Performance of Instrumental Speech Quality Measures for Next Generation Wireless Networks

Blazej Lewcio, Marcel Wältermann, Pablo Vidales, Alexander Raake, and Sebastian Möller
Deutsche Telekom Laboratories, Germany, Email: blazej.lewcio@telekom.de

Introduction

Next Generation Wireless Networks will provide a seamless access to VoIP services for nomadic users. Being on the move, user devices will be enabled to roam across heterogeneous wireless networks during active VoIP calls. The mobile VoIP service use will introduce a time-varying aspect to the VoIP user perception. As the user moves towards or away from the point of attachment to the network, the speech quality conditions change. On one hand, leaving the coverage area of a wireless network will require a heterogeneous network handover to maintain an active VoIP call. On the other hand, getting closer to a good-performance network will provide a possibility to improve the user experience.

As a consequence of a network handover, the network circumstances may suddenly change, which in turn may require a speech codec changeover. The effectiveness of narrowband codec adaptation to compensate for packet loss conditions was already studied in [1]. The authors show that a changeover from the ITU-T G.711 to the GSM codec, when certain packet loss value is exceeded, can improve the perceived call quality. Another study of user perception in heterogeneous NGN is presented in [2]. Here, the authors provide guidelines for heterogeneous mobility and wideband-narrowband codec changeover planning, which allows to achieve maximal user satisfaction in NGNs.

In this context of VoIP service distribution, the quality monitoring is an important element. To enable speech quality prediction, instrumental models have been developed [3]. According to the functional principle of these models, two main groups can be separated: parametric and signal-based models. (1) The parametric models, such as the E-Model [4], predict the user perception based on transmission characteristics measured in a system under test, or estimated from planning assumptions. According to the parameters, such as packet loss or network delay, a non-intrusive quality monitoring is possible, which enables to predict user experience without having direct access to the audio signal. (2) In contrast to the parametric quality prediction, a reliable signal-based quality estimation requires access to the audio signal [5]. The signal-based quality prediction models, such as PESQ [6] and TOSQA [7], compare the input and output audio signal of a system under test to provide a speech quality estimate.

Both parametric and signal-based models were inherently designed for quality prediction of narrowband speech quality. To capture the quality degradation of the being nowadays widely deployed wideband VoIP services, extensions of these models have been proposed in [8], [9], and [10]. Nevertheless, the current prediction models have not been designed for estimating of time-varying conditions, nor have they been fully validated for application in a wireless NGN environment.

In this paper, we evaluate the applicability of the current instrumental quality prediction methods in NGNs. We show how accurate they can handle time-varying quality parameters. We compare the instrumental quality estimates with the real user opinion obtained during conducted listening-only tests. In this way, we are able to validate the model's accuracy and disclose their limitations in quality monitoring for the forthcoming Next Generation Networks.

The paper is structured as follows. The next section presents the test set-up, the structure of the test material, and the listening test procedure. As next, the evaluation study of the current quality prediction models is described. The last section concludes the paper and presents the plan for the future work.

NGN Environment

For the evaluation of the current quality prediction models under NGN conditions (wideband-narrowband switching, network handover, and packet loss changes), a twofold approach has been used in order to map transmission characteristics to user perception. In the first evaluation part – perceptual evaluation – a set of speech samples was generated, and judged in a listening experiment to obtain Mean Opinion Scores (MOSs). In the second evaluation part – networking evaluation – network transmission metrics were measured. The recorded audio signals and captured network parameters were used for quality estimation using the quality prediction models under test:

- wideband E-Model (WB-E-Model)
- wideband PESQ (WB-PESQ)
- wideband TOSQA (WB-TOSQA)

Finally, the results of both preceding parts were merged to evaluate the obtained correlation of all quality estimates.

The speech samples were generated in the Mobisense testbed [11], where the transmission characteristics can be controlled in real-time during experiments. The audio stream was transmitted between two stations Mobile and Correspondent Node (MN and CN) on an emulated NGN platform. During ongoing VoIP calls the MN

could roam across heterogeneous access networks, and initiate a seamless wideband-narrowband speech codec switching [12]. In addition, the network characteristics in particular networks could be degraded by artificially generated packet loss.

The listening test conducted to assess user perception in NGNs consisted of two parts. In the first part, a call quality test according to ETSI TR 102 506 v.1.1.1 was performed. This test consisted of 2 sessions of approximately 25 min. Thereafter, a quality rating of short samples (1 session corresponding to approx. 25 min and 26 test conditions), according to the ITU-T Rec. P.800, was carried out. For the evaluation presented in this paper, results of the short sample quality tests are used. The call quality test is not further explained in this document. However, in [2] more insights are presented.

The short (6 s) test samples consist of two quality levels, which are equal in duration, but represent different transmission technology, network characteristics, and speech codecs applied for the transmission. The networking parameters were manipulated by the handovers between heterogeneous wireless networks (WiFi and HSDPA), and have been additionally degraded by packet loss in the first network for selected conditions. According to the networking situations, ITU-T G.722.2 at 23.05 kbit/s was employed as the wideband speech codec, whereas ITU-T G.711 was applied as the narrowband codec; interested reader may find in [13] the detailed list of test conditions.

