Speech Quality Prediction for Artificial Bandwidth Extension Algorithms

Sebastian Möller [1], *Emilia Kelaidi* [1], *Friedemann Köster* [1], *Nicolas Côté* [2], *Patrick Bauer* [3], *Tim Fingscheidt* [3], *Thomas Schlien* [4], *Hannu Pulakka* [5,6], *Paavo Alku* [5]

[1] Quality and Usability Lab, Telekom Innovation Labs, TU Berlin, Germany
[2] Institute of Electronics, Microelectronics and Nanotechnology, ISEN Department, Lille, France
[3] Institute for Communications Technology, Technische Universität Braunschweig, Germany
[4] Institut für Nachrichtengeräte und Datenverarbeitung, RWTH Aachen, Germany
[5] Dept. of Signal Processing and Acoustics, Aalto University, Finland
[6] Nokia Smart Devices, Finland

{sebastian.moeller|friedemann.koester}@telekom.de, kelaidi.aimilia.23@gmail.com, nicolas.cote@isen.fr, {patrick.bauer|t.fingscheidt}@tu-bs.de, schlien@ind.rwth-aachen.de, {hannu.pulakka|paavo.alku}@aalto.fi

Abstract

During the transition period from narrowband to wideband speech transmission services, Artificial Bandwidth Extension (ABE) algorithms are able to reduce the perceptual degradation of narrowband-transmitted speech signals by extending the audio bandwidth. In this paper, we analyze whether the resulting speech quality can be predicted reliably with instrumental models. Estimations from the new ITU standard POLQA, its predecessor WB-PESQ and the diagnostic DIAL model are compared to subjective listener judgments. This comparison reveals that the instrumental measures are not fully able to cope with ABE-processed speech, particularly in predicting ABE rank orders reliably. Reasons for this finding and corresponding diagnoses are discussed.

Index Terms: speech quality, artificial bandwidth extension, instrumental quality prediction, speech transmission, diagnosis

1. Introduction

Due to the large-scale introduction of Voice-over-Internet-Protocol (VoIP), speech transmission services are currently undergoing a substantial change. Whereas the transmitted audio bandwidth in the Public Switched Telephone Network was, with some exceptions, nearly always limited to the 300-3400 Hz bandwidth (called narrowband in the following), this limitation is no longer necessary in VoIP. Instead, speech may be transmitted with wideband (50-7000 Hz), super-wideband (50-14000 Hz) or even fullband (20-20000 Hz) audio bandwidth. Subjective listening tests have shown that the resulting speech quality can be increased by approx. 30% for wideband [1][2] and up to 79% for super-wideband transmission [3] compared to the narrow audio bandwidth.

However, users of wideband services can only take advantage of the improved quality in case their interlocutors also use the same type of service. This makes wideband telephony a “critical-mass service”, being only adopted if a minimum number of users subscribe to it. For reducing the adoption threshold, Artificial Bandwidth Extension (ABE) algorithms have been developed which are able to artificially generate higher and/or lower frequency components on the basis of a narrowband-transmitted speech signal. While the reconstruction is not perfect in the artificially-generated frequency band, the resulting perceptual effect is notable and was shown to considerably improve quality [4][5][6][7] and intelligibility [4][8][9][10].

In order to judge the quality of ABE algorithms, subjective listening-only or conversational tests are the only recommended methods [11][12]. Such tests provide valid and reliable values for the achieved quality; however they require controlled test laboratory conditions and are time-consuming and expensive to carry out. To cope with the disadvantages of subjective testing, instrumental models have been developed which are able to estimate quality – as it would be judged in a subjective experiment – on the basis of signals or parameters. The most well-known type of model is based on a perceptual comparison of the input and output signals of the transmission channel which shall be judged regarding its quality. This type of model has been recommended since a long time by the Telecommunication Standardization Sector of the International Telecommunication Union, ITU-T [13][14][15] but its predictive power has to our knowledge never been analyzed for ABE algorithms. Other types of models aim at diagnosing perceptual impairments [16], but also their behavior for ABE algorithms is unknown. Still, instrumental quality prediction of ABE-enhanced speech would be very useful for ABE developers (allowing tuning of the algorithms) and telephone service providers (allowing to plan and monitor quality) alike.

