INTERSPEECH 2016
September 8–12, 2016, San Francisco, USA

EVS Channel Aware Mode Robustness to Frame Erasures

Anssi Rämö[1], *Antti Kurittu*[2], *Henri Toukomaa*[1]

[1]Nokia Technologies
[2]Nokia Networks

anssi.ramo@nokia.com, antti.kurittu@nokia.com

Abstract

This paper discusses the voice and audio quality characteristics of the EVS, the recently standardized 3GPP Enhanced Voice Services codec. Especially frame erasure conditions with and without channel aware mode were evaluated. The test consisted of two extended range MOS listening tests. The tests contained both clean and noisy speech in clean channel as well as with four frame erasure rates (5%, 10%, 20% and 30%) for selected codecs and bitrates. In addition to subjective test results some additional objective results are presented. The results show that EVS channel aware mode performs better than EVS native mode in high FER rates. For comparison also AMR, AMR-WB and Opus codecs were included to the listening tests.

Index Terms: speech coding, listening testing, voice quality evaluation

1. Introduction to EVS

The 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. The 5.9 kbit/s mode uses VBR (Variable BitRate) with discontinuous transmission (DTX) always enabled [1]. All other bitrates are CBR (Constant BitRate) [2]. Audio and speech coding modes are switched internally in realtime by the EVS codec depending on the input signal characteristics. The EVS codec is designed to be inherently robust to channel errors [3]. Further information about the EVS codec can be found from specification, and ICASSP and GlobalSIP special session papers [4] [5] [6] [7].

There are several different coding modes that help with robustness to frame erasures. There is for example a specific transition mode that codes speech and audio onsets so that individual lost frames do not adversely affect audio quality [8]. For tonal music and stable speech vowels there are separate coding modes that handle these type of stable segments very well in case of frame loss [9]. Global gains are used independently from the previous frames in order to reduce error propagation [10]. LSF spectral coefficients are coded with a vector quantizer that takes into account possible lost frames by inserting non-predictive frames at most critical places [11][12]. Additional information about native EVS packet loss concealment (PLC) can be found in [13] and the EVS standard [14].

2. EVS Channel Aware Mode

In addition to the EVS native mode, EVS contains a specific Channel Aware (CA) mode [15][16] (Chapter 5). The CA mode is meant to be used in IP based best effort networks such as VoIP, VoWifi and VoLTE, where speech frames could be delayed or be lost totally in transmission. Additional error resilience is achieved using a form of in-band Forward Error Correction (FEC). In case of delayed or lost packet EVS-CA mode uses partially coded frames embedded to the EVS bitstream to conceal the lost frames with less artifacts. This requires a novel usage of jitter buffer [17]. Naturally partial redundant frames reduce the full rate bitrate somewhat. Depending on the criticality of the frame, the partial redundancy is dynamically enabled or disabled for a particular frame, while keeping a fixed bit budget of 13.2 kbit/s. The amount of redundant information varies frame by frame and ranges from 0 to 3.6 kbit/s. Source-controlled coding techniques are used to identify candidate speech frames for bitrate reduction, leaving spare bits for transmission of partial copies of prior frames such that a constant bit rate is maintained and very little clean channel quality is lost. This CA functionality is specified only for 13.2 kbit/s *wideband* (WB) and *superwideband* (SWB) modes with or without DTX. All other bitrates and signal bandwidths use standard PLC defined in [14].

2.1. CA Mode Details

The difference in time units (in EVS case the frame length is 20 ms) between the transmit time of the primary copy of a frame and the transmit time of the redundant copy of the frame (piggy backed onto a future frame) is called the FEC offset. If the depth of the jitter buffer at any given time is at least equal to the FEC offset, then it is possible that the future frame is available in the de-jitter buffer at the current time instance. The FEC offset is a configurable parameter at the encoder and it can be dynamically adjusted depending on the network conditions and it is sent in the bitstream with 2 bits. Allowed values are 2, 3, 5 or 7. In general the FEC offset should be longer for a very erroneous channel and shorter for a good channel.

In addition there is a frame erasure rate indicator in the encoder having the following values: *low* for FER rates less than 5% or *high* for FER higher than 5%. The *high* setting adjusts the criticality threshold to classify more frames as critical to transmit redundant information as compared to the *low* setting.

