PREDICTION OF BACKGROUND NOISE DEGRADATIONS IN FULLBAND SPEECH COMMUNICATION SCENARIOS

Sebastian Möller[1,2]*, Andreas Bütow*[1]

[1]*Quality and Usability Lab, TU Berlin,* [2]*Speech and Language Technology Lab, DFKI Berlin*
sebastian.moeller@tu-berlin.de, andreas.a.buetow@gmail.com

Abstract: In this paper, which is mostly based on the Master thesis of the second author [2], we analyze the impact of background noise at the sending and receiving side on perceived speech quality in fullband (20-20,000 Hz) speech communication scenarios. Whereas the effect of noise has been a research topic in narrowband (300-3,400 Hz) speech telephony for a long time, experimental data addressing wider speech transmission bandwidths are still scarce. In order to fill this gap, a listening-only experiment has been carried out in which background noise levels were carefully controlled at both sending and receiving side. Three types of noise have been considered, including speech-like babble as well as white and pink noise. Overall speech quality was assessed by 25 normal-hearing participants following ITU-T Rec. P.800. The results were analyzed regarding the effect of noise level and noise type on speech quality, revealing significant impacts of both factors. More precisely, babble noise showed a much smaller impact on overall quality than white and pink noise. Finally, the results were used to augment a standardized network planning model, the fullband E-model described in ITU-T Rec. G.107.2. The augmented model showed a very good correlation to the average test results, for all noise levels at sending and receiving side, while still neglecting the type of noise which might not be known during network planning applications anyhow.

1 Introduction

With the advent of IP-based communication protocols, speech communication services are no longer restricted to the traditional narrowband (300-3,400 Hz) channel, but offer wideband (50-7,000 Hz), super-wideband (20-14,000 Hz) or fullband (20-20,000 Hz) transmission of the speech signal. In turn, also background noise may impair speech quality over the full audio bandwidth, be it at the sending or receiving side of the communication channel. Whereas background noise degradations are rather well understood in narrowband scenarios [1][11], corresponding empirical data is scarce in fullband scenarios. Such empirical data is also necessary for developing prediction models anticipating the expected degradations, e.g. for network planning purposes.

In order to fill this gap, a listening-only experiment has been carried out in which background noise levels were carefully controlled at both sending and receiving side. Three types of noise have been considered, including speech-like babble which was however non-intelligible for test participants, as well as white and pink noise. 25 normal-hearing participants rated the overall speech quality following the guidelines given in ITU-T Rec. P.800 [10]. The results were first analyzed regarding the effect of noise level and noise type on speech quality. They were then used as a basis for augmenting a network planning model. For this purpose, we extended the fullband E-model which is currently standardized in ITU-T Rec. G.107.2 [7] to include a quantification of the impairment related to ambient noise. Two ways of calculating the related impairment were compared, one following a previously recommended version of the wideband E-model of ITU-T Rec. G.107.1 [5], and a modified version specifically addressing the full audio channel.

In the following sections, we will first describe the current narrowband and fullband E-model (Section 2), and then the set-up of the listening test (Section 3). The auditory results are analyzed in Section 4, and modifications to the fullband E-model are proposed in Section 5. Finally, Section 6 concludes on the results and makes proposals for future work.

2 E-model

Models for predicting speech quality are in common use during planning, set-up and operation of speech communication services. Whereas during operation speech signals can be measured at the terminals or in the network to serve as a basis for the prediction, this is impossible during service and network planning, as the corresponding equipment has not yet been set up. In those cases, parametric planning tools are used to estimate speech quality under different types of degradations. The most common and only standardized tool for this purpose is the E-model, which is defined in ITU-T Rec. G.107 [4] for narrowband speech transmission scenarios, in ITU-T Rec. G.107.1 [6] for wideband scenarios, and in ITU-T Rec. G.107.2 [7] for fullband scenarios. The model makes use of parametric descriptions of different types of impairments which are anticipated during the planning process, and calculates an estimation of the overall conversational speech quality on that basis.

