Quantifying the Quality Difference between Narrow-Band and Wideband Speech Codecs

Sebastian Möller[1], Alexander Raake[2]

[1] *IKA, Ruhr-Universität Bochum, D-44780 Bochum, Germany, Email: sebastian.moeller@rub.de*

[2] *Perception Située, LIMSI-CNRS, F-91403 Orsay, France, Email: alexander.raake@limsi.fr*

Introduction

When speech is transmitted over wideband (wb, 50-7000 Hz) instead of narrow-band (nb, 300-3400 Hz) telephone channels, it can be expected that the overall quality perceived by the human listener rises. However, degradations due to coding distortions and noise impact speech quality in both nb and wb situations. In order to judge the overall quality of a telephone connection, degradations occurring on the entire transmission channel from the mouth of the talker to the ear of the listener have to be taken into account. The degradations result in different perceptual quality dimensions which have to be integrated to form the overall quality percept.

One way of integrating different degradations is the *impairment factor principle* applied by the International Telecommunication Union (ITU-T) for predicting the quality of nb telephone networks. The impairment factor principle assumes that all types of degradations occurring jointly on a transmission channel can be quantified on a perceptual scale, in terms of individual *impairment factors* (degradation factors). This perceptual scale is called transmission rating scale (*R*-scale) and ranges from 0 for lowest possible to 100 for optimum quality. Despite efforts described in [1], there is still no method available which would allow direct scaling on the *R*-scale; only a transformation is defined in [2], linking ratings obtained on a 5-point ACR quality scale (MOS scale) to the *R*-scale.

Impairment factors are one-dimensional values calculated either from signals which are measured on the transmission channel, or from auditory listening-only tests. The overall quality of the channel can then be determined by subtracting all impairment factors from the maximum R_{max} value obtained for a "clear" channel, i.e. the channel solely impaired by the bandwidth limitation and uncorrelated noise. This subtraction implicitly assumes additivity of impairments on the *R*-scale. The calculation of impairment factors and subsequent transformation to an overall quality scale (MOS scale) is defined by the E-model, see [2].

So far, the impairment factor principle has only been defined for nb speech transmission. When the bandwidth is extended to wb transmission, it can be expected that the *R*-scale has to be extended beyond the maximum value of 100, because test subjects tend to give higher ratings to speech stimuli transmitted over such channels (typically 1.3 to 1.5 points higher on an MOS scale). Starting from the higher overall rating, degradations have to be quantified in terms of new (wb-) impairment factors which will give an estimation of the overall quality degradation introduced on the wb transmission channel.

The present paper describes investigations to calculate impairment factors *Ie,wb* for wb speech codecs. The calculation is performed in three steps: Firstly, the *R*-scale is extended on the basis of auditory test data. Secondly, impairment factors are derived on the extended *R*-scale for a number of G.722 and AMR codecs. Thirdly, the impairment factors are checked as to whether they fulfil the additivity property, i.e. whether they can be added in case of multiple codec tandems.

Extension of the R-Scale

When test subjects rate the quality of speech stimuli transmitted over nb channels, their judgment depends on whether wb stimuli are present in the same test or not. Based on this observation, the idea for extending the *R*-scale is as follows: Quality ratings are collected for different nb speech stimuli both in a nb-only and in a mixed nb/wb test. The ratings are separately transformed to the *R*-scale, using the formula given in [2]. The *R*-scale values common in both tests can then be used to derive a theoretical maximum R_{max} value for a clean wb channel, expressed in terms of the nb *R*-scale, via a linear or curvilinear extrapolation.

This procedure has been applied to results of several nb and nb/wb tests described in the literature; references can be found in [3]. The principle is illustrated in Fig. 1: A scatter plot is produced for data points corresponding to nb conditions which were part of both (nb and mixed nb/wb) tests. An exponential line is drawn through the points minimizing the squared error, and the maximum R_{max} value of the nb scale (100) leads to the corresponding R_{max} value for the clean wb connection. For two tests carried out by France Télécom, this leads to R_{max} = 129.

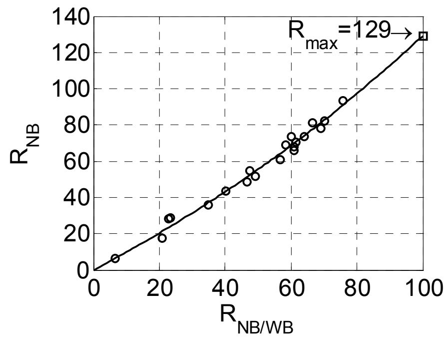

Figure 1: Extension of the R-scale based on results in [3].

Slightly lower values were derived from other tests, namely R_{max} = 120 or 112. In those cases, only two common conditions were part of both (nb and nb/wb) tests, justifying the use of a straight extrapolation line through the origin. For the following calculations, the most conservative estimate of

112 was used, i.e. a 12% extension of the *R*-scale when going from nb to wb transmission.