In this paper, we analyze the prediction accuracy of three representatives of such models on speech extended by a selection of state-of-the-art ABE algorithms. For this purpose, we carried out a subjective listening-only test and compared its results to the instrumental quality predictions. Section 2 describes the prediction models used, and Section 3 provides a short overview on the ABE algorithms tested in the study. Section 4 summarizes the subjective test set-up, and Section 5 analyzes the results. Section 6 discusses implications of our results and options for future work.

2. Quality prediction models

We used two types of models for the analysis. The first type of model predicts the overall quality of the transmitted speech signal, in terms of an average rating commonly obtained on a 5-point Absolute Category Rating (ACR) scale in a listening-only test according to [11], a Mean Opinion Score (MOS). The second type of model aims at identifying perceptual dimensions of the degradations first; these dimensions may help in diagnosing sub-optimum quality, but they also can be combined to an overall quality prediction (MOS).

2.1. Overall quality prediction

The idea of the investigated overall quality prediction models is to compare the input and the output signals of a transmission system under investigation: The larger the distance or incoherence, the lower the assumed quality. The transmission system may include the codec, other speech processing, or even the terminals including their electro-acoustic characteristics and potentially the talking/listening room. ABE algorithms have not yet been addressed by these models, so we perform strictly speaking an out-of-domain usage.

The comparison starts with a pre-processing step where input and output signals are time- and level-aligned and pre-filtered so that signal differences which do not or only marginally affect perception are ruled out before the comparison. The comparison is usually based on a simplified model of human peripheral auditory perception (non-linear frequency analysis, loudness model, spectral and/or temporal masking, etc.); perceptual versions of the input and output signals are then compared on a frame-by-frame basis, the distance (or incoherence) is integrated over the entire length of the signal, and then transformed onto the MOS scale ranging from 1 ("bad") to 5 ("excellent").

A number of models follow this approach. A popular example is former ITU-T Standard P.862, which comes in a narrowband version (PESQ, [14]) and a wideband version (WB-PESQ, [17]). This standard was recently superseded by ITU-T Rec. P.863, POLQA [15]. In addition to the perception model and distortion calculation, the POLQA model calculates six internal parameters which are then integrated into the distortion in a post-processing step.

2.2. Dimension-based quality prediction

An explicit modeling of perceptual degradations is provided by the diagnostic model DIAL [16]. This model is based on the assumption that the perceptual distances between stimuli can be displayed in a multidimensional space. Wältermann et al. [18] identified three such dimensions as "coloration", "noisiness" and "discontinuity", and later a fourth dimension "loudness" was added [19]. For each of these dimensions, Côté [16] implemented an estimator in the following way:

- *Coloration* is estimated via an Equivalent Rectangular Bandwidth (in bark) and the center frequency of the linear part of the transmission path (in Hz), following the ideas of Raake [2].
- *Noisiness* is estimated via two parameters: the noise loudness in silence, and a Noise-on-Speech parameter.
- *Discontinuity* is estimated with Weighted Spectral Slope distances and a signal-temporal loss to derive an interruption rate, an artifact rate and a clipping rate.
- *Loudness* of speech at non-optimum level is estimated via the long-term loudness of the speech signal.

The estimates of the four distortion types (MOS-C, MOS-N, MOS-D and MOS-L) are integrated with a core model estimate (analyzing distortions as a perceptual coherency between input and output signal, as explained above) to form an estimate of the overall quality, in terms of MOS. The integration is done by a *k*-nearest neighbors (*k*NN) algorithm.

