Speech Quality Evaluation of Artificial Bandwidth Extension: Comparing Subjective Judgments and Instrumental Predictions

Hannu Pulakka[1], *Ville Myllylä*[1], *Anssi Rämö*[2], *and Paavo Alku*[3]

[1]Microsoft Phone Technologies, Tampere, Finland
[2]Nokia Research Center, Tampere, Finland
[3]Department of Signal Processing and Acoustics, Aalto University, Espoo, Finland
hannu.pulakka@microsoft.com

Abstract

Artificial bandwidth extension (ABE) methods have been developed to enhance the quality and intelligibility of bandlimited speech transmitted over a telephone connection. Subjective listening tests are the most reliable way of evaluating the quality of ABE, but listening tests are time-consuming and expensive to arrange. Instrumental measures have also been used to estimate the subjective quality of ABE. This study extends the results of an earlier subjective evaluation of ABE methods by instrumental quality predictions computed with WB-PESQ (ITU-T Recommendation P.862.2) and POLQA (ITU-T Recommendation P.863). The instrumental quality predictions are compared with the subjective quality scores. The results indicate that POLQA correlates better with the subjective quality than WB-PESQ. Neither WB-PESQ nor POLQA can predict the rank order of the evaluated ABE methods in all conditions.

Index Terms: artificial bandwidth extension, subjective evaluation, listening test, instrumental quality assessment

1. Introduction

Speech transmission in telephone networks is traditionally limited to *narrowband* speech with an audio frequency band restricted below 4 kHz. For example, the adaptive multi-rate (AMR) codec [1] transmits only narrowband speech and is widely employed in mobile networks. Superior quality and intelligibility is provided by *wideband* speech transmission covering the frequency range 50–7000 Hz. Wideband speech services are increasingly available in mobile telephone networks [2] commonly using the adaptive multi-rate wideband (AMR-WB) speech codec [3]. Furthermore, several speech codecs have been developed for the transmission of *superwideband* speech with an audio frequency range up to about 14 kHz. Examples of superwideband codecs include ITU-T G.722.1 Annex C [4], ITU-T G.718 Annex B [5], and Opus [6].

Artificial bandwidth extension (ABE) methods have been developed to improve the quality and intellibility of bandlimited speech. ABE reconstructs the missing spectral content using only the bandlimited speech signal as input and can be used at the receiving terminal. A number of ABE methods have been proposed for the extension of narrowband speech to the wideband frequency range (NB-to-WB), e.g., [7, 8, 9, 10]. Recently, ABE methods for extending wideband speech to the superwideband range (WB-to-SWB) have also been proposed [11, 12, 13].

Development and deployment of ABE calls for reliable methods to assess the effect of ABE on speech quality. Subjective listening tests are the primary means of speech quality assessment. ABE is commonly evaluated with the same listening test methods that are used for the quality assessment of speech codecs, such as the absolute category rating (ACR) test and the comparison category rating (CCR) test described in [14]. However, listening tests are time-consuming and expensive to organize. Instrumental measures that estimate the subjective quality using computational models provide an attractive alternative. Reliable instrumental rank prediction of ABE variants would have high practical value in developing and optimizing ABE algorithms. Instrumental measures reported in ABE publications range from simple spectral distance metrics, such as the log-spectral distance (LSD) [8, 15, 16, 17, 18], to more advanced methods modeling the human perception [19, 20, 17, 18]. Two instrumental assessment methods are of interest in this paper: the wideband extension of the perceptual evaluation of speech quality (WB-PESQ) defined in ITU-T recommendation P.862.2 [21], and the perceptual objective listening quality assessment (POLQA) defined in ITU-T recommendation P.863 [22]. WB-PESQ has been used to evaluate ABE, e.g., in [17, 18].