Derivation of Wideband Impairment Factors

Starting from the maximum R_{max} value for the clean wb channel, impairment factors for three wb speech codecs have been determined, using results of five tests carried out at 3GPP (References in [4]). These three codecs will serve as anchors for other wb codecs. The procedure is as follows:

1. MOS values obtained in the respective test are transformed to *R* values in the nb range [0;100], using the transformation given in [2].
2. The derived *R* values are expanded to the wb range [0;112] via a linear scaling (· 1.12).
3. For each test condition, a wb impairment factor *Ie,wb* is determined by subtracting the *R* value of the test condition from the one of either
   - a clean wb channel:
     $Ie, wb = R(\text{clean wb}) - R(\text{test condition})$, or
   - a standard nb channel (including G.711 coding and a default noise floor), and then adding 18.8 which corresponds to the difference between the clear wb (112) and the standard nb channel (92.2):
     $Ie, wb = R(\text{standard nb}) - R(\text{test condition}) + 18.8$

Depending on the considered test and methodology, the procedure leads to slightly different values for the first three wb speech codecs. Averaged values are depicted in Fig. 2. It has to be noted that the derivation procedure results in a fixed relationship between nb (*Ie*) and wb (*Ie,wb*) impairment factors; thus, the new *Ie,wb* values rely on the same additivity principle that is used for nb channels.

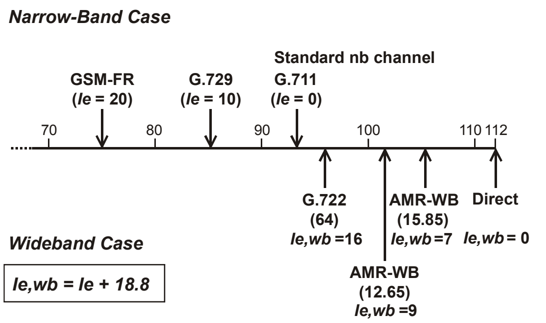

Figure 2: Anchoring of wb impairment factors on the *R*-scale.

Using the three first wb codecs as anchors, impairment factors have been derived for a number of other AMR and G.722 codecs. The following table lists the values derived for the other conditions of the mentioned 3GPP tests.

Codec	*bit-rate*	*Ie,wb*	*# tests*
G.722	64	16	5
G.722	56	18	2
G.722	48	24	4
G.722.1	24	12	3
AMR-WB	6.6	29	4
AMR-WB	8.85	18	4
AMR-WB	12.85	9	5
AMR-WB	14.25	7	4
AMR-WB	15.85	7	5
AMR-WB	18.25	5	4
AMR-WB	19.85	2	4
AMR-WB	23.05	2	3
AMR-WB	23.85	5	3

Table 1: Provisional impairment factors for wb codecs.

Additivity Check

The derived impairment factors do not necessarily satisfy the additivity assumption. Tab. 2 shows *Ie,wb* values derived from circuit conditions with codec tandems and compares them to the corresponding sum of values given in Tab. 1. From the grey-shaded conditions, it becomes obvious that the additivity is not fully satisfied. In particular, the order of codecs in a row has an impact on the overall degradation perceived by the listener. This effect is not taken into account by the impairment factor principle.

Test Condition	*Ie,wb tandem*	*Σ Ie,wb individual*
AMR-WB(12.65)*G.722(64)	33	25
AMR-WB(12.65)*G.722(48)	41	33
AMR-WB(12.65)*G.722.1(24)	29	21
G.722(48)*AMR-WB(12.65)	29	33
AMR-WB(15.85)*G.722(64)	28	23
AMR-WB(15.85)*G.722(48)	39	31
AMR-WB(15.85)*G.722.1(24)	26	19
G.722(48)*AMR-WB(15.85)	29	31

Table 2: Impairment factors for wb codec tandems.

Conclusions

We presented investigations to extend the impairment factor principle to wideband transmission, by expanding the "perceptual" quality scale underlying this principle. It was shown that integral quality may rise by about 12 to 29% when switching from a narrow-band to a wideband channel, expressed in terms of impairment factor units. On the basis of this assumption, new impairment factors were derived for several wideband speech codecs. These factors may be used to get a rough estimate of overall speech quality for network planning purposes, but they do not describe the perceptual effects in detail. For example, they disregard order effects which seem to be important for codec tandems.

Literatur

[1] Möller, S. Assessment and prediction of speech quality in telecommunications. Kluwer Academic Publ., Boston MA, 2000.

[2] ITU-T Rec. G.107. The E-model, a computational model for use in transmission planning. International Telecommunication Union, Geneva, 2005.

[3] Raake, A. Assessment and parametric modelling of speech quality in voice-over-IP networks. Doctoral dissertation (unpubl.), IKA, Ruhr-Universität Bochum, Bochum, 2005.

[4] ITU-T Del. Contr. D.029. Preliminary Equipment Impairment Factors for Wideband Speech Codecs. Federal Republic of Germany (S. Möller and A. Raake), ITU-T SG12 Meeting, Geneva, Jan. 2005.