3. ABE algorithms

Two out of the five evaluated ABE algorithms are based on a phoneme-specific Hidden Markov Model (HMM) training acc. to [20], which provides an intelligibility gain on critical fricatives [9][10]. Both ABE versions exclusively perform highband extension. Additionally, one of these versions includes a phonetic classifier to enhance speech quality by attenuating the upper frequency band during potential artifacts.

A third ABE algorithm is based on the source-filter speech production model [21] and thus split in two parts [8]: On the one hand the linear prediction residual of the narrowband signal is extracted, normalized, mixed with white noise and then used as the excitation for the artificial highband signal. On the other hand with the help of 14 narrowband features extracted from the narrowband signal (MFCCs, zero crossing rate) an HMM estimator with 128 states and 16 Gaussian mixture components per state estimates the autoregressive model filter coefficients for the spectral envelope of the highband of order 4.

ABE methods have also been studied in collaboration between Aalto University and Nokia, e.g. [4][5][6], and the two other methods evaluated in this work have been developed in this collaboration. One of the methods extends speech to the highband (4-8 kHz) using an excitation signal based on spectral folding and subsequent shaping of the highband spectrum. The other method additionally generates artificial low-frequency content below about 300 Hz.

All of the ABE algorithms have been trained on speech transcoded with the 12.2 kbps mode of the Adaptive Multirate (AMR) speech codec, equivalent to GSM-EFR. Please note that the order of presentation here is *not* related to the order of the ABE algorithms in Table 1.

4. Subjective test set-up

In order to have a ground truth for the prediction models, a subjective listening-only test according to [11] has been carried out. In the test, 20 naïve listeners rated 120 speech files which have been transmitted over channels with 15 different processing conditions. The channels used correspond to a combination of a speech codec, a simulation of packet loss, and potentially an ABE at the receiver side. The list of circuit conditions is given in Table 1.

Each condition was tested with 8 different source speech files of approx. 10 s recorded in a sound-insulated room from 4 female and 4 male speakers. Text passages consisted of German versions of the EUROM sentences [22] which are phonetically balanced, and which have been slightly shortened for the purpose of the test. These files are slightly longer than it is recommended for the prediction models, but were preferred over short sentences to give the listeners a better impression of the speech sound quality which was the main focus of the speech material.

For each circuit condition, a source speech file with super-wideband audio bandwidth was first band-limited to narrowband or wideband, depending on the codec which was applied. The narrowband signals were then pre-filtered with a G.712 bandpass filter [23] and an IRSsend filter [12], simulating a typical frequency response of a handset phone. Wideband signals were instead filtered with the corresponding wideband bandpass filter of ITU-T Rec. P.341 [24]. The pre-processed files were level-aligned to -26 dB active speech level relative to the overload point of the digital system, then coded with the corresponding codec, and in some cases random packet loss was inserted with a fixed percentage *Ppl*

Table 1: Test conditions, subjective judgments and instrumental predictions.

