SUBJECTIVE QUALITY EVALUATION OF THE 3GPP EVS CODEC

Anssi Rämö and Henri Toukomaa

Nokia Technologies, Tampere, Finland

anssi.ramo@nokia.com

ABSTRACT

This paper discusses the voice and audio quality characteristics of EVS, the recently standardized 3GPP codec. Comparison to Opus, IETF driven open source codec as well as industry standard voice codecs: 3GPP AMR and AMR-WB, and ITU-T G.718B, G.722.1C and G.719 as well as direct signals at varying bandwidths was made. Voice and audio quality was evaluated with three subjective listening tests containing clean and noisy speech in Finnish language as well as a mixed condition test containing both speech and music intermixed. Nine-scale subjective mean opinion score was calculated for all tested conditions.

Index Terms— speech coding, subjective evaluation, listening test, multi-rate codec, multi-bandwidth testing, mean opinion score

1. INTRODUCTION

In August 2014 3GPP SA4 accepted EVS (Enhanced Voice Services) codec as the next generation conversational codec for 3GPP Release 12 onwards [1][2]. The requirements for the EVS codec performance were quite strict [3], and there were tedious listening tests performed by three independent laboratories during the summer of 2014. Those results are available in the EVS selection phase GAL report [4]. However, those results were all done with a single bandwidth in each test. Also, Opus codec was not included in any of the listening tests, which is of great interest for many people in the speech coding community. Thus, Nokia performed three multi-bandwidth characterization listening tests in short notice in order to compare these two recent multi-bandwidth, multi-rate speech and audio codecs against each other. In addition, several standardized *narrowband* (AMR[5]), *wideband* (AMR-WB[6]), *superwideband* (ITU-T G.722.1 Annex C [7] and G.718B[8]) and *fullband* (ITU-T G.719[9]) codecs were tested as reference conditions. Modified 9-scale absolute category rating (ACR) test methodology was used for all experiments [10] [11].

1.1. EVS technical details

EVS codec supports four input and output sampling rates (8, 16, 32, and 48 kHz). There are also twelve bitrates ranging from 5.9 kbit/s to 128 kbit/s. 5.9 kbit/s mode is using VBR (Variable BitRate) with DTX always enabled and all other bitrates are CBR (Constant BitRate) where DTX functionality may be enabled. Frame error robustness is also optimized to a great degree providing significantly better frame error concealment performance than for example AMR-WB or G.718 [12][13]. Audio and speech coding modes are switched internally in realtime by the EVS codec depending on the input signal. Also enhanced voice quality AMR-WB interoperable mode is integrated to the EVS codec. More technical details can be found from EVS specification as well as other papers in ICASSP 2015 special session [14] [15] [16]. All of these features could not be incorporated into a single listening test. It was decided to test the most interesting EVS native bitrate range for all signal bandwidths with both clean and noisy speech as well as with generic audio signals in clean channel conditions. For example robustness to frame erasures is not discussed further in this paper. Additional listening test results from these other features of the EVS codec can be found in selection and characterization phase listening test results [4] [17].

A spectrogram in Figure 1 was concatenated from five three-second noisy speech segments showing all bandwidths of the EVS codec. For the lowest bitrates below 9.6 kbit/s only *narrowband* and *wideband* signal bandwidth is supported as can be seen from the two first spectrogram segments. For 9.6 and 13.2 kbit/s *superwideband* bandwidth of 14 kHz is supported (middle segment). For bitrates starting at 16.4 kbit/s coding upto Nyquist frequency of 16 kHz is enabled with EVS-SWB and upto 20 kHz with EVS-FB operation as shown by the last two spectrogram segments, respectively.