The defaults for the encoder are *high* and FEC offset = 3. Redundant frame type is signaled with 3 bits (NO_DATA, TCXFD, TCXTD1, TCXTD2, ALLPRED, NOPRED, GENPRED, and NELP) and is included in the bit stream for every redundant copy containing frame.

3. Description of the Listening Test

In addition to the EVS native and CA mode the reference codecs for the listening test were AMR [18][19], AMR-WB [20][21] and the Opus codec [22][23]. All codecs were tested at around 12.2-13.2 kbit/s bitrates. In addition EVS-NB was included at 8 kbit/s as well as EVS-FB and Opus at 24.4 kbit/s. Full list of tested codecs and frame error rates can be seen in Table 3.

http://dx.doi.org/10.21437/Interspeech.2016-917

Latest version of each codec as of February 2016 was used for test processing.

3.1. Extended Range ACR 9-Scale Method

A modified version of the ACR mean opinion score (MOS) method [24] was used for the multi-bandwidth listening test [25]. 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 [26]. 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. For example using a seven scale ACR was found out to give more accurate results than five scale assessment in an independent study [27]. By coincidence test results with *narrowband* characteristics often hit the traditional MOS range of 1-5[28][29]. The listening test procedure and result description is similar to that used for speech codec evaluations in [30], and [31]. Objective POLQA and WB-PESQ results correlate quite well with 9-scale subjective MOS scores [32].

3.2. Listeners and Samples

There were a total of 24 listeners in both listening tests. All of them were naive and most younger than 20 years old. Each listener listened to all conditions with eight different samples from eight different voice sample categories. The sample categories are described in Table 1. Each condition received a total of 192 votes. Samples were listened with Sennheiser HD-650 headphones. Diotic listening was conducted for improved accuracy. Separate isolated listening booths were used for listening [33].

Test	Environment	Speaker	Background noise level
1	Clean speech	Male 1	na
1	Clean speech	Female 1	na
1	Clean speech	Male 2	na
1	Clean speech	Female 2	na
2	Speech in street noise	Male 3	-15 dB
2	Speech in cafeteria noise	Female 3	-15 dB
2	Speech with classical music	Male 4	-15 dB
2	Speech in car noise	Female 4	-15 dB

Table 1: Sample types used for ACR9 listening testing

The original voice samples were recorded at 48kHz in a quiet studio environment. Office, car and street noises and background music level was set to -15 dBOv. The music sample originates from a regular CD and it was upsampled to 48kHz with a high quality resampling program [34]. All sample sequences were six to seven seconds long containing a sentence pair. The full list of tested sample types can be seen in Table 1.

3.3. Test conditions

In addition to the EVS-CA mode, EVS native mode, Opus and a selection of older 3GPP coded samples were included to the test for comparison reasons (Table 3). Additional references were bandwidth limited signals, created using ITU-Tools [35], as well as MNRU noise worsened samples [36][37]. The reference signal list can be seen in Table 2.

Reference	Bandwidth	Notes
Direct FB	20 kHz	*fullband* bandwidth
10 kHz limited	10 kHz	10 kHz limited signal
WB	8 kHz	*wideband* bandwidth
NB	4 kHz	*narrowband* bandwidth
NB MNRU 16 dB	4 kHz	MNRU noise [36]
WB MNRU 18 dB	8 kHz	MNRU noise
FB MNRU 16 dB	20 kHz	P.50 MNRU noise [37]
FB MNRU 24 dB	20 kHz	P.50 MNRU noise

Table 2: Tested reference conditions

Codec	Bandwidth	Bitrate	FER conditions
AMR	NB	12.2	0, 5, 10, 20, and 30%
AMR-WB	WB	12.65	0 - 30%
AMR-WB	WB	23.85	Only 0%
EVS	NB	8.0	0 - 30%
Opus CBR	MB*	13.2	0 - 30%
EVS	WB	13.2	0 - 30%
EVS-CA mode	WB	13.2	0 - 30%
EVS	WB	24.4	Only 0%
EVS	SWB	13.2	0 - 30%
EVS-CA mode	SWB	13.2	0 - 30%
EVS	FB	24.4	0 - 30%
Opus CBR	FB	24.4	0 - 30%

Table 3: Tested codecs and frame erasure rates. MB* *medium-band* and is approximately 6 kHz bandwidth limited.

