Predicting Speech Quality under Noise for a Wideband E-Model

Juan-Pablo Ramirez, Alexander Raake, Marcel Wältermann and Sebastian Möller
Quality and Usability Lab, Deutsche Telekom Laboratories, TU Berlin, Ernst-Reuter-Platz 7, 10587 Berlin, Germany
{ Juan-Pablo.Ramirez, Alexander.Raake, Marcel.Waeltermann, Sebastian.Möller } @ telekom.de

Introduction

Speech signal transmission over telephone networks is described by multiple parameters that depict the benefits and the impairments affecting the signal transmitted from the mouth of the talker to the ear of a listener. Each of these characteristics infers both direct and more complex impacts on the quality perceived by users. Their variety was increased while telecommunication technologies developed, now including mobile and packet-based technologies; a thorough listing can be found in [1].
Planning modern networks led to the drawing up, via subjective tests, of algorithms assessing the quality of transmissions as users would typically do. For this aim, the E-model [2] is the most commonly used in narrowband network planning (NB, $400 - 3500$ Hz), mapping the characteristics of a transmission line onto a 0 to 100 scale ($R-scale$). The assessment of the line's quality impact is split into five psychological factors whose effects, on the resulting quality scale, are additive:

$$R = R_0 - I_s - I_d - I_e + A \qquad (1)$$

Here, R_0 handles the initial score for a given line configuration, capturing the effect of the signal-to-noise ratio. The so called impairment factors I_x quantify the other impairments occurring on the transmission path [6]. A is an advantage factor. So far, only NB network planing is possible with the E.model. For future wide-band services (WB, $50 - 7000$ Hz), a WB E-model is needed. Recent works described in [1] and [5] showed that the overall quality in WB was improved by 29% in comparison with NB in case of optimal transmission. This is the starting point of our study, consisting of two listening only tests on the impact of noise in WB transmission.
This paper will describe the test procedure and the conditions submitted to its attendees, before presenting the results and their analysis. The tests' conclusion will introduce a path between modeling quality in NB and in WB transmission. In the end, some results on the behavior of codecs under noise will be discussed.

Signal–to–noise ratio and E-model

Among the most trivial features of a transmission line are the ones defining its output's basic signal-to-noise ratio, depicted by the send loudness rating (SLR), and the addition of the different noise sources impairing the signal. Their effects are quantified in the E-model by the term R_0 (1). Not that our study does not consider the case where the signal level is too high, perceived as such as impairment by the user. For NB, this effect is handled by the simultaneous impairment factor I_s, requiring further work for WB.
The E-model considers the influences of noise and of signal attenuation as being linear on the NB $R - scale$, according to equation (2).

$$R_0 = 15 - 1.5 \cdot (SLR + N_0), \qquad (2)$$

where

$$N_0 = 10 \log_{10}[10^{\frac{N_c}{10}} + 10^{\frac{N_{os}}{10}} + 10^{\frac{N_{or}}{10}} + 10^{\frac{N_{fo}}{10}}]. \qquad (3)$$

Here, N_c is the sum of all the circuit noise powers. N_{os} and N_{or} are transformed ambient noises at send and receive side, respectively. N_{fo} is the "noise floor" at receive side. All these noise sources are expressed in $dBm0p$. For further details on these characteristics, refer to [2] and [3].

Test procedure

Two listening only tests, each involving 24 students without hearing impairments, were performed according to ITU-T Rec. P.800 (1996), using the 5-point "Absolute Category Rating" method. Short sentences (Eurom K) were recorded (2 female, 2 male) and processed using simple PCM coding, as well as the codecs mentioned in figure 2. A set of references (WB codecs with different bit rates, packet loss, NB $MNRU$ and $G.711$ type coding) was included for data transformation based on quality expectations found in literature. The Nc and $Nfor$ noise sources were generated using the PSTN/ISDN/Mobile Network Simulation unit described in [4]. Ps noise sources were recorded in a cafeteria and a car's passenger cell.

Results and discussion

The raw output of the test consisted of a set of integer scores ranging between 1 and 5 for each condition, which were averaged over the subjects. The scores were normalized to lie between 1 and 4.5 according to [1] Annex D.3, and mapped onto an intermediate scale between 0 and 100 using the transformation rule given in [2] Annex B. A linear function minimizing the distance between expected and obtained scores of the reference conditions was calculated. It scaled the overall scores on the interval 0 to 129, on which WB transmission quality is to be assessed, while adjusting the scores of well known conditions to the values commonly agreed to the ITU-T. Despite their various natures, noise sources presented similar impacts on the quality perceived by the users. The test showed that the cafeteria atmosphere impaired the transmission by 10

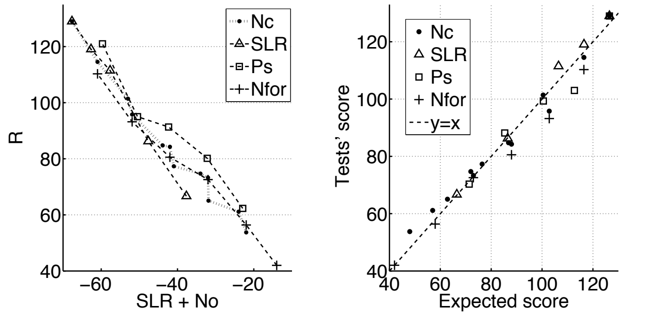

Figure 1: Results of the test. Comparison with the expected scores given by the model in equation (4).