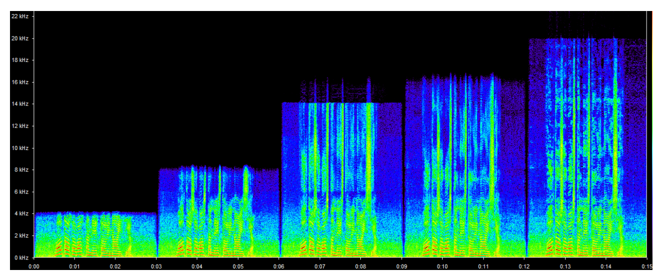

Fig. 1. *Spectrogram of a noisy speech segment with EVS codec at NB 7.2, WB 8.0, SWB 9.6, SWB 16.4, and FB 24.4 kbit/s.*

1.2. Opus technical details

Opus codec supports an approximately similar operation range as EVS codec. According to the Opus specification [18] supported bitrates range from 6 to 510 kbit/s including stereo and low delay options. For mono signals, 64 kbit/s should be enough for near transparent quality with a 20 ms frame size. Several versions of VBR and CBR operation are also available. Opus codec frame length can be adjusted between 2.5 and 60 ms. Opus supports 8, 12, 16, 24 and 48 kHz sampling rates with automatic resampling where needed.

All Opus processing in this listening test was done with 48 kHz sampling rate using 20 ms frame length. Opus codec internally decided at which bandwidth the codec operated at each requested bitrate. Opus was tested with both VBR and hard CBR modes. VBR mode was tested with requested bitrates from 5.9 to 32 kbit/s and CBR was tested from 13.2 to 64 kbit/s. At overlapping bitrates 13.2, 16.4, 24.4 and 32 kbit/s it can be seen how much the variable bitrate operation improves the subjective quality. The Opus complexity setting was set to the default value of 10 (maximum).

978-1-4673-6997-8/15/$31.00 ©2015 IEEE

Table 1. Opus VBR measured average and peak bitrates (in kbit/s) with different signal types. Total duration of the test databases was 14 minutes.

Bitrate	Clean speech	Noisy speech	Mixed music	Minimum bitrate	Maximum bitrate
5.9	6.95	7.098	7.69	2.4	13.2
7.2	7.234	7.465	7.811	2.4	13.2
8.0	7.609	7.881	8.247	2.4	12.8
9.6	8.751	9.418	9.474	2.4	14.0
13.2	12.04	13.02	12.97	3.2	22.0
16.4	15.36	16.34	16.28	6.8	25.6
24.4	23.2	24.59	24.62	1.2	40.0
32	30.37	32.65	32.23	1.2	63.6

Table 1 shows that Opus VBR actually quite often works slightly higher average bitrate than requested. Especially the lowest bitrates of 5.9, 7.2 and 8.0 kbit/s seem hard to reach especially with mixed music signals. In practice the lowest possible Opus VBR bitrate is around 7 kbit/s for speech and nearly 8 kbit/s for music signals. Also bitrate variation is quite high from frame to frame. For example when the requested average bitrate is 32 kbit/s, some of the frames are coded with 1.2 kbit/s and some as high as 63.6 kbit/s.

Coded signal bandwidths of Opus were measured with the listening test database. All signal types kept the constant bandwidth all the time and Table 2 was experimentally constructed. It seems that CBR requires about 1 kbit/s higher bitrate before enabling higher bandwidth operation.

Table 2. Opus bitrates and bandwidths with VBR and CBR modes.

Abbr	Bandwidth	Sampling rate in kHz	bitrate range in VBR mode in kbit/s	bitrate range in CBR mode in kbit/s
NB	narrowband	8	7- 10	7- 11
MB	mediumband	12	10.5- 12.5	11.5- 13.5
WB	wideband	16	13- 15.5	14- 16.5
SWB	superwideband*	24	16- 19	17- 20
FB	fullband	48	>=19.5	>=20.5

* in other context (ITU-T, 3GPP) SWB sampling rate is 32 kHz

Similarly to the EVS spectrogram Figure 1 a spectrogram showing all Opus bandwidths can be seen in Figure 2. As can be seen depending on the requested bitrate, the bandwidth can vary from *narrowband* to *fullband*. The possible bandwidths are 4 kHz (NB), 6 kHz (MB), 8 kHz (WB), 12 kHz (SWB) and 20 kHz (FB). There are also some aliasing artifacts apparent in the high frequencies above 16 kHz.