3.4. Channel aware mode configuration

According to the EVS specification [16], the CA mode configuration should be adapted to the channel conditions. Thus we tested the EVS-CA mode so that in lower frame erasure rates, the redundant frame (RF) indicator was set to *low* and FEC offset was set to values 2,3 or 5. With increasing frame erasure rate the FEC offset was increased (5 and 7) and for FER rates 20 % and 30 % the RF indicator was set to *high*. Full set of configurations can be seen in Table 4.

Frame erasure rate	RF indicator	FEC offset
0 %	*low*	2
5 %	*low*	3
10 %	*low*	5
20 %	*high*	5
30 %	*high*	7

Table 4: EVS-CA mode tested configurations

4. Listening test results

Codec and or reference conditions are collected to a X-Y line graph, where bullets point to individual MOS results and a line connects the bullets, when FER scalability can be seen (e.g. Figure 1). On the left y-axis of the table the ACR9 MOS scale is shown. On the bottom x-axis the FER rate is shown. All results are represented on a linear scale. The colors and markers of individual codecs are kept consistent over different graphs. Traditional bar graphs with confidence intervals are also presented for overall results in Figure 3. The 95% confidence interval in this listening test was around 0.2. Thus any difference larger than 0.2 can be considered statistically significant.

Results in Figure 1 show that the CA mode works as expected. In clean channel or at low FER rate EVS-CA is statistically equivalent to EVS native mode both in *wideband* and *superwideband*. Especially at high FER rates of over 10 % the CA mode improves subjective voice quality significantly. Notably, the AMR-WB codec performs quite poorly with increasing FER rate, although it shows good performance in clean channel. Opus 13.2 kbit/s is significantly worse than AMR-WB 12.65 kbit/s mainly due to reduced bandwidth (6 kHz compared to over 7 kHz of the AMR-WB) in clean channel. At FER rates higher than 10 % Opus is better than AMR-WB 12.65 kbit/s. However, even native mode EVS-WB is significantly better in all different FER rates than Opus at the same bitrate. EVS-CA mode is more than 1.1 MOS points better than Opus at all FER rates.

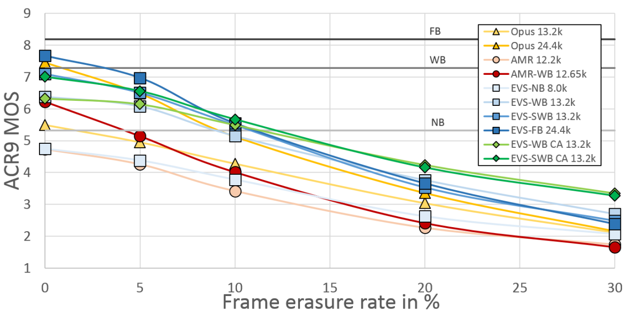

Figure 1: Voice quality with increasing frame erasure rate

EVS-NB 8.0 kbit/s is better than AMR 12.2 kbit/s at every tested FER rate. Although at low FER rates the difference is not significant.

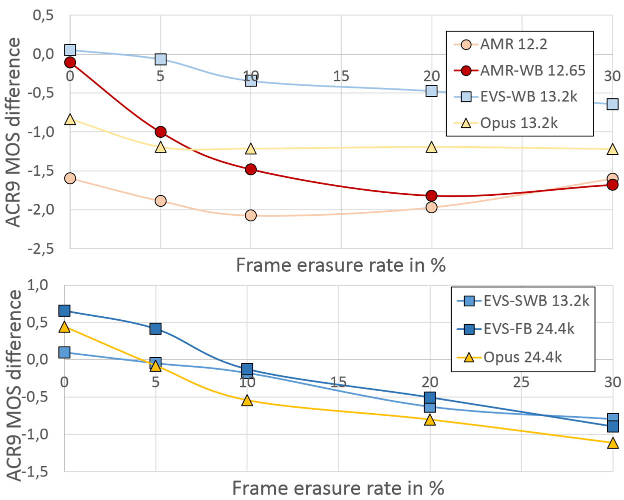

Figure 2: Difference of selected NB/WB codecs to EVS-WB CA and SWB/FB codecs to EVS-SWB CA

EVS-FB at 24.4 kbit/s and Opus 24.4 kbit/s perform almost identically in clean channel conditions and at all FER rates. The performance of Opus at 24.4 kbit/s is not surprise, since we have tested it earlier [30]. The similar FER robustness shows that both Opus and EVS-FB 24.4 kbit/s have something to improve since even EVS-WB 13.2 kbit/s is working better than either of these *fullband* codecs in 20 % and 30 % FER. The reason for this relatively poor performance is likely the artifacts caused by the on/off-nature of the high frequency excitation, which sounds a bit disturbing. The more stable narrower bandwidth probably would benefit the overall voice quality in noisy channel for both *fullband* codecs.