Quality prediction in NGN

The results of the listening test and the quality estimates of the involved models have been compared on the MOS scale to evaluate their accuracy and limitations in NGNs. The test conditions were grouped as follows: (1) *Reference* conditions employ one network and one speech codec for the entire transmission, (2) *Network handover* conditions use one speech codec and network handover in the middle of a sample, (3) *Codec changeover* conditions transmit in one network and applies a wideband-narrowband codec changeover, (4) *Mixed* conditions introduce network handover and wideband-narrowband codec changeover simultaneously in the middle of a sample. In addition, selected samples were affected by certain amounts of packet loss in the first network [2].

Figure 1 presents the correlation of WB-PESQ model quality estimates and user opinion, judged on a 5-point absolute category scale according to ITU-T Rec. P.800. One can see a considerably small underestimation that affects all of the test conditions. The conditions that employ only one codec lie near the optimum line. However, for the codec switching conditions, the underestimation increases. As well, conditions that combine codec switching with network switching conditions are underestimated. In general, WB-PESQ obtains a relative high correlation, however, provides less accurate estimation for wideband-narrowband speech codec switching conditions.

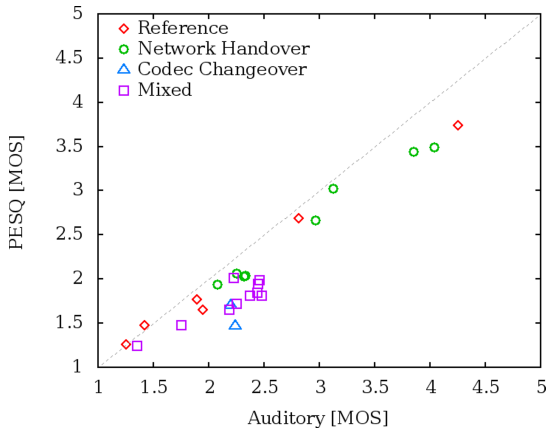

Figure 1: Comparison of WB-PESQ quality estimations and auditory quality scores.

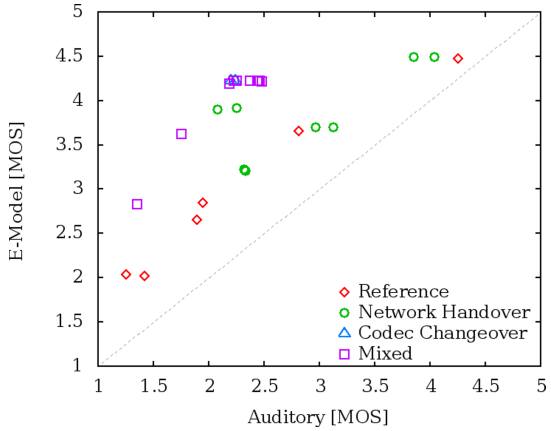

Figure 2: Comparison of WB-E-Model quality estimations and auditory quality scores.

The second model under test, the E-Model, is a planning tool designed to predict user perception when one speech codec is used only. Therefore, the quality score for this study was obtained by computation of the mean quality score over two quality levels of a speech sample (see previous section). The mean value provides unchanged quality estimates if the two quality levels are equal, and allows to take different quality levels into consideration. The resulting quality estimates have not been normalized in any way [13]. Figure 2 compares the real and the predicted quality judgments by the E-Model. An overestimation of all the test conditions can be observed. The conditions that employ only one speech codec, namely reference conditions and network handover conditions, are estimated in a best manner, and create a parallel function to the optimum line. However, if the wideband-narrowband speech codec changeover occurs, the overestimation considerably increases.

The quality estimates of the last evaluated model, WB-TOSQA, and of the conducted listening test are shown in Figure 3. A non-stable behavior of the model and relative low correlation can be observed. The model underestimates and overestimated the reference and network handover conditions, and provides a considerably large deviation from the optimum line. However, if the codec switching occurs, all of the conditions are consistently underestimated.

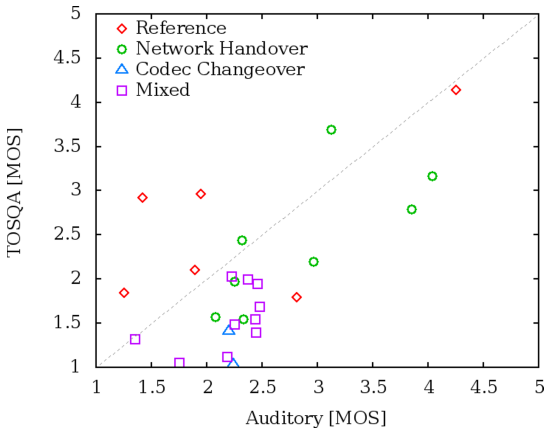

Figure 3: Comparison of WB-TOSQA quality estimations and auditory quality scores.