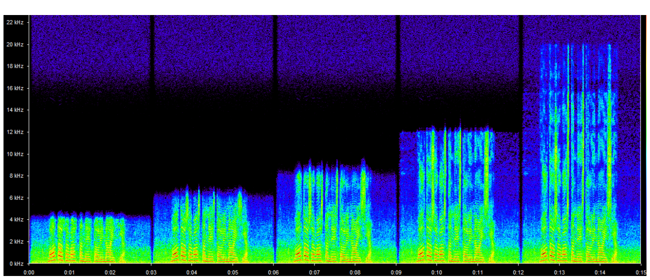

Fig. 2. *Spectrogram of a noisy speech segment with Opus codec at CBR 9.6, 13.2, 16.4, 20 and 24.4 kbit/s.*

2. LISTENING TESTING

A modified version of the ACR[10] mean opinion score (MOS) method was used for the multibandwidth listening test [19]. The MOS scale was extended to be 9 categories wide in order to get more accurate results with relatively high quality and wider than *narrowband* or *wideband* bandwidth speech and audio signals. Only the extreme categories were defined with verbal description: 1 "Very bad" and 9 "Excellent". The assessment is not free sliding, but nine different values still provide the listener more ways to discriminate the samples than five [20]. The listening test procedure and result description is similar to that used for speech codec evaluations in [21], [22] and [23].

2.1. Test conditions

The following test conditions were included in the evaluation:

-Direct reference conditions with limited audio bandwidth but no speech coding. Six lowpass cutoff frequencies were evaluated: 4 kHz, 7 kHz, 8 kHz, 10 kHz, 14 kHz and 20 kHz.

-MNRU reference conditions with artificially added distortion. NB used Q=32 dB and Q=16 dB, WB used Q=35 dB and Q=17 dB both with P.810[24]. FB used Q=31 dB and Q=17 dB with modified MNRU using P.50 shaped noise [25].

-AMR narrowband codec [5] commonly employed in mobile networks. Bitrates evaluated: 4.75, 6.4, 7.95, and 12.2 kbit/s.

-AMR-WB wideband codec [6], supported in an increasing number of mobile networks [26]. Bitrates evaluated: 6.6, 8.85, 12.65, and 23.85 kbit/s.

-EVS latest 3GPP voice and audio codec[1]. 28 operation points were tested with NB, WB, SWB and FB bandwidths 5.9- 128 kbit/s. All conditions can be seen in Figure 7.

-Opus[18], an open source codec. Eight variable bitrates (VBR) were evaluated: 5.9, 7.2, 8.0 and 9.6 kbit/s (NB, 4 kHz), 13.2 kbit/s (WB, 8 kHz), 16.4 kbit/s (SWB, 12 kHz), 24.4 and 32 kbit/s (FB, 20 kHz). Eight hard constant bitrates (CBR) were evaluated: 13.2 kbit/s (MB, 6 kHz), 16.4 kbit/s (WB, 8 kHz), 20 kbit/s (SWB, 12 kHz), 24.4, 32, 40, 48 and 64 kbit/s (FB, 20 kHz).

-ITU-T G.722.1 Annex C[7], a low-complexity superwideband voice codec with an audio bandwidth of 14 kHz. Three bit rate modes were evaluated: 24 kbit/s, 32 kbit/s and 48 kbit/s .

-ITU-T G.718 Annex B[8][27], an embedded (8–64 kbit/s) speech codec for narrowband, wideband, and superwideband services. 28 kbit/s with 14 kHz audio bandwidth was evaluated.

-ITU-T G.719[9], realtime fullband voice and audio codec. Four bitrates were evaluated: 32, 48, 64 and 96 kbit/s.

2.2. Listening tests

Three listening tests were organized:

-Clean speech 4 talkers (2 females, 2 males), sentence pairs of about 6 seconds.

-Noisy speech 4 talkers (2 females, 2 males), sentence pairs of 7 seconds. Street and car noise with signal-to-noise ratio (SNR) of 15 dB for the first pair of speakers, cafeteria and office noise (SNR 20 dB) for the second pair of speakers

-Mixed content, 4 signal types (4 excerpts for each category), Artificial and real mixed signal (each sample contains both speech with music intermixed) as well as two music only conditions (classical and modern music), duration of 8 seconds.

The tests took place in sound-proof booths in the listening test laboratory of Nokia Technologies [28]. Subjects listened to samples diotically through Sennheiser HD-650 headphones. Twenty-four native Finnish listeners participated in each test. In all the tests, 3–5 of the subjects were expert listeners (35–47 years of age). The rest of the participants were naive listeners.