The current narrowband E-model considers the effects of background noise at sending and receiving side, circuit and quantization noise, non-optimum loudness, sidetone, talker and listener echo, absolute delay, as well as the effects of digital coding techniques and packet-based transmission (packet loss or discard). Each of these degradations is first described by a set of parameters: Weighted noise levels, loudness ratings characterizing the effect of loss on perceived loudness, delay times, packet loss rates, as well as parameters characterizing digital codecs and their robustness towards packet loss. From these parameters, so-called *impairment factors* are calculated on a *transmission rating scale* ranging from *R* = 100 for optimum quality to *R* = 0 for lowest possible quality. The resulting transmission rating can then be transformed to an estimated *Mean Opinion Score* (MOS) which could be observed as an average judgment of test participants on a 5-point absolute category rating scale defined in ITU-T Rec. P.800 [10], limited to the range [1; 4.5].

For calculating the transmission rating *R*, the narrowband E-model defines a maximum transmission rating *Ro* representing the estimated signal-to-noise level of the connection, which is calculated assuming power addition of all noise sources at the virtual 0 dBr point of the connection [4]:

$$No = 10 \cdot \log_{10}\left(10^{Nc/10} + 10^{Nos/10} + 10^{Nor/10} + 10^{Nfo/10}\right) \tag{1}$$

In this equation, *Nc* represents the psophometrically weighted power of the circuit noise, and *Nfo* the corresponding power of noise on the subscriber line referred to the 0 dBr point. The corresponding equivalent noise powers resulting from the ambient noise at the sending side with A-weighted powers *Ps*, and at the receiving side with A-weighted power *Pr*, can be calculated using the send loudness rating *SLR* and the receive loudness rating *RLR* as follows:

$$Nos = Ps - SLR - Ds - 100 + 0.004 \cdot (Ps - OLR - Ds - 14)^2 \tag{2}$$

$$Nor = RLR - 121 + Pre + 0.008 \cdot (Pre - 35)^2 \tag{3}$$

with the overall loudness rating *OLR* = *SLR* + *RLR*, *Ds* and *Dr* the differences in sensitivities of the used sending and receiving terminal devices for speech compared to background noises, and

$$Pre = Pr + 10 \cdot \log_{10}\left(1 + 10^{\frac{10-LSTR}{10}}\right) \tag{4}$$

the effective room noise enhanced by the listener sidetone path with the corresponding loudness rating *LSTR*. From the value of *No* resulting from Eq. (1), the maximum value Ro is calculated by

$$Ro = 15 - 1.5 \cdot (SLR + No) \tag{5}$$

Finally, for calculating the overall transmission rating *R*, further equipment factors *Is* representing simultaneous impairments (quantizing noise, too loud/soft connection), *Id* representing delayed impairments (talker and listener echo, delay) and *Ie,eff* non-linear impairments (codecs, packet loss) are subtracted:

$$R = Ro - Is - Id - Ie,eff + A \tag{6}$$

A is an expectation factor catering for the psychological advantage effects occurring in rare communication scenarios, and is commonly set to zero.

When extending the E-model to wideband transmission scenarios (index *wb*), the transmission rating scale has been extended from its maximum value of 100 in narrowband to 129 in wideband. Further, the calculations of *Nos* and *Ro* have been changed in the 2015 version of ITU-T Rec. G.107.1 [5] compared to Eq. (2) and (5) as follows:

$$Nos,wb = Ps - SLR - Ds - 97 \tag{7}$$

$$Ro,wb = 20 - 1.5 \cdot (SLR + No) \tag{8}$$

Because doubts arose with this modified calculation of *Ro,wb*, the currently recommended version of ITU-T Rec. G.107.1 [6] uses a fixed value of *Ro,wb* = 129 instead. The resulting value of *R* is again transformed to a MOS value in the same range [1; 4.5].

