Derivation of Equipment Impairment Factors from ITU-T P.863 for Fullband Speech Codecs

Thilo Michael[1], Gabriel Mittag[1], Sebastian Möller[1,2]

[1] *Quality and Usability Lab, Technische Universität Berlin, Germany, [firstname.lastname]@tu-berlin.de*

[2] *German Research Center for Artificial Intelligence (DFKI), Berlin, Germany*

Abstract

Recently, the fullband version of the E-Model, a parametric tool for speech communication service planning, has been standardized by the International Telecommunication Union ITU-T. In order to use the E-model for a given codec, the equipment impairment factor Ie has to be defined for the speech codec under study. This procedure standardized in ITU-T Recommendation P.833 describes a methodology to derive Ie by conducting listening-only tests with that codec, together with other codecs for which the Ie is known. In ITU-T Rec. P.834, a similar methodology is described to derive the equipment impairment factor based on predictions of instrumental models. With the extension of the E-model to fullband communication scenarios, this methodology can be extended for fullband speech codecs. In this paper, we present our first efforts for this extension. We use the ITU-T P.863 instrumental model in fullband mode to estimate the quality of different fullband, wideband, and narrowband coded speech files to obtain instrumental model-based equipment impairment factors Ie,FB. We then apply a linear interpolation between the expected Ie,FB values, and the instrumental model-based Ie,FB predictions and use them to predict the Ie,FB for the Opus codec in fullband mode, for which currently no Ie,FB values are available.

Introduction

When service providers plan the quality of future speech communication services they are usually not able to instrumentally measure Quality of Service (QoS) indicators in order to estimate the Quality of Experience (QoE), which is perceived later by users of the speech service. Instead, these service providers make use of parametric models which can predict one or several indicators of QoE based on planning assumptions of the respective network and terminals. These assumptions relate to the characteristics of the transmission network and quantify the characteristics via single-dimensional parameters, such as the coding processes involved, the assumed delay and attenuation, the assumed rate of lost packets, the assumed noise level in the surroundings of both interlocutors, etc. Based on these parameters, the overall speech quality experienced by both interlocutors of the conversation is predicted by an algorithmic model. The most popular such model is the E-model, which has een standardized by the International Telecommunications Union as ITU-T Rec. G.107 in 1996, and extended several times since then [1].

With the advent of packet-switched voice over IP (VoIP) transmission networks, the full audible frequency and (0-20,000 Hz) can easily be transmitted. Compared to a standard narrowband (300-3,400 Hz) speech communication channel, fullband transmission improves the experienced quality. It has been quantified that the amount of improvement reaches roughly 48% when measured on the so-called transmission rating scale R underlying the E-model [2–4]. As a consequence, the E-model has recently been extended to predict such elevated quality levels in ITU-T Rec. G.107.2 [5]. In addition, this recent extension predicts the effects of several EVS codecs with and without random packet loss, as well as the effects of pure delay.

When calculating impairments related to the coding of speech and packet-loss, codec related parameters like the equipment impairment factor Ie or the burstiness robustness factor Bpl are used. These parameters are specific to each codec and a list of these values for most established codecs are available in ITU-T Rec. G.113 [6]. For codecs that are not included in that recommendation, ITU-T Rec. P.833 describes a methodology to derive the equipment impairment factors for such codecs based on subjective listening tests [7]. ITU-T Rec. P.834 extends this methods for derive the equipment impairment factors based on instrumental models [8]. While these recommendations have been update for the wideband E-model [9], there is currently no standardized method to derive the equpment impairment factors for super-wideband and fullband speech codecs.

In the present paper, we use the instrumental model ITU-T Rec. P.863 "POLQA" [10] in fullband mode to estimate the quality of the fullband Opus codec. Thereby, we propose a way to extend the Methodology to derive eqipment impairment factors from instrumental models ITU-T P.834 to work in a fullband context.