No.	Circuit	Subjective judgm.		Overall quality predictions			Dimension predictions			
		MOS	std.	WB-PESQ	POLQA	DIAL	MOS-C	MOS-N	MOS-D	MOS-L
1	G.711	3.00	0.80	2.88	3.21	3.21	3.02	4.38	4.31	4.39
2	GSM-EFR	2.70	0.84	2.35	2.83	2.91	3.00	4.36	3.61	4.38
3	Clean WB	4.24	0.90	4.40	4.06	4.11	3.97	4.44	4.31	4.40
4	G.722@64	3.82	0.95	4.32	4.16	3.96	4.04	4.38	4.31	4.39
5	G.722.2@23.05	3.59	0.88	3.73	3.87	3.95	3.91	4.34	4.24	4.39
6	G.722.2@6.6	1.96	0.79	2.53	2.91	2.72	3.66	4.12	3.32	4.36
7	G.711+ ABE3 +*Ppl* (10%)	1.59	0.90	1.83	2.16	2.94	3.46	4.16	3.58	4.39
8	G.722+ *Ppl* (10%)	1.99	1.04	1.99	2.43	2.98	4.02	4.19	3.18	4.38
9	G.711+ ABE1 +*Ppl* (10%)	1.98	0.90	1.79	2.01	2.71	3.24	3.97	2.87	4.24
10	G.711+ABE1	3.00	0.82	2.57	2.66	3.18	3.37	4.23	3.94	4.25
11	G.711+ABE2	1.96	0.84	1.63	2.34	2.86	3.26	4.20	3.06	4.33
12	G.711+ABE3	2.25	0.90	2.31	2.60	3.21	3.46	4.18	4.00	4.39
13	G.711+ABE4	3.04	0.76	2.87	2.92	3.21	3.07	4.36	4.25	4.39
14	G.711+ABE5	2.88	0.78	2.71	3.04	3.20	3.17	4.28	4.03	4.39
15	G.711+*Ppl* (10%)	1.96	0.78	1.99	2.47	2.97	3.00	4.35	3.86	4.39

of lost packets. No packet loss concealment was applied at the decoder side. The coded data streams were then decoded, in some cases, where G.711 has been used as codec, the ABE algorithm was applied, and then the files were presented to the listeners. Note that for *all* of the ABE approaches this means a codec mismatch between training of the ABE algorithms and their test conditions.

The test took place in a low-noise test room at the premises of Deutsche Telekom. Test participants were welcomed, informed about the topic of the test, and instructed about the usage of the test GUI which allowed to play each stimulus once, and then to rate the stimulus on a continuous rating scale with the five attributes of the MOS scale, as described in [25]. The continuous scale layout was preferred over the standard category rating layout as the layout prevents saturation effects otherwise occurring at the scale extremities, see discussion in [25]. Before the main test started, five speech files of different processing conditions and speakers were presented to the participants in order to accustom them to the range of qualities to be experienced in the test, and to the rating procedure. Then, participants rated the 120 files each in a differently randomized order. Listening took place diotically through a Sennheiser HMD headphone at 73 dB sound pressure level. A test session was split into two parts separated by a short break, and lasted less than one hour. Test participants were compensated for their service with a gift voucher.

5. Results

In this section, we will first analyze the results of the subjective test which can be regarded as a ground truth for the prediction models. Then, the performance of the prediction models will be assessed, separating the overall quality predictions from the diagnostic ones.

5.1. Subjective judgments

Subjective judgments have first been analyzed with respect to their distribution and outliers. Two participants have been excluded from further analysis, as they did not rate several test conditions, leaving 18 ratings per file, and 144 ratings per test condition. The ratings have then been averaged over samples and over test conditions, see Table 1. A one-way ANOVA shows that both text material ($F=12.62$, $p<0.01$) and speaker ($F=12.33$, $p<0.01$) have a significant impact on the judgments; this is why a variety of speakers and text materials was necessary. The stronger influence was however caused by the processing conditions ($F=121.65$, $p<0.01$).

The subjective judgments show a clear advantage of the wideband over the standard narrowband (G.711) channel. This advantage shrinks when wideband speech coding such as G.722 or G.722.2 at 23.05 kbit/s is applied, and the wideband channel may even become worse than G.711 with the G.722.2 codec at its lowest (6.6 kbit/s) bitrate. Similarly, the quality drops when narrowband GSM-EFR coding is applied, or when packet loss occurs (in both narrowband and wideband).