The usability of WB-PESQ and POLQA for the quality assessment of ABE was investigated in [23] and [24]. The experiments indicated significant correlations between subjective and instrumental scores in general. However, the correlations were clearly lower when only ABE conditions were considered. Moreover, the instrumental methods were unable to reliably rank the evaluated ABE methods, which limits the applicability of the quality prediction methods for selecting and optimizing ABE algorithms. This paper provides additional results on the feasibility of WB-PESQ and POLQA for the assessment of ABE. The results of the subjective evaluation presented in [25] are compared with instrumental predictions of speech quality based on WB-PESQ and POLQA. The quality assessments are performed in a context of different audio bandwidths and a variety of standardized speech codecs. This study also includes test conditions for WB-to-SWB ABE as well as background noise conditions that were not investigated in [23] or [24].

2. Subjective quality assessment

Subjective listening tests were arranged to evaluate the quality of ABE-processed speech in relation to narrowband, wideband, and superwideband speech codecs. The subjective evaluation and its results were presented in [25]. The listening test procedure was similar to that used for codec evaluation in [26, 27].

2.1. ABE methods

The following ABE methods were evaluated:

ABE1 is the NB-to-WB ABE method described in [10]. A neu-

ral network is used to estimate the spectral shape of the extension band from input features. An excitation signal is generated from the linear prediction residual of the narrowband input by spectral folding, and a time-domain filter bank technique is used to shape the spectrum.

ABE2 is based on the structure of ABE1, but the neural network was replaced by a hidden Markov model and piecewise linear mapping from input features to the spectral shape parameters of the extension band. Also, input features and the synthesis filter bank were modified.

SWB-ABE is a WB-to-SWB ABE method based on ABE2 with some modifications: The input features were selected for the WB-to-SWB ABE task and the synthesis filter bank was designed for the extension band 7–15 kHz. The excitation is generated from spectrally replicated linear prediction residual and white noise.

2.2. Test conditions

The following test conditions were included in the evaluation:

Direct reference conditions with limited audio bandwidth but no speech coding. Four lowpass cutoff frequencies were evaluated: 4 kHz, 7 kHz, 10 kHz, and 14 kHz.

AMR narrowband codec [1] commonly employed in mobile networks. Four bit rate modes were evaluated: 4.75 kbit/s, 7.95 kbit/s, 10.2 kbit/s, and 12.2 kbit/s.

AMR + ABE: AMR codec followed by ABE processing. Three ABE variants were tested: ABE1, ABE2, and ABE2b that refers to the ABE2 method with the extension band attenuated by 5 dB. Each ABE variant was evaluated with two bit rate modes of the AMR codec: 7.95 kbit/s and 12.2 kbit/s.

AMR-WB codec [3] for wideband speech, currently supported in an increasing number of mobile networks [2]. Four bit rate modes were evaluated: 6.6 kbit/s, 8.8 kbit/s, 12.65 kbit/s, and 23.85 kbit/s.

AMR-WB + SWB-ABE: AMR-WB codec followed by SWB-ABE processing. Three variants of the SWB-ABE method were generated by varying the attenuation of the extension band: SWB-ABEa (0 dB attenuation), SWB-ABEb (5 dB attenuation), and SWB-ABEc (10 dB attenuation). All three variants were evaluated in combination with two bit rate modes of the AMR-WB codec: 12.65 kbit/s and 23.85 kbit/s.

Opus [6], an open source codec supporting both variable and fixed bit rates. Four constant bit rates (CBR) were evaluated, and the corresponding audio bandwidths were determined by the codec: 10.2 kbit/s (narrowband, 4 kHz), 12.65 kbit/s (mediumband, 6 kHz), 16 kbit/s (wideband, 8 kHz), and 20 kbit/s (superwideband, 12 kHz).

ITU-T G.722.1 Annex C [4], a low-complexity superwideband voice codec with an audio bandwidth of 14 kHz. Two bit rate modes were tested: 24 kbit/s and 32 kbit/s.

ITU-T G.718 Annex B [5], an embedded (8–64 kbit/s) speech codec for narrowband, wideband, and superwideband services. Two bit rate modes with 14-kHz audio bandwidth were evaluated: 28 kbit/s and 40 kbit/s.

2.3. Listening tests

Three listening tests were organized with different background noise conditions and highpass filter types:

Test 1: Clean speech, highpass cutoff 150 Hz, 8 talkers (4 females, 4 males), sentence pairs of about 6 seconds.