3. RESULTS

Clean speech results in Figure 3 show that EVS is significantly better than either AMR or AMR-WB at all operation points. Also in SWB and FB conditions EVS is significantly better than either G.722.1C or G.719 at bitrates below 64 kbit/s, where the quality saturates near to the direct. Comparison to Opus indicates at bitrates below 24.4 kbit/s; EVS is significantly better than Opus. At 32 kbit/ Opus VBR shows very good performance, but due to variable bitrate operation, the actual bitrate may be as high as 64 kbit/s as can be seen from Table 1, this naturally helps the subjective quality. Notably EVS-SWB 9.6 kbit/s is better than AMR-WB 23.85 kbit/s or Opus 20 kbit/s providing better voice quality at less than half the bitrate.

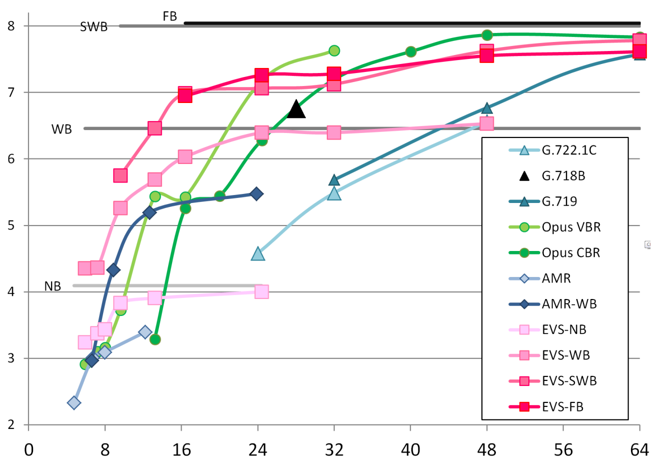

Fig. 3. *Clean speech MOS scores with increasing bitrate (in kbit/s)*

Noisy speech results in Figure 4 are very similar to the clean speech results, with the exception that G.722.1C and G.719 perform somewhat better in noisy speech. This phenomenon is well known from previous listening testing [22] [23]. Opus VBR has some quality issues at 16.4 kbit/s, where subjective quality degrades compared to 13.2 kbit/s.

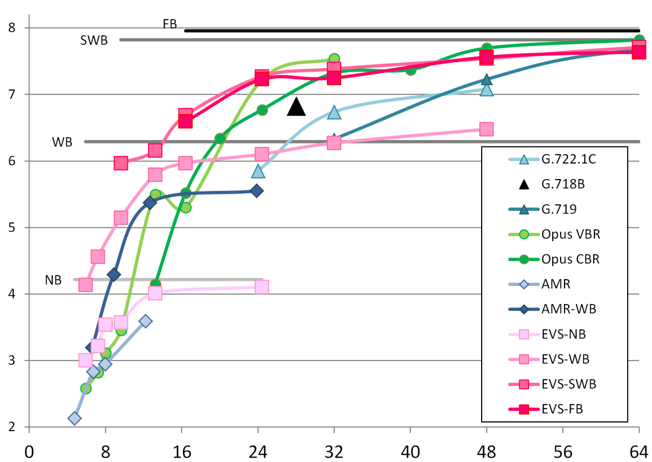

Fig. 4. *Noisy speech MOS scores with increasing bitrate (in kbit/s)*

Mixed content results in Figure 5 are quite similar to the noisy speech results, but now the quality scaling problem with increasing bitrate with Opus is visible in both VBR and CBR operation at 16.4 kbit/s and 20 kbit/s respectively. The performance drop most likely happens due to too fast bandwidth increase with bitrate. These bitrates are the only ones using Opus SWB (12 kHz) mode, so it could also be related to some quality issue in the actual SWB coding algorithm. With mixed speech and music signal EVS shows excellent performance compared to Opus and older standardized codecs. At each bitrate 5.9- 24.4 kbit/s EVS is upto 2 MOS scores better than any other codec. The reason for good performance at low bitrates is likely due to the seamless, frame-by-frame switching between ACELP and MDCT based coding cores of the EVS [29][30].