Recently, the E-model has also been extended to fullband scenarios (index *fb*), see ITU-T Rec. G.107.2 [7]. This extension results in a newly extended range for $R \in [0; 148]$, with a fixed value of *Ro,fb* = 148. The delayed impairment factor *Id,fb* only includes the effects of pure delay (no echoes), and *Is,fb* = 0 in the fullband version. Thus, this model requires an extension to also include the effects of background noise at send and receive side, which is the aim of our investigations reported hereafter.

3 Listening experiment

In order to quantify the impact of background noise at send and receive side on overall quality, a listening experiment has been set up in a controlled laboratory environment. The room design followed ETSI EG 202 396-1 V1.2.2 [3] with respect to its size, equipment and noise floor (below 32 dB(A)). Four loudspeakers positioned at 1.25 m height in a square around the test participant were used to generate a rather diffuse background noise around the test participant.

In addition to the noise-free condition, three types of noise adjusted to levels of 55, 65 and 75 dB(A) have been considered, including speech-like babble which was however non-intelligible for test participants, as well as white and pink noise. Speech stimuli were presented via an open headset (Plantronics Audio 355) at approx. 74 dB(A) sound pressure level (SPL) and consisted of 12 double or triple sentences of approx. 10 s length, each recorded in fullband audio in a sound-insulated cabin from 2 female and 2 male speakers. The clean source files were degraded in the noisy environment by first calibrating the noise level at the position of the listener, then re-recording the noise with the headset by positioning it on a head-and-torso simulator HEAD Acoustics GmbH HMS II.3, and finally mixing the speech signal with the re-recorded noise at an active speech level of -26 dB relative to the overload point of the digital system, corresponding to an acoustic SPL of approx. 82 dB at the headset microphone. This procedure was repeated for all speech files using all available sending noise conditions (3 noise types x 3 noise levels), and a clean fullband, a wideband (following ITU-T Rec. P.341 [9]) and a narrowband version (following ITU-T Rec. P.48 [8]) of each file were added, resulting in 12 sending noise conditions (c1 to c12) with four stimuli each. The 48 stimuli were then presented either in quiet (background noise condition r1) or with pink noise of level 55, 65 or 75 dB(A) at the receiving side (background noise condition r2, r3 and r4).

25 participants reporting non-impaired listening rated the overall speech quality following the guidelines given in ITU-T Rec. P.800 [10] on a 5-point ACR scale. Before the start of the experiment, participants were informed about the purpose of the experiment, and listened to a range of 6 noise conditions to get accustomed to the quality levels to be expected in the experiment. They then carried out four separate sessions corresponding to the background noise conditions r1…r4, with randomized order of sending (c1…c12) and receiving (r1…r4) noise conditions. Participants were renumerated for their effort.

4 Analysis of results

The subjective ratings were first screened for outliers, resulting in no suspicious cases. Thus, 100 judgments distributed over 4 speech files form the basis for each noise condition average. Fig. 1 shows the average ratings for each sending and receiving noise condition.