Test 2: Clean speech, highpass cutoff 50 Hz, 8 talkers (4 females, 4 males), single sentences of about 4 seconds.

Test 3: Noisy speech, highpass cutoff 50 Hz, 4 talkers (2 females, 2 males), sentence pairs of about 7 seconds. Four noise types with signal-to-noise ratios of 15–20 dB.

Both highpass filters have a flat response in the passband. The filter with a 150-Hz cutoff simulates the response of a mobile terminal in the far end with highpass filtering to reduce low-frequency noise. A 50-Hz cutoff causes minimal low-frequency limitation and is commonly used in codec characterization tests.

A modified ACR test type with a discrete 9-point scale was used. The 9-point scale has been found to saturate less easily than the standard 5-point scale [26]. The tests took place in sound-proof booths in the listening test laboratory of Nokia Research Center [28]. Subjects listened to samples diotically through Sennheiser HD-650 headphones. The listening level was set to a sound pressure level of 76 dB and could not be adjusted by the listeners. A training session with 12 samples preceded each test. All speech samples were in Finnish. Twenty-eight listeners participated in each test.

3. Instrumental quality assessment

This study extends the results of the subjective evaluation by instrumental speech quality predictions of the test conditions. The speech quality of the listening test samples was estimated with the instrumental methods WB-PESQ [21] and POLQA [22].

3.1. WB-PESQ

ITU-T Recommendation P.862 [29] defines the perceptual evaluation of speech quality (PESQ) algorithm. PESQ computes an estimate of the subjective speech quality by comparing a degraded speech signal with the corresponding reference signal. The algorithm is based on a perceptual model motivated by the human auditory system and it generates a MOS-LQO value (mean opinion score, listening quality, objective) on a scale from 1 to 5. This is a prediction of a listening quality score that would be obtained in a subjective ACR listening test. A wideband extension (WB-PESQ) of the PESQ algorithm is described in ITU-T Recommendation P.862.2 [21]. The extension allows the evaluation of the frequency band 50–7000 Hz and predicts the subjective quality in a context of wideband speech.

WB-PESQ was used in this work to estimate the quality of listening test samples with bandwidth up to 7 kHz. Clean wideband speech samples with a frequency range of 50–7000 Hz were generated with the P.341 filter [30] and were used as reference signals (also for tests 1 and 3). According to [21], WB-PESQ should be used only with clean speech samples. Consequently, the scores calculated for test 3 have to be considered experimental due to out-of-domain usage of WB-PESQ.

3.2. POLQA

ITU-T recommendation P.863 [22] defines the perceptual objective listening quality assessment (POLQA) method for predicting the subjective speech quality of telephony systems. POLQA is the successor of PESQ and also based on a perceptual model. POLQA has two operation modes: narrowband (300–3400 Hz) and superwideband (50–14000 Hz). In the superwideband mode, a limitation of the audio band below the superwideband range is regarded as a degradation and scored accordingly. The

output of POLQA is a MOS-LQO score on a scale from 1 to 5. POLQA can be used to test also noisy speech, but the reference signal is always expected to be noise-free.

The superwideband mode of POLQA was used in this study. The reference signals were prepared from clean speech samples with the 50–14000-Hz bandpass filter available in [30]. Noise-free references were used also for test 3. Part of the samples did not fulfill the minimum duration recommended in [22].

4. Results

Table 2 presents the ACR listening tests results (MOS9) and the corresponding instrumental quality estimates computed with WB-PESQ and POLQA. Each instrumental score is the mean value over 32 speech samples in tests 1 and 2 and over 16 speech samples in test 3. Ninety-five percent confidence intervals (CI) are given. MOS9, WB-PESQ, and POLQA scores are not directly comparable due to different scales, and WB-PESQ scores are not available for superwideband conditions. Correlation coefficients between condition MOS9 values and condition-averaged instrumental scores are presented in Table 1. Correlations have been calculated for all test conditions and separately for only the NB-to-WB ABE conditions.