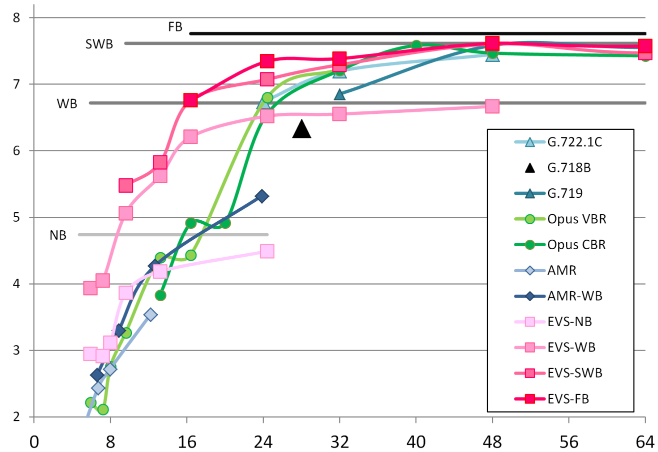

Fig. 5. *Mixed content MOS scores with increasing bitrate (in kbit/s)*

Finally all three listening test results were combined and a single overall results Figure 6 containing all 72 listeners was generated. Some additional results for higher bitrates 96 and 128 kbit/s are visible in the higher bitrates. Combined results together with confidence intervals are shown in block diagram form in Figure 7.

From the overall results in Figures 6 and 7 it can be seen that EVS is better than or equivalent to Opus CBR or VBR at all bitrates. EVS-NB is statistically equivalent to NB direct at 24.4 kbit/s. EVS-WB is statistically equivalent to wideband direct at 32 kbit/s. Also EVS-SWB and EVS-FB reach statistical equivalence to direct SWB and FB at 64 and 128 kbit/s, respectively.

4. CONCLUSIONS

A subjective quality evaluation was conducted with three listening tests in Nokia Technologies listening facilities. From the results it can be seen that the 3GPP EVS codec produces state-of-the-art voice and audio quality across all tested bitrates and bandwidths. Compared to Opus and other standardized codecs, EVS provides the same quality at about half the bitrate in low bitrates. For example EVS-SWB 9.6 kbit/s is significantly better than Opus CBR 20 kbit/s and EVS-SWB 16.4 kbit/s provides the same overall quality as G.722.1C 32 kbit/s. If we consider that EVS-SWB 13.2 kbit/s provides MOS 6.15, AMR-WB 12.65 kbit/s MOS 4.95 and AMR 12.2 kbit/s MOS 3.51, the improvement from WB to SWB (1.2 points) is almost as large as the improvement was from NB to WB (1.44 points). Further, if we consider a slightly higher operation point, EVS-SWB 24.4 kbit/s versus AMR-WB 23.85 kbit/s, the improvement is almost 1.7 MOS points and thus a more notable than NB to WB improvement was in the previous voice codec generation upgrade.