When we subtract selected codec results from the EVS-CA mode results we get Figure 2, where NB/WB conditions and SWB/FB conditions are shown in upper and lower figures respectively. As can be seen in both figures, the EVS-CA mode is statistically significantly better than any of the reference codecs in NB/WB at FER rates starting at 10 % and in SWB/FB starting at around 15 %.

5. Objective test results

In addition to the subjective listening test, objective measurements were conducted with P.863 a.k.a POLQA [38]. Recommendation ITU-T P.863 describes an objective method for predicting overall listening speech quality from *narrowband* to *superwideband* telecommunication scenarios as perceived by the user in an absolute category rating (ACR) listening-only test [39]. The *superwideband* mode of POLQA was used, providing predicted scores on a MOS ACR *superwideband* scale for multi-bandwidth scenarios. The maximum objective MOS-LQO$_{swb}$ score is 4.75. POLQA enables fast processing of large amount of files, making it possible to assess the performance of codecs in more conditions than possible in a subjective test.

5.1. Comparison between objective POLQA and subjective listening test results

The same speech samples that were used in the listening test were processed with the objective processing device POLQA. The *fullband* conditions were omitted as POLQA does not support it. It can be seen in Table 5 that POLQA ranks the 0 % FER conditions almost in the same order as the listeners in the performed listening test. There are some differences however. For example POLQA ranks AMR-WB 23.85 kbit/s as better than EVS-SWB CA 13.2 kbit/s, whereas in the subjective test it is vice versa. Also EVS-NB 8.0 kbit/s is significantly worse than AMR-NB 12.2 kbit/s in the objective measurement, whereas in the listening test they were scored equally.

Condition	MOS	MOS-LQO$_{swb}$
10kHz limited	7.90	4.75
WB 8kHz	7.28	4.74
EVS-WB 24.4k	7.17	4.70
EVS-SWB 13.2k	7.10	4.51
EVS-SWB CA 13.2k	7.01	4.28
AMR-WB 23.85k	6.40	4.43
EVS-WB 13.2k	6.38	4.31
EVS-WB CA 13.2k	6.33	4.16
AMR-WB 12.65k	6.22	4.26
Opus 13.2k	5.49	4.18
NB 4kHz	5.32	3.84
EVS-NB 8.0k	4.74	3.27
AMR 12.2k	4.73	3.61

Table 5: Subjective MOS vs. objective MOS-LQO$_{swb}$ at 0 % FER

Moreover, it seems that POLQA is not able to predict differences in audio bandwidth as in the subjective test. POLQA scores the WB and 10 kHz limited reference conditions equally, while there is a clear difference in subjective scores. POLQA also scores most of the *wideband* and *superwideband* conditions similarly, while there is a difference of over 0.7 MOS in the subjective test. The Opus 13.2 kbit/s *mediumband* codec is scored equally to EVS-WB CA 13.2 kbit/s, while there is a difference of 0.8 MOS in the subjective score.