Table 1: *Correlation coefficients between subjective MOS9 values and instrumental predictions of WB-PESQ and POLQA.*

	all conditions			**NB-to-WB ABE**		
	test 1	test 2	test 3	test 1	test 2	test 3
WB-PESQ	0.821	0.847	0.359	0.937	0.888	0.701
POLQA	0.959	0.950	0.953	0.981	0.968	0.954

Figure 2 illustrates the relationship between subjective ACR scores and instrumental predictions. For a further comparison between subjective and instrumental scores, Figure 1 shows both subjective and instrumental scores of ABE-related conditions in test 1. ABE conditions with the same ABE algorithm but different attenuation of the extension band allows a comparison between changes in subjective and instrumental quality scores as a result of varying extension band level.

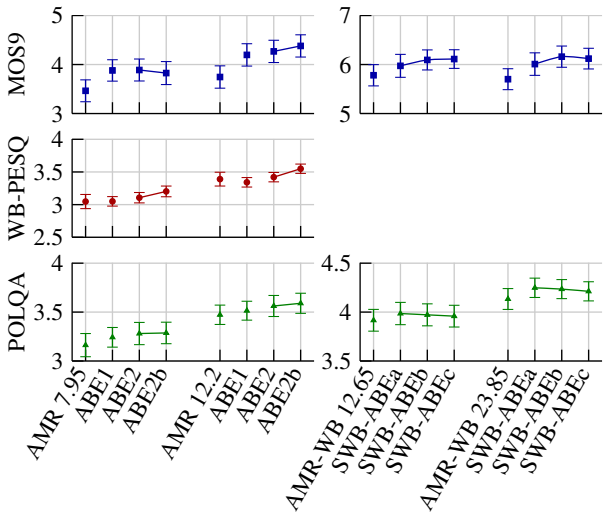

Figure 1: *Subjective scores (MOS9) and instrumental predictions of ABE conditions in test 1. The codec shown in the leftmost condition in each group is used also in the ABE conditions of the group. Conditions using the same ABE method but different extension band attenuation are connected with lines.*

5. Discussion

Subjective ACR scores (scale 1–9, superwideband context), WB-PESQ scores (1–5, wideband context), and POLQA scores (1–5, superwideband context) are not directly comparable. However, they should yield the same rank order between conditions. No mapping between the scales was used in this study.

Correlation coefficients presented in Table 1 indicate that POLQA outperforms WB-PESQ in terms of correlation with subjective scores. This is also reflected in Figure 2. Also, the estimation capability of WB-PESQ degrades remarkably for noisy speech in test 3 (Figure 2, top-right plot), but this was to be expected since WB-PESQ should be applied to clean speech [21].

Figure 2 suggests that the quality estimates of ABE conditions are in line with those of other conditions. However, the rank order of ABE variants is not reliably predicted by WB-PESQ or POLQA. For example, in test 3, POLQA indicates improved quality for increased level of SWB-ABE, whereas the ACR scores show an opposite trend. Moreover, WB-PESQ and POLQA do not always succeed in indicating whether ABE processing improves the subjective quality. For instance, the WB-PESQ scores of ABE1 in test 1 and the POLQA scores of SWB-ABE in test 3 suggest different preference than the ACR scores. However, the rank orders need to be considered with care because the score differences between ABE variants are small and many of the differences in scores are not statistically significant.

The instrumental quality estimates improve consistently with increasing bit rate of each codec, but quality estimates between codecs are not always consistent with subjective ratings. For example, subjective scores indicate that listeners preferred all AMR-WB conditions over all narrowband AMR conditions on average. Both WB-PESQ and POLQA, however, predict lower scores for AMR-WB at a low bit rate than for AMR at a high bit rate. This observation, together with ABE rank order differences, suggest that the instrumental methods weight bandwidth limitations and other degradations in a somewhat different way from human listeners in this study. It is worth noting that combining different kinds of degradations into a quality score is not straightforward for listeners, and the listening context and the instructions given to listeners may affect the results.

In both [23] and [24], WB-PESQ had a higher correlation with the subjective ratings of ABE conditions than POLQA. In this study, however, POLQA was found to correlate better with subjective ratings than WB-PESQ also in the NB-to-WB ABE conditions. However, the number of ABE conditions in this study is small and their quality scores are concentrated in a relatively small range of values. Furthermore, the quality scores are affected more by the codec bit rate than by the ABE variant.