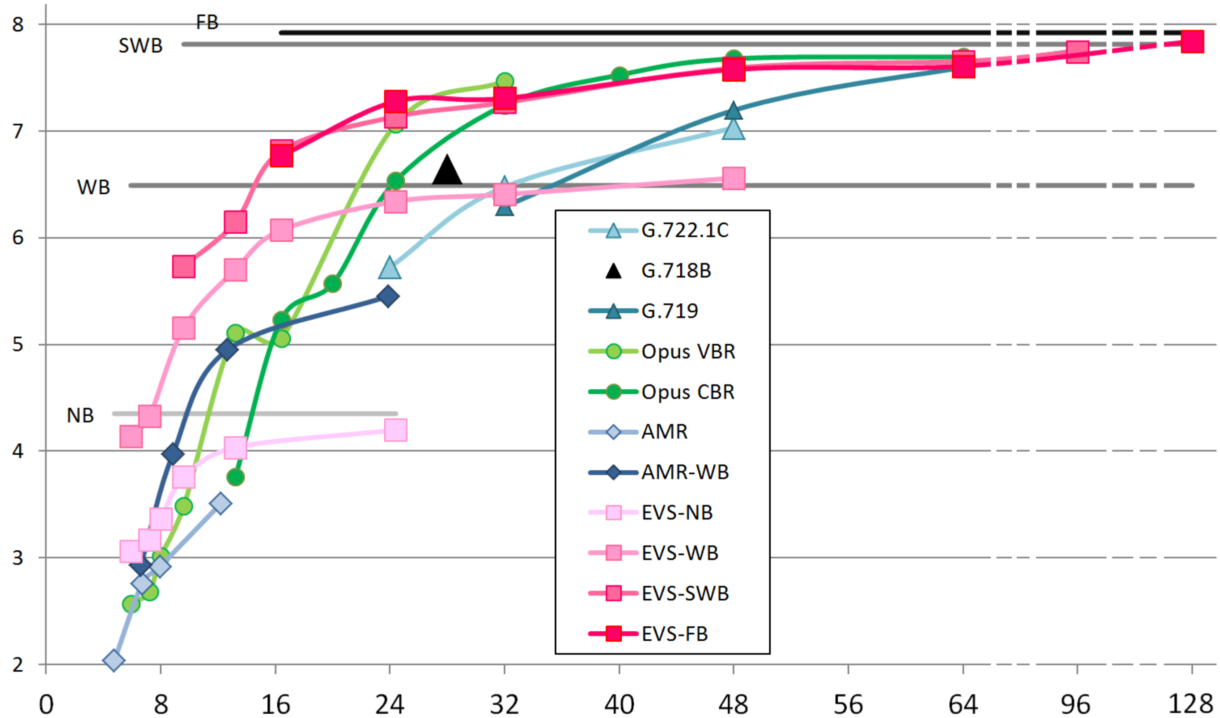

Fig. 6. *Combined results with all 72 listeners and all signal types with increasing bitrate (in kbit/s). Note that the two highest bitrates of 96 and 128 kbit/s are not in linear bitrate scale.*