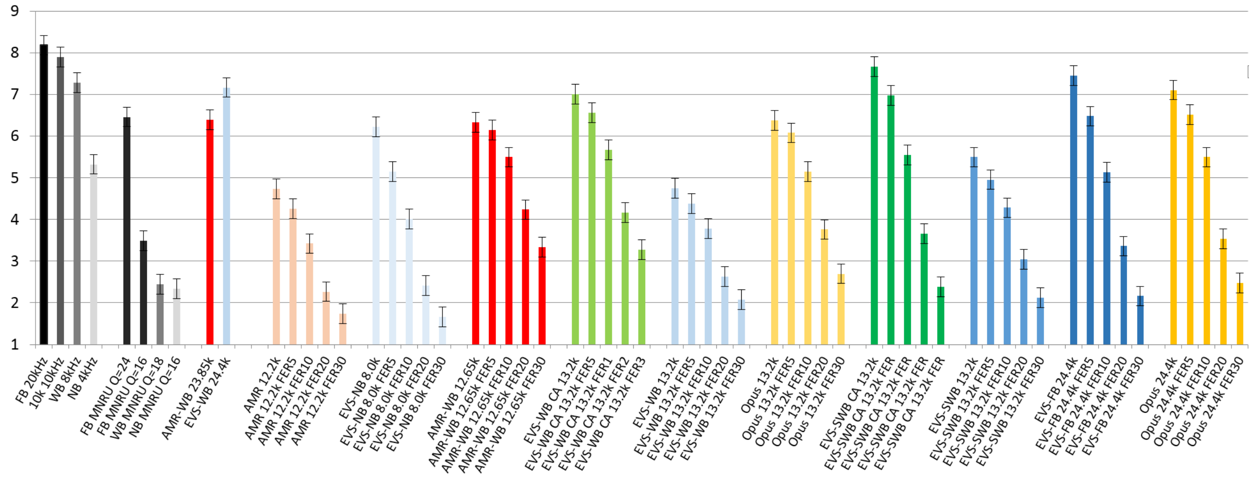

Figure 3: *All tested conditions with 95% confidence intervals.*

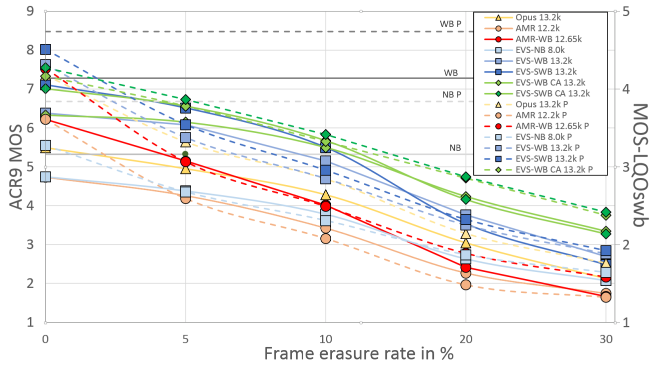

Figure 4: Comparison of ACR9 and objective MOSLQO$_{SWB}$

Figure 4 shows the objective and subjective MOS scores for the same conditions as in the subjective test in Figure 1. The EVS channel aware mode's high quality at higher FER (>5 %) ratios is apparent. One slight difference is that POLQA actually evaluates the quality degradation between clean channel and 5 % FER channel larger than it actually is in the subjective test.

5.2. POLQA in bursty FER

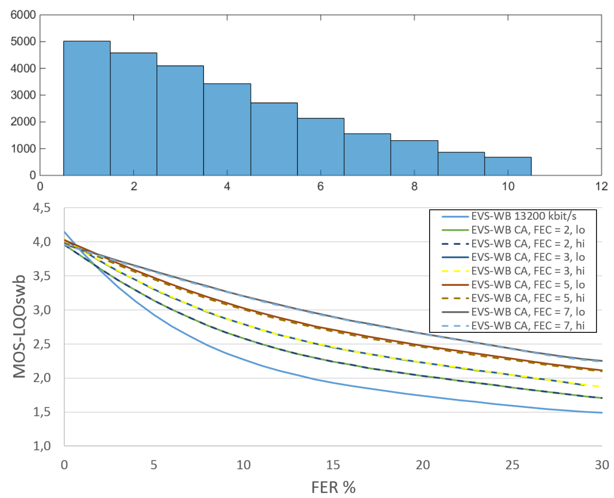

Figure 5: Burst length distribution in FER profiles. Effect of FEC offset and FER indicator setting on voice quality

Heavily bursty FER profiles were generated from 0 % to 30 % with statistics shown in Figure 5. Figure 5 illustrates how the performance changes with different EVS-CA mode parameters in bursty FER scenarios. The assumption was that with increasing FEC offset the performance should be better in high FER scenarios. It was also expected that the *high* FER indicator would be better than the *low* in high FER. The results show that with higher FER, increasing the FEC offset helps the voice quality. The downside of increasing FEC offset is increasing delay in decoding. The FER indicator setting had little effect on the results.

6. Conclusion

The EVS Channel Aware mode provides a nice boost in quality with high FER ratios. Currently the EVS-CA mode supports only 13.2 kbit/s bitrate with either *wideband* or *superwideband* bandwidth (*narrowband* or *fullband* are not supported). It also requires that the packet transmission network together with jitter buffer on the decoder side is used. This means that the CA mode benefits cannot always be realized, but still 13.2 kbit/s SWB will likely be the most used mode in 3GPP VoLTE networks. Comparison to older generation 3GPP codecs or Opus at the same bitrate shows that there are significant improvements in both subjective and objective voice quality at all FER rates.