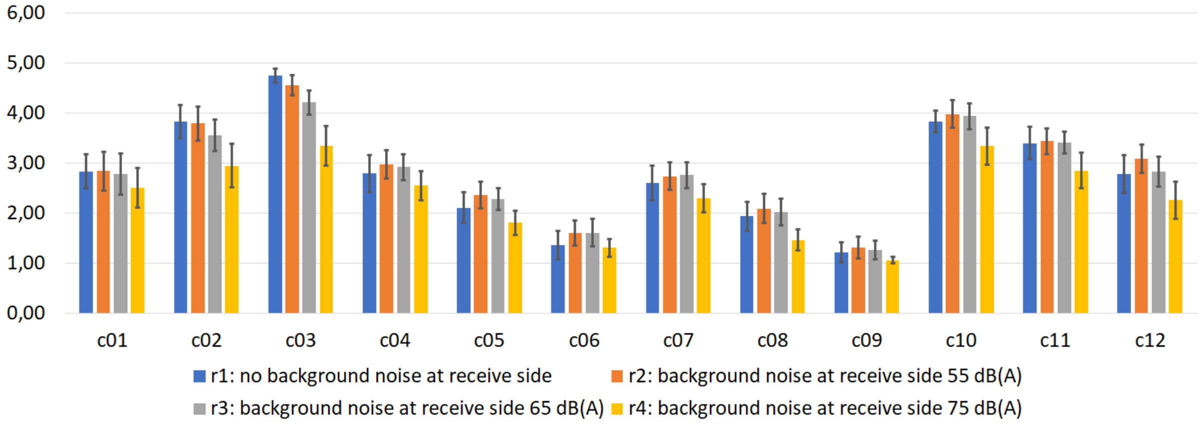

Figure 1 - Average values and 95% confidence intervals for each noise condition.

Comparing the results for condition r1, the advantage of fullband over narrowband and wideband transmission as well as the degrading effect of background noise become apparent. White sending noise (conditions c7…c9) seems to degrade overall quality slightly more than pink noise (c4…c6), but considerably less than babble noise (c10…c12). Noise on the receiving side (r2…r4) degrades quality in some cases compared to the noise-free case; the degradation is most apparent for the high-quality sending conditions and for the high receiving background noise conditions, whereas it seems to saturate at the low-quality sending conditions (narrowband and loud sending noise). Statistical significance tests could only in a few cases prove a statistically significant difference (e.g. for c3). Still, the observed effects all seem reasonable and follow the hypothesis of degrading quality with increasing noise level.

5 E-model extension

The results were finally used to augment the fullband E-model described in ITU-T Rec. G.107.2 [7]. In a first approached, we followed the 2015 version of the wideband E-model [5] described in Eq. (7) and (8), and using values of -96 dBm0p for *Nc* and *Nfor*. Fig. 2 shows the resulting predictions for different levels of sending and receiving noise, in comparison to the auditory test results for the three types of noise. It should be noted that the E-model does not take the type of noise into account in its predictions, as this is expected to be unknown at the time of network planning.