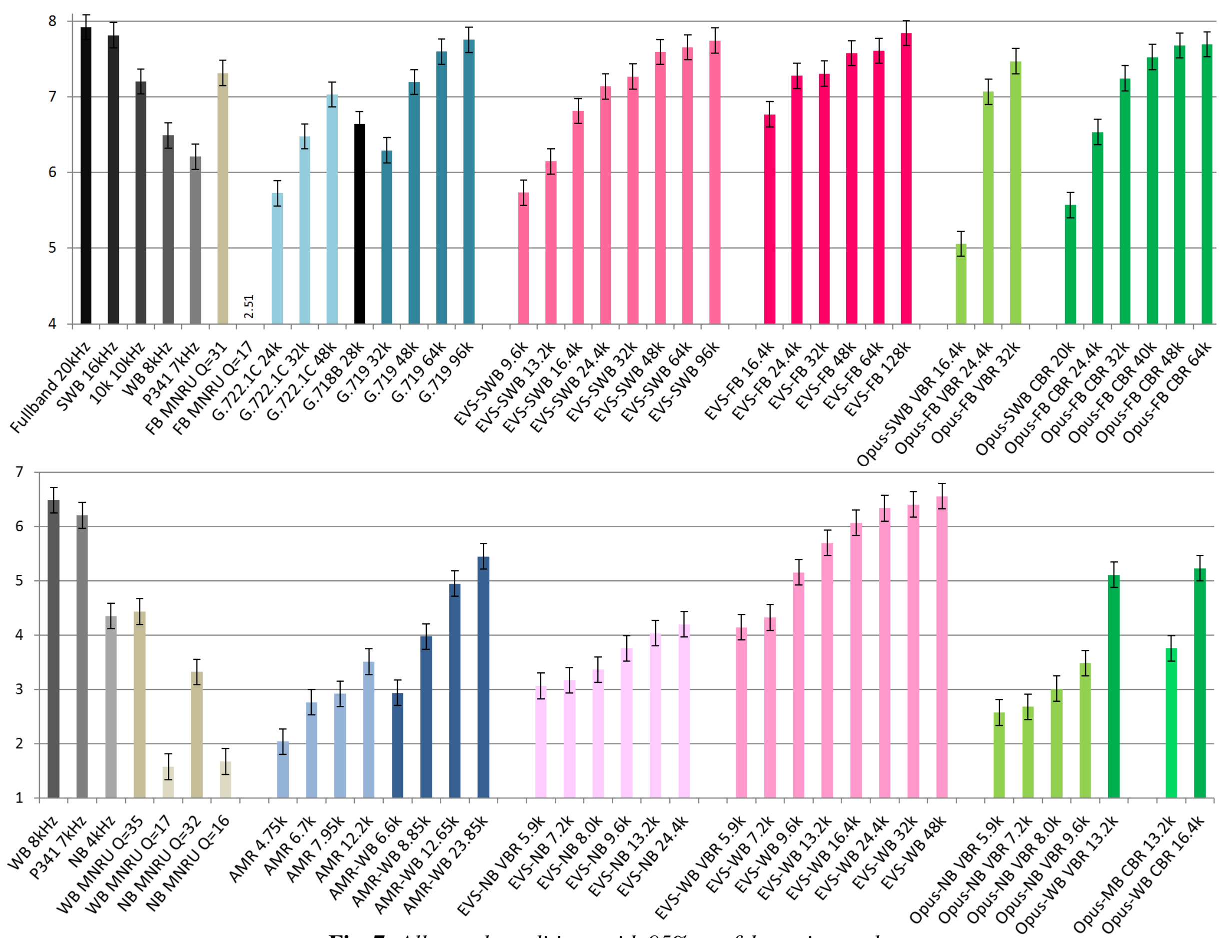

Fig. 7. *All tested conditions with 95% confidence intervals.*

5. REFERENCES

[1] Stefan Bruhn, Harald Ploboth, Markus Schnell, Bernhard Grill, Jon Gibbs, Lei Miao, Kari Järvinen, Lasse Laaksonen, Noboru Harada, Nobuhiko Naka, Stefane Ragot, Stefane Proust, Takako Sanda, Imre Varga, Craig Greer, Milan Jelinek, and Minjie Xie Paolo Usai, "Standardization of the new 3GPP EVS codec," in *Proc. ICASSP*, Brisbane, Australia, Apr. 2015.

[2] Kari Järvinen, Imed Bouazizi, Lasse Laaksonen, Pasi Ojala, and Anssi Rämö, "Media coding for the next generation mobile system LTE," *Elsevier Computer Communications*, vol. 33, no. 16, pp. 1916–1927, Oct. 2010.

[3] 3GPP Tdoc S4-130522, *EVS Permanent Document (EVS-3): EVS performance requirements, Version 1.4*, 3GPP, Apr. 2013, online: http://www.3gpp.org/ftp/tsg_sa/WG4_CODEC/TSGS4_73/Docs/S4-130522.zip.

[4] 3GPP Tdoc S4-141065, *Report of the Global Analysis Lab for the EVS Selection Phase*, 3GPP, Aug. 2014, online: http://www.3gpp.org/ftp/tsg_sa/WG4_CODEC/TSGS4_80bis/Docs/S4-141065.zip.

[5] 3GPP TS 26.090, *Adaptive multi-rate (AMR) speech codec; Transcoding functions*, 3GPP, Sept. 2012.

[6] 3GPP TS 26.190, *Adaptive multi-rate wideband (AMR-WB) speech codec; Transcoding functions*, 3GPP, Sept. 2012.

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

[8] 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 superwideband scalable extension for ITU-T G.718 and corrections to main body fixed-point C-code and description text*, ITU, Mar. 2010.

[9] ITU-T G.719, *Low-complexity, full-band audio coding for high-quality, conversational applications*, ITU, June 2008.

[10] ITU-T P.800, *Methods for subjective determination of transmission quality*, ITU, Aug. 1996, online: http://www.itu.int/rec/T-REC-P.800-199608-I/en.

[11] Anssi Rämö, "Voice quality evaluation of various codecs," in *Proc. ICASSP*, Dallas, TX, USA, Mar. 2010, pp. 4662–4665.

[12] Jeremie Lecomte, Tommy Vaillancourt, Stefan Bruhn, Hosang Sung, Ke Peng, Kei Kikuiri, Bin Wang, Shaminda Subasingha, and Julien Faure, "Packet loss concealment technology advances in EVS," in *Proc. ICASSP*, Brisbane, Australia, Apr. 2015.