Related Work

The E-model is the most popular parametric model for the prediction of conversational quality. While initially only applicable to narrowband voice services in ITU-T Rec. G.107 [1], it has been extended for wideband communication in ITU-T Rec. G.107.1 [9] and recently fullband communication in ITU-T Rec. G.107.2 [5]. The E-model assumes that impairments are independent of each other and can be quantified as impairment factors on a so-called "psychological scale", that is the scale of the transmission rating R. The transmission rating R itself can be transformed into an average judgment of a test participant on a 5-point absolute category rating scale, the Mean Opinion Score (MOS) using an S-shaped monotonous relationship. While in the narrow-

band version, the maximum transmission rating is 100, it increases to 148 in the fullband version.

Codec related impairment factors are standardized in ITU-T Rec. G.113 [6] for narrowband codecs and the Amendments 1 and 2 extend the recommendation to include wideband and fullband codecs respectively. Because the equipment impairment factors and packet loss robustness factors have not been standardized for all possible codecs, they can be calculated based on subective ratings. ITU-T Rec. P.833 [7] describes this process based on the narrowband E-model, while ITU-T Rec. P.833.1 [11] extends the methodology for wideband codecs. With instrumental quality prediction models like ITU-T Rec. P.862 "PESQ" [12] and ITU-T Rec. P.863 "POLQA" [10] it is possible to derive the impairment factors without the need for subjective tests. This methodology is standardized in ITU-T Rec. P.834 [8] for narrowband and in ITU-T Rec. P.834.1 for wideband.

While recent work has dervied the equipment impairment factors of the Opus codecs for narrowband and wideband [13] and even for super-wideband [14], these approaches do not utilize the fullband version of the E-model and thus do not take into account the increase in the transmission rating R.

Test setup

As a first step towards this extension, we process 28 different fullband speech samples from ITU-T Rec. P.501 Annex C [15] with 6 different codecs at various different bitrates. While one codec, the Opus Codec [16] at version 1.3-RC, has no standardized Ie,FB value, the five other codecs have standardized equipment impairment factors in Rec. G.113, Amendment 1 and Amendment 2 [6]. We predict the MOS of the speech samples using ITU-T P.863 model version 3 in fullband mode [10] and transform the results into the R rating scale of the fullband E-model. We then used the least squares method to linearly interpolate the calculated Ie,FB value of the P.863 prediction with the expected Ie,FB. We finally used this linear model to predict the equipment impairment factor of the Opus codec at different bitrates.

For the input speech material, we used the speech files provided by Rec. P.501 Annex C [15]. We selected the fullband speech files for the seven languages Chinese, English, Finish, French, German, Italian, and Japanese. The Dutch files were left out as they are only available in super-wideband. For each language two male and two female speech files are present, resulting in 28 clean fullband speech samples.

Because the target codec Opus has no set bitrates [16], we selected the five bitrates 13, 16, 24, 32, and 48 kbit/s to be comparable to the EVS codec at similar bitrates. Because there are not yet any reference conditions defined for fullband codecs, we used different fullband, wideband and narrowband codec as reference conditions. We selected EVS with 6 bitrates, AMR-WB with 9 bitrates, G.722 with 3 bitrates, AMR-NB, and G.711 (see Table 1). For non-fullband codecs we transformed the Ie,WB and Ie,NB values from G.113 and G.113 Amendment 1 into Ie,FB values by adding the differences between the maximum R value of the fullband scale (148) and the maximum R value of the narrowband (100) or wideband (129) scale. We used every codec and bitrate combination in a tandem with itself and calculated the Ie,FB values accordingly. Table 1 shows all codecs, bitrates and the equipment im-pairment factors of the codecs as defined by Rec. G.113 Amendment 2 and the calculated impairment factors from G.113 and G.113 Amendment 1, as well as the calculated Ie,FB value of the self-tandem according to the additivity principle. We coded each of the 28 speech samples with the 25 codec-bitrate-combinations and their tandems, resulting in 1400 coded speech files.

Table 1: Codecs, bitrate (in kbit/s), expected Ie,FB values and Ie,FB values of their tandems.

