formula for *Ie,eff*. The model has been analyzed with the help of 7 subjective databases as well as POLQA predictions, showing a reasonable performance for coding and packet loss effects, as well as for the degradation due to pure delay.

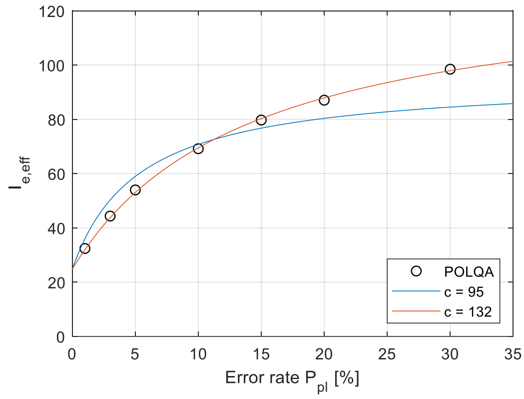

Figure 2: *Comparison of E-model predictions for Ie,eff to P.OLQA scores, using Eq. (6) and (9) (from [23]).*

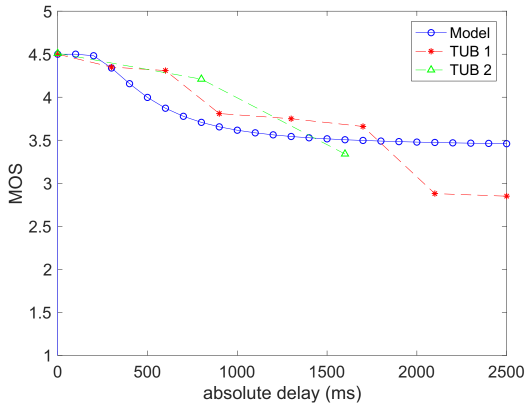

Figure 3: *Comparison of extended E-model predictions and subjective test results for the effects of pure delay.*

It is emphasized that the performed analysis is not a strict validation, as the subjective test results have been partially used to derive values for *Ie* and *Bpl*. In addition, other degradations which are covered by the NB and WB E-model are still missing in the extension, such as the effects of background noise, talker echo. As soon as subjective test results become available, it needs to be checked whether the corresponding formulae for *Ro* and *Id,te* from the NB E-model are still valid. Nevertheless, the achieved extension may already be applicable by network operators.

It has to be noted that the delay effects will most probably depend on the conversation test scenarios. More interactive scenarios (such as random number verification [19]) may lead to more substantial degradations of the quality ratings. As the SCT scenarios used in our conversation tests seem to represent everyday conversations quite well, we think that they may be reasonable as a basis for a first modelling in the FB E-model. As soon as further test results become available, the formula might be extended to provide different types of models for different conversational structures.

7. References

[1] ITU-T Recommendation P.862, *Perceptual evaluation of speech quality (PESQ): An objective method for end-to-end speech quality assessment of narrow-band telephone networks and speech codecs*, Geneva: Int. Telecomm. Union, 2001.

[2] ITU-T Recommendation P.863, *Perceptual objective listening quality prediction*, Geneva: Int. Telecomm. Union, 2018.

[3] ITU-T Recommendation G.107, *The E-model: a computational model for use in transmission planning*, Geneva: Int. Telecomm. Union, 2015.

[4] ITU-T Recommendation G.107.1, *Wideband E-model*, Geneva: Int. Telecomm. Union, 2015.

[5] ITU-T Recommendation P.800, *Methods for subjective determination of transmission quality*, Geneva: Int. Telecomm. Union, 1996.

[6] ITU-T Recommendation G.113, *Transmission impairments due to speech processing*, Geneva: Int. Telecomm. Union, 2007.

[7] S. Möller, *Assessment and Prediction of Speech Quality in Telecommunications*, Boston: Springer, 2000.

[8] V. Barriac, J.-Y. Le Saout, C. Lockwood C., "Discussion on Unified Methodologies for the Comparison of Voice Quality of Narrowband and Wideband Scenarios", *Proc. ETSI Workshop on Wideband Speech Quality in Terminals and Networks: Assessment and Prediction*, pp. 75-79, 2004.

[9] A. Takahashi, A. Kurashima, H. Yoshino, "Subjective Quality Index for Compatibly Evaluating Narrowband and Wideband Speech", *Proc. MESAQIN*, Prague, Czech Republic, 2005.

[10] ITU-T Contr. SG12-C79, *Super-wideband extension of R-value,* Source: Nippon Telegraph and Telephone Corporation (NTT), Geneva: Int. Telecomm. Union, 2017.

[11] ITU-T Contr. SG12-C94, *Maximum R-scale value estimations of a super-wideband E-model extension*, Source: Deutsche Telekom, Geneva: Int. Telecomm. Union, 2017.

[12] S. Möller, G. Mittag, F. Köster, V. Barriac, N. Côté, „Erweiterung der Netzwerkplanung auf super-breitbandige Sprachübertragung", in: *Fortschritte der Akustik – DAGA 2018: Plenarvortr. u. Fachbeitr. d. 44. Dtsch. Jahrestg. f. Akust.*, München, Berlin: Dtsch. Ges. Akust., 2018.

[13] M. Wältermann, A. Raake, S. Möller, "Extension of the E-Model Towards Super-Wideband Speech Transmission", in: *Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP 2010)*, Dallas TX, 14-19 Mar. 2010.

[14] ITU-T Contr. SG12-C260, *Fullband extension of R-value,* Source: Nippon Telegraph and Telephone Corporation (NTT), Geneva: Int. Telecomm. Union, 2018.

[15] ITU-T Contr. SG12-C22, *A subjective ACR LOT testing fullband speech coding and prediction by P.863*, Source: Rohde & Schwarz, Geneva: Int. Telecomm. Union, 2017.

[16] ITU-T Recommendation P.Imp863, *Implementer's Guides for ITU-T P.863*, Geneva: Int. Telecomm. Union, 2016.

[17] ITU-T Contr. SG12-C158, *Equipment impairment factors for EVS codecs*, Source: Nippon Telegraph and Telephone Corporation (NTT), Geneva: Int. Telecomm. Union, 2018.

[18] F. Köster, *Multidimensional analysis of conversational telephone speech*, Singapore: Springer Nature Singapore Pte Ltd., 2018.

[19] ITU-T Recommendation ITU-T P.805, *Subjective evaluation of conversational quality*, Geneva: Int. Telecomm. Union, 2007.

[20] S. Uhrig, T. Michael, S. Möller, P. Keller, J.-N. Voigt-Antons, "Effects of delay on perceived quality, behavior and oscillatory brain activity in dyadic telephone conversations", in: *Proc. 10th Int. Conf. on Quality of Multimedia Experience (QoMEX 2018)*, 29 - 31 May 2018, Sardinia.

[21] ITU-T Contr. SG12-C306, *Merged results of EVS-SWB equipment impairment factor estimations*, Source: Deutsche Telekom AG, Orange, Geneva: Int. Telecomm. Union, 2018.

[22] ITU-T Recommendation ITU-T P.501, *Test signals for use in telephonometry*, Geneva: Int. Telecomm. Union, 2017.

[23] G. Mittag, S. Möller, V. Barriac, S. Ragot, "Quantifying quality degradation of the EVS super-wideband speech codec", in: *Proc. 10th Int. Conf. on Quality of Multimedia Experience (QoMEX 2018)*, Sardinia, 29 - 31 May 2018.