[13] Venkatraman Atti, Daniel J. Sinder, Shaminda Subasingha, Vivek Rajendran, Duminda Dewasurendra, Venkata Chebiyyam, Imre Varga, Venkatesh Krishnan, Benjamin Schubert, Jeremie Lecomte, Xingtao Zhang, and Lei Miao, "Improved error resilience for VOLTE and VOIP with 3GPP EVS channel aware coding," in *Proc. ICASSP*, Brisbane, Australia, Apr. 2015.

[14] 3GPP TS 26.441, *EVS Codec General Overview*, 3GPP, Aug. 2014, online: http://www.3gpp.org/DynaReport/26441.htm.

[15] 3GPP TS 26.445, *Codec for Enhanced Voice Services (EVS); Detailed algorithmic description*, 3GPP, Sept. 2014, online: http://www.3gpp.org/DynaReport/26445.htm.

[16] Martin Dietz, Markus Multrus, Vaclav Eksler, Vladimir Malenovsky, Erik Norvell, Harald Ploboth, Lei Miao, Zhe Wang, Lasse Laaksonen, Adrian Vasilache, Yutaka Kamamoto, Kei Kikuiri, Stephane Ragot, Hiroyuki Ehara, Vivek Rajendran, Venkatraman Atti, Hosang Sung, Eunmi Oh, Hao Yuan, and Changbao Zhu, "Overview of the EVS codec architecture," in *Proc. ICASSP*, Brisbane, Australia, Apr. 2015.

[17] 3GPP TR 26.952, *Performance Characterization of the EVS codec*, 3GPP, Dec. 2014, online: http://www.3gpp.org/DynaReport/26952.htm.

[18] Jean-Marc Valin, Koen Vos, and Timothy B. Terriberry, "Definition of the Opus audio codec," IETF RFC 6716, Sept. 2012, online: http://tools.ietf.org/html/rfc6716.

[19] Anssi Rämö and Henri Toukomaa, "On comparing speech quality of various narrow- and wideband speech codecs," Sydney, Australia, Aug. 2005, pp. 603–606.

[20] Kerrie Lee, Phillip Dermody, and Daniel Woo, "Evaluation of a method for subjective assessment of speech quality in telecommunication applications," in *ASTA*, Apr. 1996, pp. 199–203, online http://www.assta.org/?q=sst-1996-proceedings.

[21] Anssi Rämö and Henri Toukomaa, "Voice quality evaluation of recent open source codecs," in *Proc. Interspeech*, Tokyo, Japan, Sept. 2010, pp. 2390–2393.

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

[23] Hannu Pulakka, Anssi Rämö, Ville Myllylä, Henri Toukomaa, and Paavo 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.

[24] ITU-T P.810, *Telephone transmission quality: methods for objective and subjective assessment of quality. Modulated noise reference unit (MNRU)*, ITU, Feb. 1996.

[25] ITU-T P.50, *Telephone transmission quality, telephone installations, local line networks: Objective measuring apparatus*, ITU, Sept. 1999, online: http://www.itu.int/rec/T-REC-P.50-199909-I.

[26] Global mobile suppliers association (GSA), "Mobile HD voice: Global update report," Sept. 2014.

[27] Lasse Laaksonen, Mikko Tammi, Vladimir Malenovsky, Tommy Vaillancourt, Mi Suk Lee, Tomofumi Yamanashi, Masahiro Oshikiri, Claude Lamblin, Balazs Kovesi, Lei Miao, Deming Zhang, Jon Gibbs, and Holly Francois, "Superwideband extension of G.718 and G.729.1 speech codecs," in *Proc. Interspeech*, Tokyo, Japan, Sept. 2010, pp. 2382–2385.

[28] Mikko Kylliäinen, Heikki Helimäki, Nick Zacharov, and John Cozens, "Compact high performance listening spaces," in *Proc. Euronoise*, Naples, Italy, May 2003.

[29] Tommy Vaillancourt, Wang Zhe, Kihyun Choo, and Venkatraman Atti, "Two-stage speech/music classifier with decision smoothing and sharpening in the EVS codec," in *Proc. ICASSP*, Brisbane, Australia, Apr. 2015.

[30] Emmanuel Ravelli, Christian R. Helmrich, Guillaume Fuchs, , and Markus Multrus, "Open loop switching decision based on evaluation of coding distortions for audio codecs," in *Proc. ICASSP*, Brisbane, Australia, Apr. 2015.