Codec	Bitrate	Ie,FB	Tandem Ie,FB
Opus	13.0	-	-
	16.0	-	-
	24.0	-	-
	32.0	-	-
	48.0	-	-
EVS	9.6	22.7	45.4
	13.2	17.1	34.2
	16.4	10.8	21.6
	23.4	7.2	14.4
	32.0	8.7	17.4
	48.0	10.2	20.4
AMR-WB	6.6	60.0	101.0
	8.85	45.0	71.0
	12.65	32.0	45.0
	14.25	29.0	39.0
	15.85	26.0	33.0
	18.25	24.0	29.0
	19.85	22.0	25.0
	23.05	20.0	21.0
	23.85	27.0	35.0
G.722	48.0	50.0	81.0
	56.0	39.0	59.0
	64.0	32.0	45.0
AMR-NB	12.2	53.0	58.0
G.711	64.0	48.0	48.0

Derivation of fullband equipment impairment factors

For the instrumental assessment with ITU-T P.863, we used the SQuadAnalyzer software in its full-band mode, and POLQA version 3. We linearly scaled the resulting MOS estimations to range from 1 to 4.5.

We inverted the equations 1a and 1b to transform the scaled MOS estimations from ITU-T P.863 to the R scale of the fullband E-model (R_FB):

$$Rx = \frac{R_FB}{1.48} \tag{1a}$$

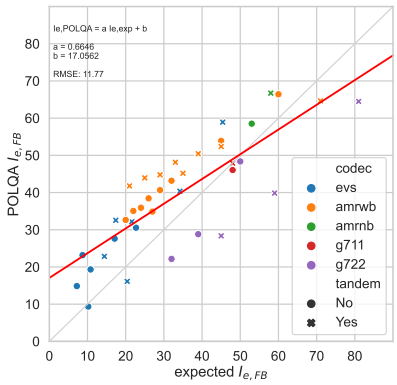

Figure 1: Scatter plot of expected Ie, FB values and predictions by ITU-T P.863. Colored dots correspond to the different type of codecs and crosses show the tandems.

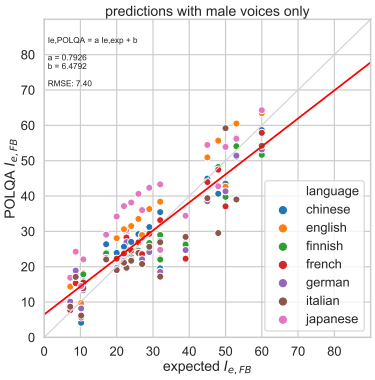

Figure 2: Scatter plot of only male voices, divided by the language of the speaker without tandem codecs.

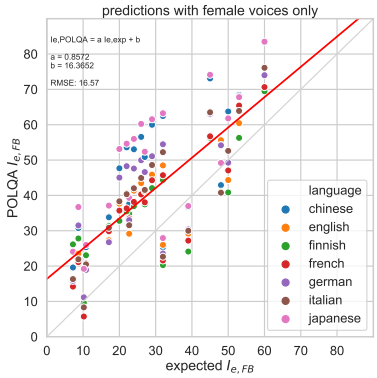

Figure 3: Scatter plot of only female voices, divided by the language of the speaker without tandem codecs.

$$\begin{aligned} For\ Rx < 0\text{: } & MOS_{CQE}{=}1 \\ For\ 0 < Rx < 100\text{: } & MOS_{CQE}{=}1{+}0.035Rx \\ & {+}Rx(Rx{-}60){\cdot}(100{-}Rx){\cdot}7{\cdot}10^{-6} \\ For\ Rx > 100\text{: } & MOS_{CQE}{=}4.5] \end{aligned} \tag{1b}$$

Assuming no other degradations are present in the coded speech signals, we can transform the R values directly into estimates for the fullband equipment impairment factor (K):

$$K = 148 - R_{FB} \tag{2}$$

From these P.863-based K values and the defined Ie, FB, def values we calculated an interpolation line as shown in equation 3. The coefficients a and b are fitted in a least-squares sense.

$$K = a \cdot Ie, FB, def + b \tag{3}$$

Results

Figure 1 shows the scatterplot of the defined Ie, Fb, def values and the P.863-based K values (denoted as "POLQA Ie,FB" in the plot). Figure 2 and 3 split the data set up into the language and gender of the speaker. The differences in languages are similar for both male and female voices. However, for female speakers, the P.863 predictions result in a higher estimated Ie, FB than for the male speakers. The least squares fit in Figure 1 results in $a = 0.82$ and $b = 11.42$.

