Enhanced AMR-WB Bandwidth Extension in 3GPP EVS Codec

Magdalena Kaniewska,
Stéphane Ragot
Orange Labs, France

Zexin Liu, Lei Miao, Xingtao
Zhang, Jon Gibbs
Huawei Technologies Co. Ltd, China

Vaclav Eksler
VoiceAge Corp., QC, Canada

***Abstract*— This paper presents the bandwidth extension (BWE) method developed for the AMR-WB interoperable (AMR-WB IO) modes of the 3GPP EVS codec. The low-band signal (0-6.4 kHz) is coded using an enhanced version of ACELP as in AMR-WB and post-processed; the high-band (above 6.4 kHz) in contrast to AMR-WB is represented with a new BWE method. The decoded low-band excitation is adaptively extended to high frequencies and filtered in the DCT domain. The extended excitation is scaled by subframe gains and shaped by a weighted LPC synthesis filter. Test results show that the AMR-WB IO BWE contributes to the performance advantage of EVS AMR-WB IO over the original AMR-WB, with limited complexity increase and no extra delay.**

Index Terms— **BWE, AMR-WB, speech coding, EVS**

I. INTRODUCTION

The AMR-WB codec [1], which is widely deployed and a key enabler for HD voice, was standardized in 2001 for GSM and UMTS. This codec was also standardized in 2003 as ITU-T G.722.2 [2]. The usage of AMR-WB has been extended to packet-switched services, including Voice over LTE (VoLTE). The Enhanced Voice Services (EVS) codec, standardized by 3GPP in Sept. 2014, is a state-of-the-art multi-rate speech/audio coder that provides, in addition to the EVS primary modes, backward compatibility with AMR-WB [3]. The backward compatible modes of EVS are referred to as the AMR-WB interoperable (AMR-WB IO) modes. The objective of this paper is to describe the enhancements to the AMR-WB bandwidth extension (BWE) in the EVS.

BWE is a technique in which the signal is split into two frequency bands; a low-band and a high-band, where the high-band is coded separately or estimated to extend the audio bandwidth of the low-band. BWE is a key method employed to code audio signals efficiently at low bit rates where the BWE associated bit budget is usually very low compared to the low-band coding, or even zero. We focus in this paper on BWE for codecs used in conversational services. Early variants of BWE considered either coding the high-band with waveform coding [4] or extrapolation of the high-band without side information, e.g. based on the linear prediction coding (LPC) of the residual signal [5]. Many time-domain BWE methods have been proposed based on LPC techniques; and in most cases a high-band excitation is generated in the time-domain, scaled and then shaped by LPC synthesis filtering [6, 22]. Variants of excitation shaping have been proposed [7] and noise excitation has also been considered [1, 8]. Frequency-domain BWE methods operate in the transform-domain. Low-delay examples can be found in [23] using an FFT or in [9, 10, 24] using the modified discrete cosine transform (MDCT). Subband-domain BWE is an alternative approach that has been successfully used in conversational applications, e.g. in [11]. BWE methods may be adapted to perform blind BWE [12, 23] where no side information is transmitted. BWE methods used in non-conversational applications are not considered here but examples are described in [13, 14, 25].

In this paper we address BWE as one of the backward compatible enhancements to AMR-WB developed to satisfy a key objective of the 3GPP EVS standardization [15]. Due to design constraints on bitstream interoperability with the original AMR-WB, the (enhanced) AMR-WB IO operation must remain blind for bit rates from 6.6 to 23.05 kbit/s; with side information (at 0.8 kbit/s) only available at 23.85 kbit/s. We describe in this paper how the original AMR-WB BWE has been improved in EVS given these interoperability constraints.

The paper is organized as follows. The original BWE method used in AMR-WB is reviewed in Section II. The design of the improved AMR-WB IO BWE method is described in Section III. Quality, complexity, and delay are discussed in Section IV, before concluding in Section V.

II. AMR-WB BWE: REVIEW AND DISCUSSION

AMR-WB [1] is based on ACELP coding of a low-band signal (0-6.4 kHz band) after decimating the input signal from 16 to 12.8 kHz. As illustrated in Fig. 1, the high-band (6.4-7 kHz) of the AMR-WB BWE is generated by shaping a white noise signal with a time envelope (subframe gains) and a frequency envelope (LPC synthesis filter). An artificial excitation is generated every 5 ms as white noise sampled at 16 kHz. This excitation is scaled in each 5 ms subframe by a gain which is composed of two sub-factors: one factor brings the white noise (at 16 kHz) to a level comparable with the ACELP decoded excitation in the low-band (at 12.8 kHz sampling rate), including an implicit attenuation by a 12.8/16 ratio; the other factor is a gain correction factor which is either estimated based on the spectral tilt of the low-band signal (at bit rates of 23.05 kbit/s and below) or decoded from the bitstream at 23.85 kbit/s. The artificial excitation, shaped in the time-domain, is then filtered by a weighted LPC synthesis filter operating at 16 kHz. The LPC filter at 16 kHz is obtained by applying perceptual weighting to the low-band ACELP LPC filter. The reconstructed signal in the high-band is then band-pass filtered by a FIR filter (with 0.9375 ms delay) to keep just the 6-7 kHz frequency region (giving an overlap with low-band, to avoid an energy drop at 6-6.4 kHz, where the low-band signal is already rolling off). At 23.85 kbit/s, an extra low-pass FIR filter (with 0.9375 ms delay) is used to further attenuate frequencies above 7 kHz. The resulting high-band signal is then added to the decoded low-band signal to obtain the final synthesis signal at 16 kHz.