Interestingly, none of the ABE algorithms manages to significantly improve the MOS over that observed for G.711. In the optimum case, ABE4 reaches a slight (but not significant) increase over the MOS of 3.0 observed for G.711; ABE1 obtains the same MOS as G.711, and all other ABEs obtain lower MOS ratings. This finding may be due to the particular set-up of the test which includes – apart from different audio bandwidths – also a pre-filtering and codec mismatch to the pre-processing in the ABE trainings and packet-loss degradations. As a result, it is probable that the test participants have not only concentrated to the coloration of the speech signal, but also on dimensions such as noisiness and discontinuity, which might have been impacted by the ABEs. The results might have been different in a paired-comparison judgment situation, where differences in only one perceptual dimension might have been at the basis of the judgments. This hypothesis will be further analyzed in Section 5.3.

5.2. Overall quality predictions

For each pair of source and processed speech files, instrumental MOS predictions have been calculated with WB-PESQ, POLQA and DIAL. For WB-PESQ, the estimations of the model have been transformed to the MOS scale using the transformation law given in [17]. For POLQA, the native transformation rule has been applied. No 3rd-order mapping was used. The requirements for the source files according to Appendix II of [15] have been met.

The results from all WB-PESQ, POLQA and DIAL confirm the superiority of the clean wideband compared to the

narrowband channel. All models also predict the degradations due to coding in a more-or-less correct rank order, with slight order changes for POLQA (clean WB/G.722@64 and G.722.2@6.6/ GSM-EFR) and WB-PESQ (G.722.2@6.6/GSM-EFR). Furthermore, all models correctly predict the degrading effect of packet loss.

However, all WB-PESQ, POLQA and DIAL show problems in correctly predicting the effects of different ABE algorithms, such as their rank order. WB-PESQ slightly overestimates degradations generated by ABE, in providing a little too pessimistic predictions. POLQA estimations are more realistic, but still are not able to predict the rank order to different ABEs in all cases correctly. DIAL seems to have problems in differentiating between the ABE algorithms; 4 out of 5 algorithms are predicted to be very similar to G.711 regarding their overall quality.

Overall, WB-PESQ reveals a significant Pearson correlation of r=0.942 ($p<0.01$) with the subjective test results, whereas POLQA and DIAL yield slightly lower correlations of r=0.916 ($p<0.01$) and r=0.904 ($p<0.01$), respectively. Focusing on the ABE conditions 7 and 9-14 in Table 1, WB-PESQ achieves a higher correlation (r=0.925) than POLQA (r=0.868) and DIAL (r=0.753). This proves that in comparison to POLQA and DIAL, WB-PESQ seems to be slightly better able to estimate the effects of ABE algorithms. In spite of the correlation above 0.90, indicting an acceptable prediction performance on the entire data set, ABE rank orders were not predicted reliably. This fact severely limits the practical value of instrumental models for optimizing or selecting ABE algorithms.

5.3. Diagnostic quality predictions

In addition to overall quality estimates, DIAL also provides estimations of 4 perceptual dimensions. As these dimensions have not been assessed in the auditory test, there is no ground truth which would validate the predictions. However, some of the prediction results are intuitive. Exemplarily, the narrowband conditions mainly score below optimum on coloration, and packet-loss conditions on discontinuity. G.722.2@6.6 kbit/s is an example for a codec that also causes discontinuity.

The perceptual dimensions which are most affected by all ABEs are coloration and discontinuity. As expected, DIAL predicts the coloration of ABE algorithms to range between all narrowband and wideband channels; thus, DIAL acknowledges the extended bandwidth with a higher score compared to narrowband, but still below optimum wideband. All ABE algorithms seem to introduce only a bit noisiness, strongest for ABE3. ABE2 is predicted to cause high discontinuity, whereas ABE1 scores slightly worse than optimum on loudness.

6. Conclusions and future work

The aim of this paper was to analyze the prediction accuracy of different types of quality prediction models for speech enhanced by state-of-the-art ABE algorithms. For this purpose, a standard listening-only test was carried out, and the results were compared to model predictions.