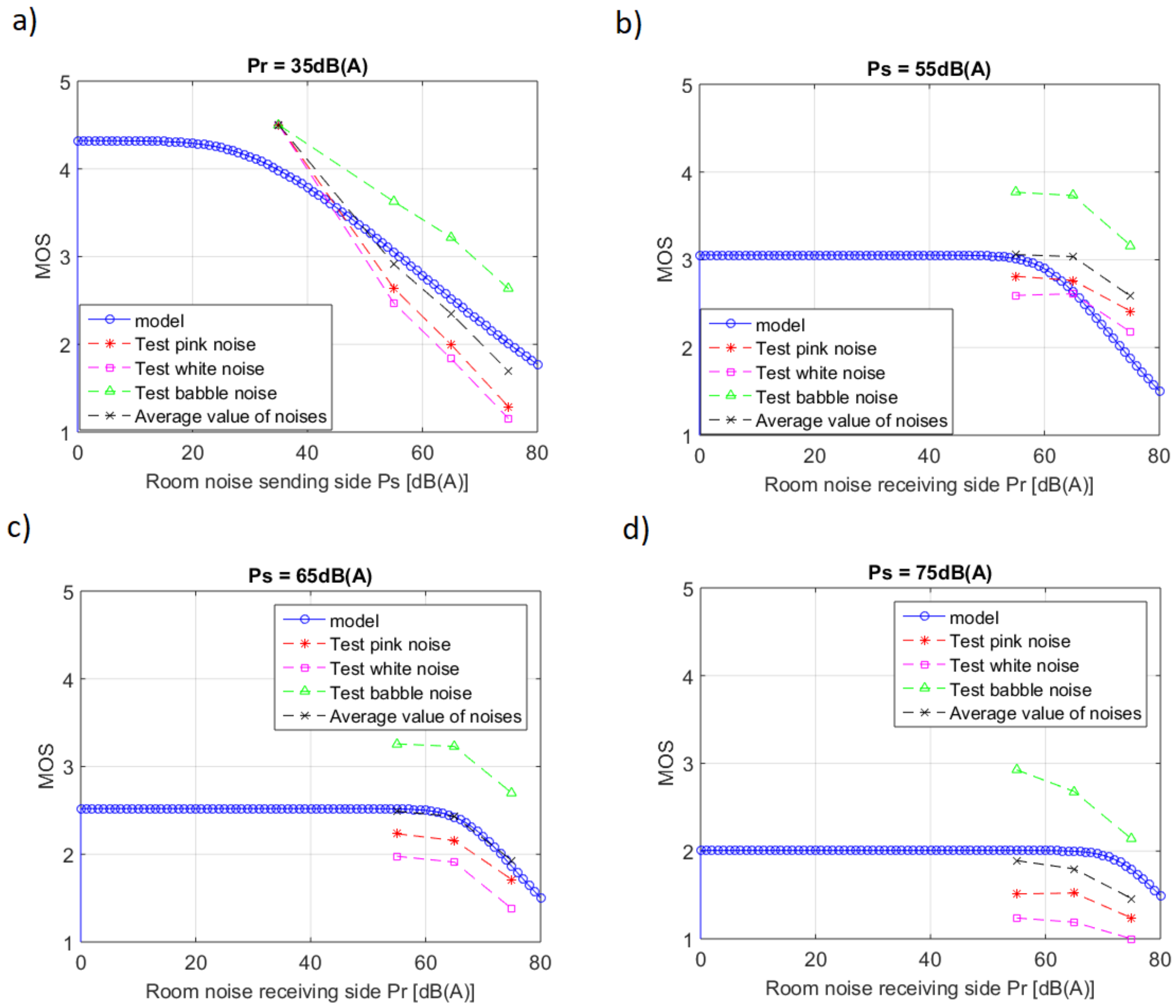

Figure 2 – Fullband E-model predictions using the approach of the wideband model with Eq. (7) and (8). Panel a) shows the impact of sending side noise, whereas panels b) to d) address receiving side noise.

The figure shows that the E-model predictions are quite well in line with the auditory test results when averaging over all noise types. As it was observed in the auditory test, the actual impact of babble noise is less strong, and the impact of pink and white noise is stronger than predicted. As white noise can be considered to be quite un-representative for real-life noise, this behavior seems acceptable for a planning model.

In a second approach, we used again Eq. (2) from the narrowband E-model for calculating *Nos*, and setting again *Nc* = *Nfor* = -96 dBm0p. For calculating *Nor*, we modified Eq. (3) to

$$Nor = RLR - 147 + 1.12 \cdot Pre + 0.009 \cdot (Pre - 25)^2 \qquad (9)$$

The resulting values were used in Eq. (1) for calculating *No*, and then calculating *Ro,fb* with

$$Ro,fb = 20 - 1.5 \cdot (SLR + No) \qquad (10)$$

A comparison between of the thus-modified fullband E-model with the auditory test results is shown in Fig. 3. The predictions of the model are now in even better alignment with the auditory results. Again, the E-model predicts the averages over all noise types quite well, whereas the differences between the noise types are not taken into account by the model.