points more than the inside car environment. Alike the NB case, the dependence of the quality (R_0) as a function of the noise and the speech levels can reasonably be considered as linear. If the SLR is set to 8, its default value, the impact of noise follows the law $R_0 = c_1 - 1.5 \cdot N_0$. Considering the sound level only, an accentuated slope allows more accuracy: $R_0 = c_2 - 2 \cdot SLR$. This second assumption calls into question the linear dependence between signal-to-noise ratio and users' quality judgment, presumed in the NB E-model. The combination of both approximations leads to the equation (4). Adding a coefficient of 29 lead to the best approximation, minimizing the $RMSE$ ($RMSE = 4$, $Pearson = 0.99$).

$$R_0 = 29 - 2 \cdot SLR - 1.5 \cdot N_0 \qquad (4)$$

This linear modeling conceives network planning algorithms in WB transmission in a similar fashion as the NB model did, thus initiating an easily computable up-dating of the E-model to WB, implying eased comparability of both bandwidth performances. Regarding effect of intermediate bandwidths on quality, see works depicted in [7]. Path with models of speech intelligibility under noise and bandwidth restrictions (SII, AI, STI) are under study.

NB and WB transmission performances, when submitted to noise, can now be compared in terms of quality perceived by the users. If SLR is set to its default value, R_0 takes its maximum value for N_0 set at the threshold $T_{WB} = -77dBm0p$ and will saturate at 129 if the noise level is decreased. In the NB case, R_0 would saturate at 93.2 for a noise level below $T_{NB} = -60dBm0p$. The ear of the user is found to be more receptive to noise in enlarged bandwidth. For $T_{WB} > N_0 > T_{NB}$, the reward of WB over G.711 type of coding decreases linearly from 29 to 10 points. Once N_0 is above T_{NB}, we consider the gain in quality as constant and equal to 10. The asset of WB over NB described above is illustrated by the set of conditions processed through G.711, as shows figure 2.

The second test included conditions with signal coding. The E-model estimates the impairment due to codecs quality distortion as an additive factor (I_e in (1)) independent from the other characteristics of the transmission, e.g. noise or bandwidth. The test validates the impairment factors attributed to each codecs in unnoised conditions in current ITU-T Rec. G.113, App. IV ($RMSE = 2, Pearson = 0.99$). It shows as well that codecs' performances are reduced in presence of noise, whose influence on quality is not anymore linear. Adding I_e, independent of the line's other characteristics, does not present a valid approximation.

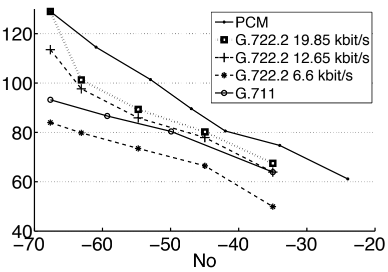

Figure 2: Scores given to several codecs in circuit noise.

Conclusion

A model adapted to WB transmission with line noise and ambient noise at send side was proposed, in the same line of thoughts as the E-model in NB. Subjective assessment of quality in NB and WB bandwidth showed high similarities, with the difference that noise, being more perceptible in enlarged bandwidth, impacts the users' opinion in a wider range. The model proposed distinguishes the impacts of noise and signal attenuation on the quality perceived by the users, breaking off with the traditional signal-to-noise ratio linear dependance considered in the NB E-model.

References

[1] Raake A.: *Speech Quality of VoIP – Assessement and Prediction.* Wiley, UK–Chichester, West Sussex, 2006.

[2] ITU-T Rec. G.107: *The E-model, a Computational Model for Use in Transmission Planning.* International Telecommunication Union, CH-Geneva, 2005.

[3] ETSI Technical report ETR 250: *Transmission and Multiplexing (TM); Speech Communication Quality from Mouth to Ear for 3,1 kHz Handset Telephony across Networks.* European Telecommunications Standards Institute, F-Sophia Antipolis, 1996.

[4] Möller S., *Assessment and Prediction of Speech Quality in Telecommunications.* Boston MA: Kluwer Academic Publ., 2000.

[5] Möller S.; Raake A.; Kitawaki N.; Takahashi A. and Wältermann M.: *Impairment Factor Framework for Wide-Band Speech Codecs.* In: IEEE Trans. Audio Speech and Language Proc.

[6] Möller S.: *Wie gut ist breitbandige Sprachübertragung wirklich? Ein Lagebericht zur Erweiterung des E-Modells.* In: Proc. DAGA 2008.

[7] Wältermann M. und Raake A.: *Ein neuer Qualitätsfaktor für die Modellierung linearer Verzerrungen bei der Übertragung von Sprache.* In: Proc. DAGA 2008.