Interestingly, none of the ABE algorithms reached a significant improvement compared to G.711 (condition 1) regarding the overall MOS. This result is in contrast to results presented earlier, which showed an improvement of ABE-processed speech compared to narrowband. A potential explanation is that – when ABEs are judged subjectively – they are commonly presented in comparison to the clean speech or coded speech only, frequently in a paired-comparison paradigm. In our test, ABE conditions have been put into a context of NB, WB and NB-extended speech files, as well as in a context of packet-loss speech. Thus, the judgments of our test may reflect an integration of various types of distortions, and not only the effect of improved bandwidth.

Another potential explanation could be that we used G.712 bandpass filtering and IRSsend filtering of the narrowband conditions for simulating the frequency response of the sending terminal in NB; this is not always the case in other subjective tests on ABE-processed speech and can produce a mismatch to the pre-processing in ABE trainings, which might employ different or additional filter masks. Another significant issue and explanation is that all of the ABE algorithms have been trained with narrowband speech material that was transcoded using the AMR/GSM-EFR speech codec, as opposed to the G.711 which was used in this test. Consequently, 3 out of 5 ABEs reached a higher MOS than GSM-EFR. The fact that ABEs may reach higher scores than GSM-EFR is confirmed by other subjective tests (paired comparison and ACR type, see [4][5][9]).

The subjective test results were – to some extent – confirmed with WB-PESQ, POLQA, and DIAL. As a matter of fact, WB-PESQ seems to better cope with ABE effects than POLQA and DIAL. The overall correlations between subjective quality judgments and instrumental predictions are all above 0.90 for all models. Still, none of the models is able to correctly predict the rank order of the ABEs. This severely limits their practical value for ABE optimization and selection.

Helpful information for ABE optimization may come from the diagnostic predictors of DIAL. They indicate that the ABE-enhanced signals are still sub-optimal with respect to the coloration and discontinuity. Although this result could not be explicitly validated with a subjective ground truth, we expect such diagnostic predictions to have substantial practical value.

The question arises whether instrumental predictions can be improved for the use with ABEs. One potential enhancement could be not to use the clean wideband channel (condition 3) as the input of the prediction model, but to use a band-limited version of it. In addition, the integration of diagnostic predictors into the overall quality prediction of the DIAL model might be improved.

In the future, we consider repeating the analysis with results from conversation tests. In such tests, the situation is more realistic, and the subjective results will reflect the “true” quality ranking even better than in a listening-only situation. We would also like to validate the dimension predictions on DIAL by comparing them to subjective dimension ratings, using the procedure proposed in [19]. This way, we hope to produce an instrumental model which provides valid and reliable estimations of overall quality, as well as diagnostic information for ABE developers and appliers.

7. References

[1] Möller, S., Raake, A., Kitawaki, N., Takahashi, A., Wältermann, M., "Impairment Factor Framework for Wideband Speech Codecs", IEEE Trans. Audio, Speech and Language Processing 14(6):1969- 1976, 2006.

[2] Raake, A., Speech Quality of VoIP — Assessment and Prediction, John Wiley & Sons, Chichester, West Sussex, 2006.

[3] Wältermann, M., Raake, A., Möller, S., "Extension of the E-Model Towards Super-Wideband Speech Transmission", in: Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP 2010), Dallas TX, 14-19 Mar, 2010.

[4] Laaksonen, L., Kontio, J., Alku, P., "Artificial Bandwidth Expansion Method to Improve Intelligibility and Quality of AMR-Coded Narrowband Speech", in: Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP 2005), Philadelphia, PA, USA, 18-23 Mar, 2005.

[5] Pulakka, H., Laaksonen, L., Vainio, M., Pohjalainen, J., Alku, P., "Evaluation of an Artificial Speech Bandwidth Extension Method in Three Languages", IEEE Trans. Audio, Speech and Language Processing 16(6): 1124-1137, 2008.

[6] Kontio, J., Laaksonen, L., Alku, P., "Neural Network-based Artificial Bandwidth Expansion of Speech", IEEE Trans. Audio, Speech, Language Process., vol. 15, no. 3, pp. 873-881, Mar. 2007.