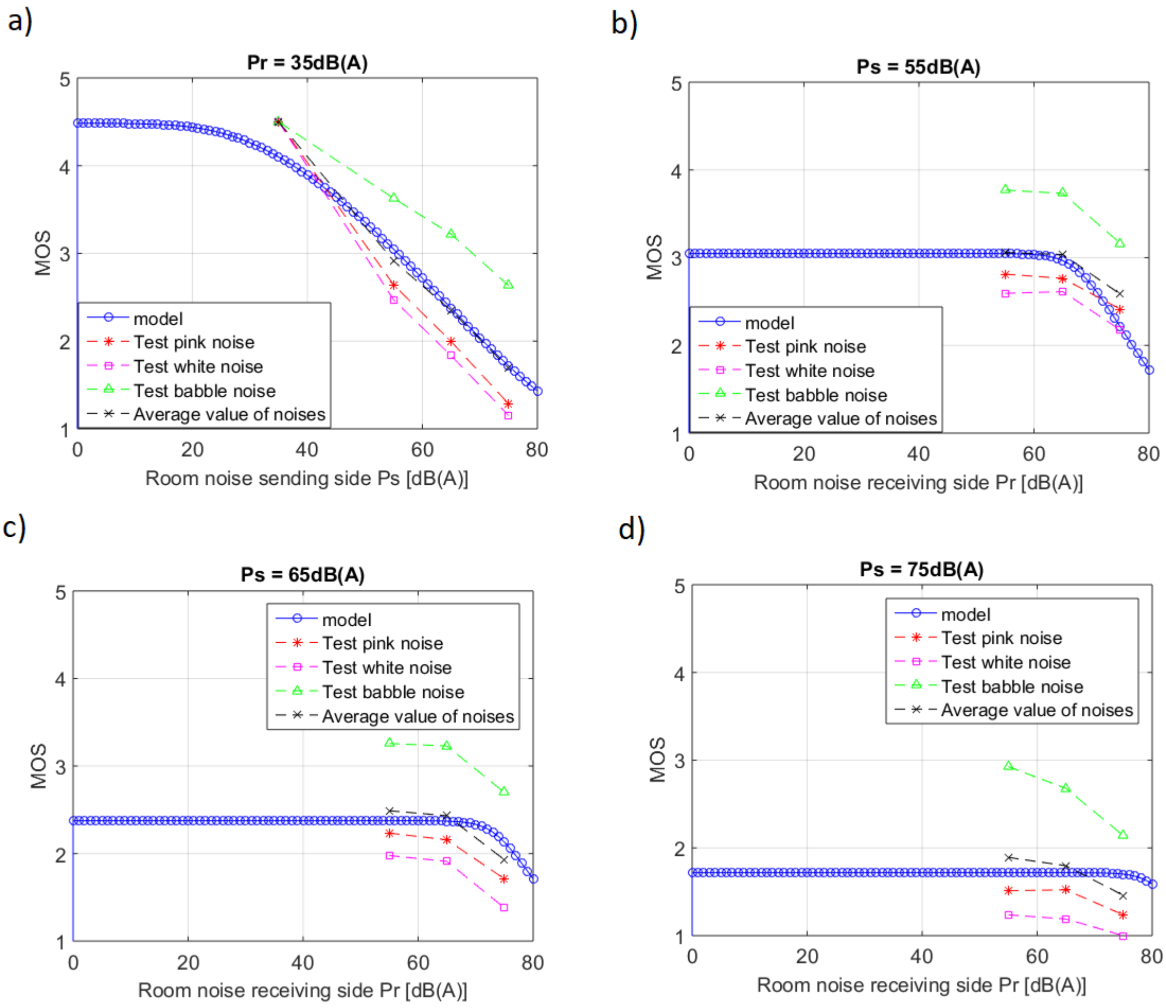

Figure 3 - Fullband E-model predictions using the modified approach with Eq. (9) and (10). Panel a) shows the impact of sending side noise, whereas panels b) to d) address receiving side noise.