To estimate the Ie, FB value for the range of Opus codecs under study, we first calculated the K values from the scaled P.863 predictions based on the equations 1a, 1b, and 2. We then used the linear fit from equation 3 to calculate the estimates of the fullband equipment impairment factors:

$$Ie, FB = \frac{K - b}{a} \tag{4}$$

Table 2 shows the calculated Ie, FB for Opus bitrates 13, 16, 24, 32, and 48 kbit/s and their self-tandems. For the conditions in which the calculated Ie, FB value is negative, the value was set to zero instead.

Table 2: Estimated Ie, FB values for Opus at various bitrates (in kbit/s) and with self-tandem. For negative calculated values, the Ie, FB is set to 0*.

Codec	Bitrate	Ie,FB	Tandem Ie,FB
Opus	13.0	37.85	52.73
	16.0	21.96	34.93
	24.0	9.82	21.27
	32.0	2.58	12.4
	48.0	0*	2.92

Conclusion

This paper illustrates the extension of the methodology defined in ITU-T Rec. P.834 and P.834.1 to fullband speech codecs, and a first exemplary determination of a fullband equipment impairment factor for the Opus codec. The methodology uses the fullband codec EVS, the wideband codecs AMR-WB and G.722, as well as the narrowband codecs AMR-NB and G.711 as reference conditions. The relationship between the estimated, P.863-based Ie, FB values and the defined Ie, FB, def values is largely dependent on the language of the speakers and their gender. In general, the predictions from P.863 are more pessimistic than the defined impairment factors, which results in a linear interpolation with a positive y-intercept of 11.42. Future research should expand on this research and identify more references conditions and tandems that can be reliably used in the fullband scenario. Also, an estimation of the fullband impairment factor from subjective tests should be conducted. Additionally, a method to derive packet loss robustness factors should be defined.

References

[1] ITU-T Recommandation G.107, *The E-model: a computational model for use in transmission planning.* Geneva: International Telecommunication Union, 2011.

[2] ITU-T Contribution SG12-C79, "Super-wideband extension of r-value." Source: Nippon Telegraph and Telephone Corporation (NTT), 2017.

[3] ITU-T Contribution SG12-C94, "Maximum r-scale value estimations of a super wideband e-model extension." Source: Deutsche Telekom, 2017.

[4] S. Möller, G. Mittag, T. Michael, V. Barriac, and H. Aoki, "Extending the e-model towards super-wideband and fullband speech communication scenarios," in *Proceedings of INTERSPEECH 2019*, (Graz, Austria), pp. 3436–3440, 2019.

[5] ITU-T Recommandation G.107.2, *Fullband E-model*. Geneva: International Telecommunication Union, 2019.

[6] ITU-T Recommandation G.113, *Transmission impairments due to speech processing*. International Telecommunication Union, 2007.

[7] ITU-T Recommandation P.833, *Methodology for derivation of equipment impairment factors from subjective listening-only tests*. International Telecommunication Union, 2001.

[8] ITU-T Recommandation P.834, *Methodology for the derivation of equipment impairment factors from instrumental models*. International Telecommunication Union, 2001.

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

[10] ITU-T Recommendation P.863, *Perceptual objective listening quality assessment*. International Telecommunication Union, 2014.

[11] ITU-T Recommandation P.833.1, *Methodology for derivation of equipment impairment factors from subjective listening-only tests for wideband speech codecs*. International Telecommunication Union, 2009.

[12] ITU-T Recommandation 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*. International Telecommunication Union, 2001.

[13] M. Al-Ahmadi, P. Pocta, and H. Melvin, "Derivation of e-model equipment impairment factors for narrowband and wideband opus codec using the instrumental method," in *2020 31st Irish Signals and Systems Conference (ISSC)*, pp. 1–6, IEEE, 2020.

[14] M. Al-Ahmadi, P. Pocta, and H. Melvin, "Instrumental estimation of e-model equipment impairment factor parameters for super-wideband opus codec," in *2019 30th Irish Signals and Systems Conference (ISSC)*, pp. 1–5, IEEE, 2019.

[15] ITU-T Recommandation P.501, *Test signals for use in telephony and other speech-based applications*. International Telecommunication Union, 2020.

[16] J.-M. Valin, K. Vos, and T. Terriberry, "Definition of the opus audio codec," *IETF, September*, 2012.