6. Conclusions

This paper extends the results of the subjective evaluation presented in [25] by instrumental quality predictions computed with WB-PESQ and POLQA. In particular, the applicability of the instrumental measures in the assessment of NB-to-WB and WB-to-SWB ABE techniques is considered. In general, both WB-PESQ and POLQA have a reasonable correlation with subjective scores in clean speech conditions. POLQA correlates better with subjective scores than WQ-PESQ, and POLQA also gives reasonable results for noisy conditions. However, neither WB-PESQ nor POLQA can reliably predict the preference for using ABE or the rank order of ABE variants. Consequently, these instrumental measures cannot satisfactorily replace subjective tests in the evaluation of ABE algorithms.

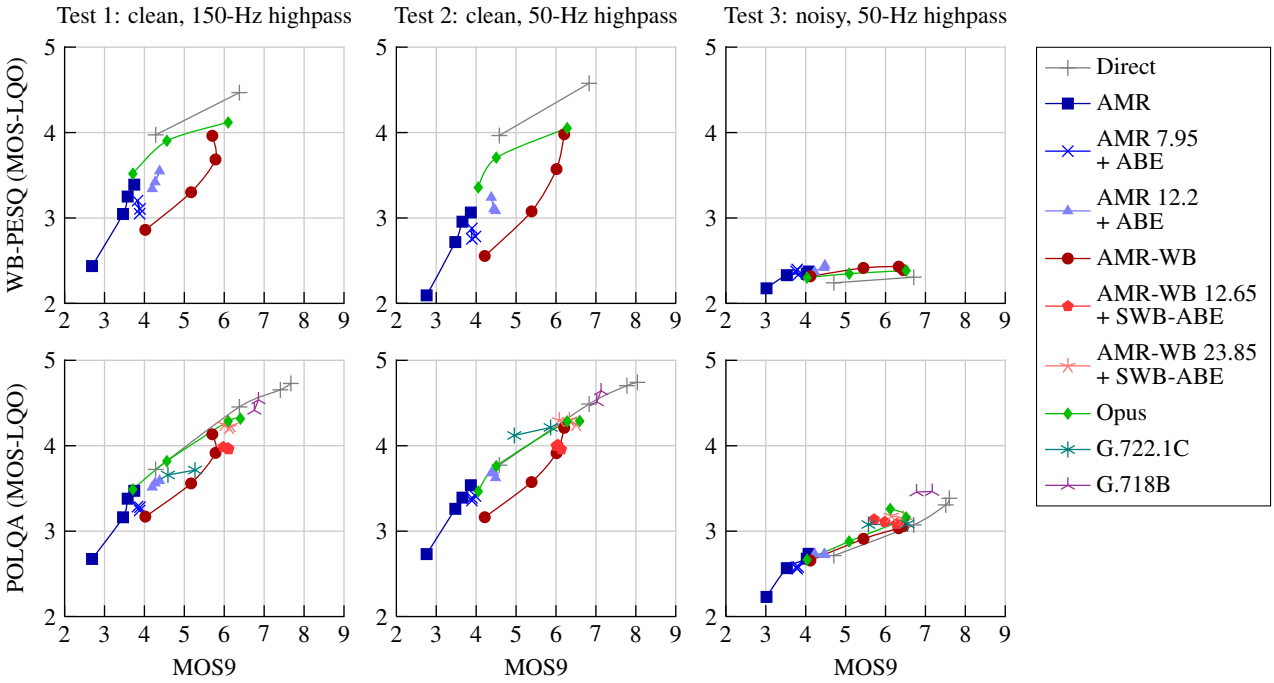

Figure 2: *Subjective scores (MOS9) and instrumental predictions (WB-PESQ and POLQA). Related conditions are connected by lines.*