6 Conclusions and future work

In this paper, we presented an empirical analysis of the effect of ambient background noise at the sending and receiving side of a fullband speech communication scenario. In a listening-only laboratory test paradigm, three types of noise at the sending and one type of noise on the receiving side were carefully controlled, and overall quality judgments were solicited from 25 normally-hearing test participants. In addition to the noise degradations, also narrowband and wideband speech files were included in the test. The results followed the expected behavior: Both sending and receiving side noise degraded overall quality, the more the higher the noise level is. Interestingly, the type of noise seems to play a major role for the judgments: Speech-like babble noise caused considerably less degradation compared to pink and (especially) white noise. The quality advantage of fullband compared to both narrowband and wideband speech transmission was also clearly visible.

The auditory test results were subsequently used for extending the fullband version of the E-model, a popular planning tool for speech communication services, to include the effects of sending and receiving side noise. Using a previous version of the wideband E-model, already quite acceptable predictions could be obtained, which come close to the observations when averaging across noise types. The predictions could be further improved by modifying the formula for calculating *Nor*, the equivalent power of receiving side noise at the 0 dBr point of the connection, from the narrowband E-model, and by using the formula for calculating the maximum transmission rating *Ro,fb* from the wideband E-model. Using these formulae, an excellent fit to the average test results could be obtained for all tested sending and receiving noise levels.

It should be noted that the experimental paradigm was listening-only. It may happen that the degradation experienced in a real conversation differs from the one in a listening-only situation. As conversational experiments require longer interactions compared to the stimuli used in listening-only tests, we preferred to use the listening paradigm to be able to test more noise levels and noise types. Future conversation experiments should test whether the subjective conversational ratings for some anchor conditions differ to the ones observed here.

The formulae addressing background noise effects can now be integrated into the fullband E-model, and potentially lead to an update of ITU-T Rec. G.107.2 [7]. It needs to be assessed whether the predictions stay valid also when noise is combined with other types of degradations, such as coding distortions, packet loss or delay. Further empirical studies have to address such combinations.

7 Acknowledgements

The authors gratefully acknowledge the support from Falk Schiffner, Rafael Zequeira Jiménez, Clemens Feldhaus and Gabriel Mittag during the set-up of the experiment. We further thank our test participants who volunteered despite the restrictions of the Covid-19 pandemic.

8 Bibliography

[1] BODDEN, M., JEKOSCH, U.: *Entwicklung und Durchführung von Tests mit Versuchspersonen zur Verifizierung von Modellen zur Berechnung der Sprachübertragungsqualität*. Final report to a project funded by Deutsche Telekom AG (unpublished), Institut für Kommunikationsakustik, Ruhr-Universität, Bochum, 1996.

[2] BÜTOW, A.: *Vorhersage der Qualität verrauschter Sprache bei vollbandiger Telefonie*. Master thesis, Quality and Usability Lab, Technische Universität Berlin, 2020

[3] ETSI Guide EG 202 396-1 V1.2.2: *Speech Processing, Transmission and Quality Aspects (STQ); Speech Quality Performance in the Presence of Background Noise; Part 1: Background Noise Simulation Technique and Background Noise Database*. European Telecommunications Standards Institute, Sophia Antipolis, 2008.

[4] ITU-T REC. G.107: *The E-model, a Computational Model for Use in Transmission Planning*. International Telecommunication Union, Geneva, 2015.

[5] ITU-T REC. G.107.1: *Wideband E-model*. International Telecommunication Union, Geneva, 2015.

[6] ITU-T REC. G.107.1: *Wideband E-model*. International Telecommunication Union, Geneva, 2019.

[7] ITU-T REC. G.107.2: *Full E-model*. International Telecommunication Union, Geneva, 2019.

[8] ITU-T REC. P.48: *Specification for an Intermediate Reference System*. International Telecommunication Union, Geneva, 1988.

[9] ITU-T REC. P.341: *Transmission Characteristics for Wideband Digital Loudspeaking and Handsfree Telephony Terminals*. International Telecommunication Union, Geneva, 2011.

[10] ITU-T REC. P.800: *Methods for Subjective Determination of Transmission Quality*. International Telecommunication Union, Geneva, 1996.

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