Finally Figure 3 shows subjective listening test results for all conditions in bar format with confidence intervals, allowing for a more detailed comparison of codecs and FER conditions. An especially interesting fact is that reasonable communications quality (MOS>3 similar to e.g. AMR 12.2 kbit/s in 10% FER) is achieved with EVS-CA mode with 30% FER allowing communications even when almost one third of the frames are lost or delayed in the network.

The objective tests showed that with bursty FER, increasing the FEC offset helps the voice quality. The tests also revealed that the FER indicator setting *low* / *high* has very little effect on voice quality. The objective and subjective results correlate quite nicely as can be seen in figure 4, although POLQA had some trouble consistently predicting the quality of speech between different bandwidths. However, objective scores depend heavily on the used speech samples so the result may have been different with a different sample set.

7. References

[1] K. Järvinen, A. Rämö, J. Hagqvist, O. Kirla, A. Vasilache, and L. Laaksonen, "Nokia networks white paper: The 3GPP Enhanced Voice Services (EVS) codec," http://networks.nokia.com/file/41551/the-3gpp-enhanced-voice-services-evs-codec, June 2015.

[2] 3GPP TS 26.450, "Codec for enhanced voice services (evs); discontinuous transmission (dtx)," 2015.

[3] J. Lecomte, T. Vaillancourt, S. Bruhn, H. Sung, K. Peng, K. Kikuiri, B. Wang, S. Subasingha, and J. Faure, "Packet loss concealment technology advances in EVS," in *Proc. ICASSP*, Brisbane, Australia, April 2015.

[4] 3GPP TS 26.441, "Codec for Enhanced Voice Services (EVS); General overview," 2015.

[5] S. Bruhn, H. Ploboth, M. Schnell, B. Grill, J. Gibbs, L. Miao, K. Järvinen, L. Laaksonen, N. Harada, N. Naka, S. Ragot, S. Proust, T. Sanda, I. Varga, C. Greer, M. Jelinek, M. Xie, and P. Usai, "Standardization of the new 3GPP EVS codec," in *Proc. ICASSP*, Brisbane, Australia, April 2015.

[6] M. Dietz, M. Multrus, V. Eksler, V. Malenovsky, E. Norvell, H. Ploboth, L. Miao, Z. Wang, L. Laaksonen, A. Vasilache, Y. Kamamoto, K. Kikuiri, S. Ragot, H. Ehara, V. Rajendran, V. Atti, H. Sung, E. Oh, H. Yuan, and C. Zhu, "Overview of the EVS codec architecture," in *Proc. ICASSP*, Brisbane, Australia, April 2015.

[7] S. Bruhn, T. Frankkila, F. Gabin, K. Hellwig, and M. Hultström, "System aspects of the 3GPP evolution towards enhanced voice services," in *Global-SIP 2015*, Orlando, Florida, USA, December 2015.

[8] V. Eksler, M. Jelnek, and R. Salami, "Efficient handling of mode switching and speech transitions in the EVS codec," in *Proc. ICASSP*, Brisbane, Australia, April 2015.

[9] R. Sperschneider, J. Sukowski, and G. Markovic, "Delay-less frequency domain packet-loss concealment for tonal audio signals," in *Global-SIP 2015*, Orlando, Florida, USA, December 2015.

[10] V. Malenovsky and M. Jelinek, "Memory-less gain quantization in the evs codec," in *Global-SIP 2015*, Orlando, Florida, USA, December 2015.

[11] A. Vasilache, A. Rämö, H. Sung, S. Kang, J. Kim, and E. Oh, "Flexible spectrum coding in the 3GPP EVS codec," in *Proc. of ICASSP-2015*, Brisbane, Australia, April 2015.

[12] A. Rämö, A. Vasilache, and H. Toukomaa, "Robust speech coding with EVS," in *Proc. of Global-SIP 2015*, Orlando, Florida, USA, December 2015.

[13] J. Lecomte, A. Tomasek, G. Markovic, M. Schnabel, K. Tsutsumi, and K. Kikuiri, "Enhanced time domain packet loss concealment in switched speech/audio codec," in *Proc. ICASSP*, Brisbane, Australia, April 2015.