Table 2: *Subjective scores (MOS9), instrumental predictions (WB-PESQ and POLQA), and 95% confidence intervals (CI).*

	test 1						**test 2**						**test 3**					
condition	MOS9	CI	WB-PESQ	CI	POLQA	CI	MOS9	CI	WB-PESQ	CI	POLQA	CI	MOS9	CI	WB-PESQ	CI	POLQA	CI
direct 14 kHz	7.67	0.23			4.73	0.01	8.04	0.19			4.74	0.01	7.60	0.33			3.38	0.27
direct 10 kHz	7.41	0.20			4.65	0.04	7.78	0.20			4.70	0.03	7.51	0.29			3.31	0.24
direct 7 kHz	6.38	0.20	4.47	0.02	4.45	0.05	6.83	0.22	4.58	0.01	4.49	0.05	6.71	0.34	2.31	0.18	3.07	0.22
direct 4 kHz	4.28	0.23	3.97	0.06	3.72	0.08	4.58	0.25	3.97	0.12	3.77	0.08	4.71	0.36	2.24	0.17	2.71	0.19
AMR 4.75	2.69	0.22	2.44	0.09	2.67	0.12	2.76	0.19	2.09	0.12	2.73	0.10	3.02	0.34	2.18	0.09	2.23	0.13
AMR 7.95	3.46	0.22	3.05	0.11	3.16	0.12	3.48	0.22	2.72	0.14	3.26	0.10	3.53	0.33	2.33	0.12	2.57	0.16
AMR 10.2	3.58	0.23	3.25	0.11	3.38	0.11	3.66	0.25	2.96	0.14	3.39	0.10	4.03	0.34	2.35	0.14	2.68	0.17
AMR 12.2	3.75	0.23	3.39	0.11	3.47	0.10	3.87	0.24	3.06	0.14	3.54	0.08	4.07	0.33	2.37	0.14	2.73	0.17
AMR 7.95 ABE1	3.88	0.22	3.05	0.07	3.24	0.10	3.89	0.21	2.76	0.10	3.36	0.10	3.81	0.33	2.33	0.10	2.58	0.12
AMR 7.95 ABE2	3.89	0.22	3.11	0.08	3.28	0.11	3.98	0.22	2.78	0.11	3.41	0.10	3.76	0.34	2.38	0.12	2.58	0.13
AMR 7.95 ABE2b	3.83	0.23	3.20	0.08	3.29	0.11	3.89	0.22	2.88	0.11	3.39	0.11	3.79	0.34	2.40	0.12	2.56	0.13
AMR 12.2 ABE1	4.20	0.23	3.34	0.07	3.51	0.10	4.49	0.23	3.08	0.09	3.62	0.09	4.22	0.33	2.37	0.11	2.73	0.13
AMR 12.2 ABE2	4.27	0.23	3.42	0.07	3.56	0.11	4.44	0.23	3.11	0.10	3.69	0.09	4.46	0.34	2.42	0.13	2.73	0.14
AMR 12.2 ABE2b	4.38	0.23	3.55	0.07	3.59	0.10	4.38	0.24	3.24	0.10	3.68	0.10	4.48	0.32	2.45	0.14	2.72	0.14
AMR-WB 6.6	4.02	0.23	2.86	0.09	3.17	0.12	4.21	0.24	2.56	0.10	3.16	0.09	4.12	0.36	2.32	0.11	2.66	0.15
AMR-WB 8.8	5.17	0.23	3.30	0.10	3.56	0.13	5.39	0.23	3.08	0.11	3.57	0.11	5.45	0.32	2.42	0.13	2.91	0.20
AMR-WB 12.65	5.78	0.22	3.68	0.09	3.92	0.11	6.01	0.22	3.57	0.10	3.91	0.11	6.33	0.33	2.43	0.15	3.03	0.19
AMR-WB 23.85	5.70	0.21	3.96	0.08	4.13	0.11	6.21	0.23	3.98	0.09	4.21	0.09	6.45	0.31	2.39	0.17	3.06	0.22
AMR-WB 12.65 SWB-ABEa	5.97	0.23			3.99	0.11	6.04	0.25			4.01	0.11	5.71	0.33			3.14	0.22
AMR-WB 12.65 SWB-ABEb	6.09	0.20			3.97	0.11	6.01	0.24			4.00	0.10	5.98	0.33			3.11	0.21
AMR-WB 12.65 SWB-ABEc	6.11	0.19			3.96	0.11	6.13	0.21			3.96	0.10	6.28	0.32			3.08	0.21
AMR-WB 23.85 SWB-ABEa	6.01	0.23			4.25	0.10	6.08	0.25			4.30	0.10	6.14	0.35			3.18	0.23
AMR-WB 23.85 SWB-ABEb	6.16	0.22			4.23	0.10	6.33	0.24			4.30	0.10	6.45	0.30			3.15	0.23
AMR-WB 23.85 SWB-ABEc	6.12	0.21			4.21	0.10	6.51	0.22			4.25	0.09	6.49	0.32			3.12	0.23
Opus 10.2 narrowband	3.71	0.23	3.52	0.10	3.49	0.09	4.05	0.24	3.36	0.15	3.46	0.07	4.04	0.34	2.30	0.14	2.67	0.17
Opus 12.65 mediumband	4.56	0.23	3.90	0.06	3.82	0.08	4.50	0.25	3.71	0.09	3.76	0.07	5.09	0.32	2.35	0.15	2.88	0.18
Opus 16 wideband	6.10	0.20	4.12	0.05	4.29	0.07	6.28	0.24	4.05	0.06	4.29	0.06	6.52	0.27	2.38	0.16	3.16	0.22
Opus 20 superwideband	6.41	0.26			4.32	0.06	6.59	0.23			4.29	0.07	6.12	0.35			3.26	0.22
G.722.1C 24	4.59	0.31			3.66	0.08	4.96	0.30			4.12	0.13	5.57	0.43			3.08	0.22
G.722.1C 32	5.27	0.29			3.72	0.11	5.87	0.27			4.21	0.14	6.55	0.36			3.08	0.22
G.718B 28	6.75	0.23			4.42	0.07	7.02	0.24			4.51	0.06	6.78	0.31			3.46	0.22
G.718B 40	6.86	0.24			4.54	0.05	7.13	0.23			4.64	0.04	7.17	0.26			3.46	0.24