[7] Pham, T.V., Schaefer, F., Kubin, G., "A Novel Implementation of the Spectral Shaping Approach for Artificial Bandwidth Extension" in: Proc. Int. Conf. on Communications and Electronics (ICCE 2010), Nha Trang, Vietnam, 11-13 Aug, 2010.

[8] Heese, F., Geiser, B., Vary, P., "Intelligibility Assessment of a System for Artificial Bandwidth Extension of Telephone Speech", Proc. German Ann. Conf. on Acoustics (DAGA), 2012.

[9] Bauer, P., Jung, M.-A., Qi, J., Fingscheidt, T., "On Improving Speech Intelligibility in Automotive Hands-Free Systems", in: Proc. IEEE Int. Symp. on Consumer Electronics (ISCE 2010), 1-5, Braunschweig, Germany, 7-10 Jun, 2010.

[10] Bauer, P., Fischer, R.-L., Bellanova, M., Puder, H., Fingscheidt, T., "On Improving Telephone Speech Intelligibility for Hearing Impaired Persons", in: Proc. ITG Conf. on Speech Communication (ITG 2012), 275-278, Braunschweig, Germany, 26-28 Sep, 2012.

[11] ITU-T Rec. P.800, Methods for Subjective Determination of Transmission Quality, Int. Telecomm. Union, Geneva, 1996.

[12] ITU-T Rec. P.830, Subjective Performance Assessment of Telephone-band and Wideband Digital Codecs, Int. telecomm. Union, Geneva, 1996.

[13] ITU-T Rec. P.861, Objective Quality Measurement of Telephone-Band (300-3400 Hz) Speech Codecs. Int. Telecomm. Union, Geneva, 1998.

[14] 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, Int. Telecomm. Union, Geneva, 2001.

[15] ITU-T Rec. P.863, Perceptual Objective Listening Quality Assessment, Int. Telecomm. Union, Geneva, 2011.

[16] Côté, N., Integral and Diagnostic Intrusive Prediction of Speech Quality, Springer, Berlin, 2011.

[17] ITU-T Rec. P.862.2, Wideband Extension to Recommendation P.862 for the Assessment of Wideband Telephone Networks and Speech Codecs, Int. Telecomm. Union, Geneva, 2007.

[18] Wältermann, M., Raake, A., Möller, S., "Quality Dimensions of Narrowband and Wideband Speech Transmission", Acta Acustica united with Acustica 96(6):1090-1103, 2010.

[19] Wältermann, M., Dimension-based Quality Modeling of Transmitted Speech, Springer, Berlin, 2013.

[20] Bauer, P., Fingscheidt, T., "A Statistical Framework for Artificial Bandwidth Extension Exploiting Speech Waveform and Phonetic Transcription", in: Proc. Europ. Signal Processing Conf. (EUSIPCO 2009), 1839-1843, Glasgow, Scotland, 24-28 Aug, 2009.

[21] P. Jax and P. Vary, "On artificial bandwidth extension of telephone speech," Signal Processing, IEEE, vol. 83, no. 8, 2003.

[22] Gibbon, D., "EUROM.1 German Speech Database," ESPRIT Project 2589 Report (SAM, Multi–Lingual Speech Input/Output Assessment, Methodology and Standardization), Universität Bielefeld, Bielefeld, 1992.

[23] ITU-T Rec. G.712, Transmission Performance Characteristics of Pulse Code Modulation Channels, Int. Telecomm. Union, 2001.

[24] ITU-T Rec. P.341, Transmission Characteristics for Wideband Digital Loudspeaking and Hands-free Telephony Terminals, Int. telecomm. Union, 2011.

[25] Möller, S., Assessment and Prediction of Speech Quality in Telecommunications, Kluwer Academic Publishers, Boston MA, 2000.