[14] 3GPP TS 26.447, "Codec for Enhanced Voice Services (EVS); Error concealment of lost packets," 2015.

[15] V. Atti, D. J. Sinder, S. Subasingha, V. Rajendran, D. Dewasurendra, V. Chebiyyam, I. Varga, V. Krishnan, B. Schubert, J. Lecomte, X. Zhang, and L. Miao, "Improved error resilience for VoLTE and VoIP with 3GPP EVS channel aware coding," in *Proc. ICASSP*, Brisbane, Australia, April 2015.

[16] 3GPP TS 26.445, "Codec for Enhanced Voice Services (EVS); Detailed algorithmic description," 2015.

[17] 3GPP TS 26.448, "Codec for enhanced voice services (evs); jitter buffer management," 2015.

[18] 3GPP TS 26.071, "Mandatory speech codec speech processing functions; AMR speech codec; general description," 1999.

[19] E. Ekudden, R. Hagen, I. Johansson, and J. Svedberg, "The amr speech coder," in *IEEE Workshop on speech coding*, Porvoo, Finland, June 1999, pp. 117–119.

[20] 3GPP TS 26.171, "Speech codec speech processing functions; Adaptive Multi-Rate - Wideband (AMR-WB) speech codec; General description," 2001.

[21] B. Bessette, R. Salami, R. Lefebvre, M. Jelinek, J. Rotola-Pukkila, J. Vainio, H. Mikkola, and K. Jrvinen, "The adaptive multi-rate wideband speech codec (AMR-WB)," *IEEE Transactions on Speech and Audio Processing*, vol. 10, no. 8, pp. 620–636, November 2002.

[22] J.-M. Valin, K. Vos, and T. B. Terriberry, "Definition of the Opus audio codec," IETF RFC 6716, September 2012, online: http://tools.ietf.org/html/rfc6716.

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

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

[25] A. Rämö and H. Toukomaa, "On comparing speech quality of various narrow- and wideband speech codecs," in *"Proc. of ISSPA-2005*, Sydney, Australia, August 2005, pp. 603–606.

[26] A. Rämö, "Voice quality evaluation of various codecs," in *Proc. of ICASSP-2010*, Dallas, TX, USA, March 2010, pp. 4662–4665.

[27] K. Lee, P. Dermody, and D. Woo, "Evaluation of a method for subjective assessment of speech quality in telecommunication applications," in *ASTA*, Australia, April 1996, pp. 199–203, online http://www.assta.org/?q=sst-1996-proceedings.

[28] A. Rämö and H. Toukomaa, "Voice quality evaluation of recent open source codecs," in *Proc. of Interspeech-2010*, Tokyo, Japan, September 2010, pp. 2390–2393.

[29] K. Järvinen, I. Bouazizi, L. Laaksonen, P. Ojala, and A. Rämö, "Media coding for the next generation mobile system LTE," *Elsevier Computer Communications*, vol. 33, no. 16, pp. 1916–1927, October 2010.

[30] A. Rämö and H. Toukomaa, "Subjective quality evaluation of the 3GPP EVS codec," in *Proc. of ICASSP-2015*, Brisbane, Australia, April 2015, pp. 5157–5161.

[31] 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. of Interspeech-2014*, Singapore, September 2014, pp. 2804–2808.

[32] H. Pulakka, V. Myllylä, A. Rämö, and P. Alku, "Speech quality evaluation of artificial bandwidth extension: Comparing subjective judgements and instrumental predictions," in *Proc. of Interspeech-2015*, Dresden, Germany, September 2015.

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

[34] McGill University, "AFsp programs and routines," http://www-mmsp.ece.mcgill.ca/documents/Software/Packages/AFsp/AFsp.html.

[35] ITU-T G.191, "Software tools for speech and audio coding standardization," Geneva, November 2009.

[36] ITU-T P.810, *Telephone transmission quality: methods for objective and subjective assessment of quality. Modulated noise reference unit (MNRU)*, ITU, February 1996, online: https://www.itu.int/rec/T-REC-P.810-199602-I/en.

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

[38] ITU-T Rec. P.863, "Perceptual objective listening quality assessment," September 2014.

[39] ITU-T P.Imp863, "Implementer's guide on assessment of EVS coded speech with recommendation ITU-T P.863," January 2016.