7. References

[1] 3GPP TS 26.090, *Adaptive multi-rate (AMR) speech codec; Transcoding functions*, 3rd Generation Partnership Project, Sept. 2012, version 11.0.0.

[2] Global mobile suppliers association (GSA), "Mobile HD voice: Global update report," Sept. 2014, online: http://www.gsacom.com/downloads/pdf/GSA_mobile_hd_voice_190914.php4, accessed on 29 Sept. 2014.

[3] 3GPP TS 26.190, *Adaptive multi-rate wideband (AMR-WB) speech codec; Transcoding functions*, 3rd Generation Partnership Project, Sept. 2012, version 11.0.0.

[4] ITU-T G.722.1, *Low-complexity coding at 24 and 32 kbit/s for hands-free operation in systems with low frame loss*, Int. Telecommun. Union, May 2005.

[5] ITU-T G.718 Amendment 2, *Frame error robust narrow-band and wideband embedded variable bit-rate coding of speech and audio from 8–32 kbit/s; Amendment 2: New Annex B on super-wideband scalable extension for ITU-T G.718 and corrections to main body fixed-point C-code and description text*, Int. Telecommun. Union, Mar. 2010.

[6] J.-M. Valin, K. Vos, and T. B. Terriberry, "Definition of the Opus audio codec," IETF RFC 6716, Sept. 2012.

[7] H. Carl and U. Heute, "Bandwidth enhancement of narrow-band speech signals," in *Proc. EUSIPCO*, vol. 2, Edinburgh, UK, Sept. 1994, pp. 1178–1181.

[8] P. Jax and P. Vary, "On artificial bandwidth extension of telephone speech," *Signal Process.*, vol. 83, no. 8, pp. 1707–1719, Aug. 2003.

[9] K.-T. Kim, M.-K. Lee, and H.-G. Kang, "Speech bandwidth extension using temporal envelope modeling," *IEEE Signal Process. Lett.*, vol. 15, pp. 429–432, May 2008.

[10] H. Pulakka and P. Alku, "Bandwidth extension of telephone speech using a neural network and a filter bank implementation for highband mel spectrum," *IEEE Trans. Audio, Speech, Language Process.*, vol. 19, no. 7, pp. 2170–2183, Sept. 2011.