Table 1 compares the correlation results of the analyzed quality prediction models. The WB-PESQ provides the best quality estimation under NGN conditions, which is indicated by Pearson correlation of $r = 0.956$ and Root Mean Square Error of $RMSE = 0.471$, obtained for the conducted test. The second in the rank, the parametric E-Model, provides lower correlation of $r = 0.685$ and higher $RMSE = 1.462$, which disqualifies this model for reliable use in wireless NGNs, at least in the simple form we have applied it here, i.e. by averaging the MOS estimations of two degradation conditions which are combined in one sample. The worst correlation results are achieved by WB-TOSQA ($r = 0.597$ and $RMSE = 0.793$); this model can apparently also not be applied for quality prediction in wireless NGNs.

Model	Pearson	RMSE
WB-PESQ	0.956	0.471
E-Model	0.658	1.462
TOSQA	0.597	0.793

Table 1: Pearson correlation and the Root Mean Square Error provided by the evaluated quality prediction models.

Conclusions and Future Work

The seamless VoIP distribution in the Next Generation Wireless Networks will introduce new service quality aspects. The possibilities of dynamic VoIP service adaptation due to the variety of access networks and applied speech codecs will require monitoring instruments to provide high-quality VoIP service. In this paper, we have evaluated the accuracy of selected quality prediction models, namely Wideband PESQ, Wideband TOSQA, and the Wideband E-Model, in a wireless NGN environment. We have shown that WB-PESQ is the only candidate for restricted use in NGNs. None of the other models under test, the WB-E-Model and WB-TOSQA, provides a reliable quality estimation in wireless NGN environment. However, it should be noted that none of the models was ever designed to be used in such scenarios, and that the simple application of the E-model - by averaging two distinct MOS predictions for the different sample degradation levels - may be far too simplistic.

Because the presented study aimed to evaluate and disclose limitations of current quality prediction models in Next Generation Wireless Networks, a potential improvement of these models will be addressed in the future research. Therefore, more tests are planned to overcome the limitations of our study, and to provide more detailed output for the required modeling. In this way, the analysis helps to quantify VoIP quality perception of nomadic next generation network users, and contributes to high-quality VoIP service maintenance in Next Generation Wireless Networks.

References

[1] S. Leng Ng, S. Hoh, and D. Singh, "Effectiveness of Adaptive Codec Switching VOIP Application over Heterogeneous Networks," in *2nd International Conference on Mobile Technology, Applications and Systems*, Nov. 2005, pp. 1–7.

[2] S. Möller, M. Wältermann, B. Lewcio, N. Kirschnick, and P. Vidales, "Speech Quality while Roaming in Next Generation Networks.," (Accepted for publication at ICC 2009).

[3] Alexander Raake, *Speech Quality of VoIP: Assessment and Prediction*, John Wiley & Sons, Sept. 2006.

[4] ITU-T Rec. G.107, *The E-model, a Computational Model for Use in Transmission Planning*, International Telecommunication Union, Aug. 2008.

[5] A. W. Rix, J. G. Beerends, D.S. Kim, P. Kroon, and O. Ghitza, "Objective Assessment of Speech and Audio Quality - Technology and Applications," *IEEE Transactions on Audio, Speech & Language Processing*, vol. 14, no. 6, pp. 1890–1901, Nov. 2006.

[6] ITU-T Rec. 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, Feb. 2001.

[7] ITU-T Contribution COM 12-34-E, *TOSQA - Telecommunication Objective Speech Quality Assessment*, ITU-T SG12 Meeting, Dec. 1997.

[8] ITU-T Rec. G.107 - Amendment 1, *New Appendix II - Provisional Impairment Factor Framework for Wideband Speech Transmission*, International Telecommunication Union, June 2006.

[9] ITU-T Rec. P.862.2, *Wideband Extension to Recommendation P.862 for the Assessment of Wideband Telephone Networks and Speech Codecs*, International Telecommunication Union, Nov. 2005.

[10] ITU-T Contribution 12-19-E, *Results of Objective Speech Quality Assessment of Wideband Speech Using The Advanced TOSQA 2001*, ITU-T SG12 Meeting, Dec. 2000.

[11] P. Vidales, N. Kirschnick, F. Steuer, B. Lewcio, M. Wältermann, and S. Möller, "Mobisense Testbed: Merging User Perception and Network Performance," in *Proceedings of TRIDENTCOM 2008*, Mar. 2008.

[12] M. Wältermann, B. Lewcio, P. Vidales, and S. Möller, "A Technique for Seamless VoIP-Codec Switching in Next Generation Networks," in *Proceedings of ICC 2008*, May 2008, pp. 1772–1776.

[13] B. Lewcio, M. Wältermann, S. Möller, and P. Vidales, "E-Model Supported Switching Between Wideband and Narrowband Speech Quality," (Submitted for QoMEX 2009).