[11] B. Geiser and P. Vary, "Beyond wideband telephony – bandwidth extension for super-wideband speech," in *Proc. German Annual Conf. Acoust. (DAGA)*, Dresden, Germany, Mar. 2008, pp. 635–636.

[12] B. Geiser, "High-definition telephony over heterogeneous networks," Ph.D. dissertation, Rheinisch-Westfälische Technische Hochschule Aachen, 2012.

[13] B. Geiser and P. Vary, "Artificial bandwidth extension of wideband speech by pitch-scaling of higher frequencies," in *Workshop Audiosignal- und Sprachverarbeitung (WASP)*, Koblenz, Germany, Sept. 2013, pp. 2892–2901.

[14] ITU-T P.800, *Methods for subjective determination of transmission quality*, Int. Telecommun. Union, Aug. 1996.

[15] P. Bauer and T. Fingscheidt, "An HMM-based artificial bandwidth extension evaluated by cross-language training and test," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, Las Vegas, NV, USA, Mar. 2008, pp. 4589–4592.

[16] G.-B. Song and P. Martynovich, "A study of HMM-based bandwidth extension of speech signals," *Signal Process.*, vol. 89, no. 10, pp. 2036–2044, Oct. 2009.

[17] A. H. Nour-Eldin and P. Kabal, "Memory-based approximation of the Gaussian mixture model framework for bandwidth extension of narrowband speech," in *Proc. Interspeech*, Florence, Italy, Aug. 2011, pp. 1185–1188.

[18] C. Yağlı, M. A. T. Turan, and E. Erzin, "Artificial bandwidth extension of spectral envelope along a Viterbi path," *Speech Commun.*, vol. 55, no. 1, pp. 111–118, Jan. 2013.

[19] B. Iser and G. Schmidt, "Bandwidth extension of telephony speech," *EURASIP Newslett.*, vol. 16, no. 2, pp. 2–24, June 2005.

[20] H. Pulakka, L. Laaksonen, M. Vainio, J. Pohjalainen, and P. Alku, "Evaluation of an artificial speech bandwidth extension method in three languages," *IEEE Trans. Audio, Speech, Language Process.*, vol. 16, no. 6, pp. 1124–1137, Aug. 2008.

[21] I.-T. P.862.2, *Wideband extension to Recommendation P.862 for the assessment of wideband telephone networks and speech codecs*, Int. Telecommun. Union, Nov. 2007.

[22] ITU-T P.863, *Perceptual objective listening quality assessment*, Int. Telecommun. Union, Jan. 2011.

[23] S. Möller, E. Kelaidi, F. Köster, N. Côté, P. Bauer, T. Fingscheidt, T. Schlien, H. Pulakka, and P. Alku, "Speech quality prediction for artificial bandwidth extension algorithms," in *Proc. Interspeech*, Lyon, France, Aug. 2013.

[24] P. Bauer, C. Guillaumé, W. Tirry, and T. Fingscheidt, "On speech quality assessment of artificial bandwidth extension," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, Florence, Italy, May 2014, pp. 6082–6086.

[25] H. Pulakka, A. Rämö, V. Myllylä, H. Toukomaa, and P. Alku, "Subjective voice quality evaluation of artificial bandwidth extension: Comparing different audio bandwidths and speech codecs," in *Proc. Interspeech*, Singapore, Sept. 2014, pp. 2804–2808.

[26] A. Rämö, "Voice quality evaluation of various codecs," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, Dallas, TX, USA, Mar. 2010, pp. 4662–4665.

[27] A. Rämö and H. Toukomaa, "Voice quality characterization of IETF Opus codec," in *Proc. Interspeech*, Florence, Italy, Aug. 2011, pp. 2541–2544.

[28] M. Kylliäinen, H. Helimäki, N. Zacharov, and J. Cozens, "Compact high performance listening spaces," in *Proc. Euronoise*, Naples, Italy, May 2003.

[29] I.-T. 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. Telecommun. Union, Feb. 2001.

[30] ITU-T G.191, *Software tools for speech and audio coding standardization*, Int. Telecommun. Union, Mar. 2010.