978-1-4799-7591-4/15/$31.00 ©2015 IEEE

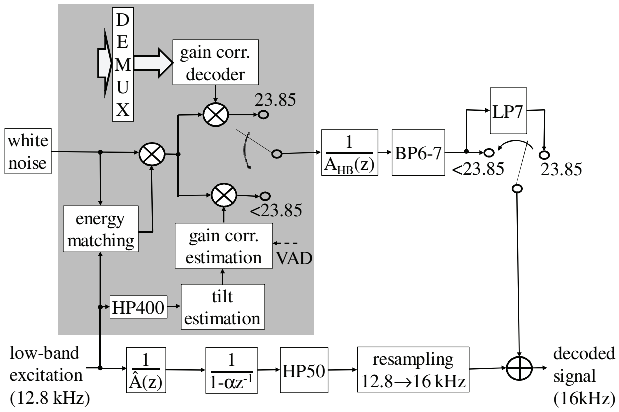

Figure 1. AMR-WB BWE decoding diagram.

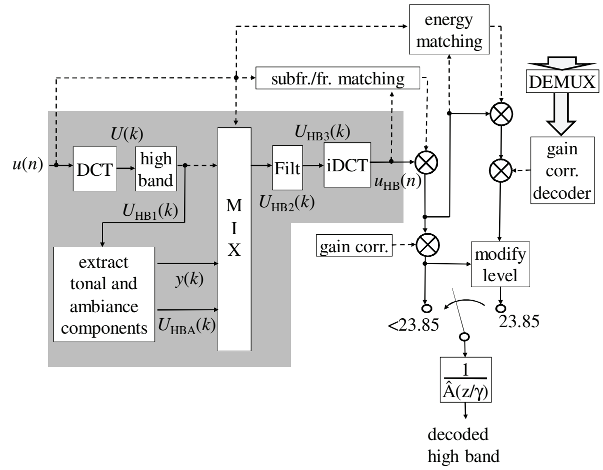

Figure 2. BWE decoding in EVS AMR-WB IO.

The AMR-WB BWE has proved to be effective at very low bit rates for speech signals. However, it has several drawbacks:

- The high-band signal model relies entirely on shaping a white noise signal in both the time and frequency domains. This signal model is too limited to represent general signals above 6.4 kHz. For example, for music signals with harmonics extending above 6.4 kHz, this model introduces artifacts.
- The additional low-pass FIR filter at 23.85 kbit/s introduces extra delay, which time shifts the high-band slightly (by 0.9375 ms) in relation to the low-band. Such misalignment is ideally to be avoided in general.
- Characterization test results of AMR-WB [16] indicated that the quality of AMR-WB at 23.85 kbit/s is lower than at 23.05 kbit/s and quite similar to 15.85 kbit/s for clean speech signals. This observation indicates that the side information (0.8 kbit/s) available at 23.85 kbit/s to code the high-band may be better exploited. It has been found that the level of the high-band artificial excitation needs to be carefully controlled. In general, the performance of the AMR-WB BWE method is very sensitive to the tuning of the grey shaded part of Fig. 1, particularly related to the scaling of the excitation by the subframe gain has to be well controlled to avoid artifacts.

Note that in ITU-T G.718 [17] the AMR-WB BWE method was re-used for the AMR-WB compatible decoding, with minor adjustments. The 6-7 kHz band-pass filtering and LPC synthesis filtering in the high-band have been swapped. The 4 bits per subframe available at 23.85 kbit/s for high-band gain correction are discarded, which suppresses the quality issue discussed previously. Finally the 7 kHz low-pass filter specific to 23.85 kbit/s is also not used. G.718 also included some post-processing of the low-band signal which may indirectly benefit the BWE. However, the quality improvement for the G.718 AMR-WB IO BWE is still not significant, particularly for mixed content and music signals, mostly due to the use of same white noise excitation as in original AMR-WB.

III. Enhanced BWE in AMR-WB IO

For interoperability reasons, the quantization of the high-band gain in each 5 ms subframe at 23.85 kbit/s is conducted in EVS AMR-WB IO using the same method as in the original AMR-WB. The following description is therefore focused on BWE decoding.

The high-frequency band generation is illustrated in Fig. 2. The principles of the main processing steps are described below; while a detailed algorithmic description of AMR-WB IO BWE can be found in [26, clause 6.8.3]. The main modification compared to the original AMR-WB BWE is the generation of the excitation signal, which is highlighted in the grey box in Fig. 2. The gain scaling and filtering steps are also modified to some extent.

A. Generation of High-Band Excitation in DCT Domain

The current frame of the decoded excitation from the low-band, $u(n)$, $n = 0,\cdots,255$, sampled at 12.8 kHz, is transformed by a DCT-IV of length 256 to obtain the spectrum, $U(k)$, $k = 0,\cdots,255$. The advantage of performing the extension in DCT domain is the possibility to control the spectral contents of the high-band by precisely choosing the portion of low-band spectrum to be copied, while maintaining low complexity. The DCT-IV transform was used to eliminate additional Program ROM (it is used in the EVS codec for low-band decoder post-processing and for MDCT computation) and to avoid extra overlap-add delay (when compared to an MDCT using a non-rectangular window).

The DCT spectrum covering the 0-6.4 kHz band is extended to 0-8 kHz, according to the start frequency bin k_{start} as follows:

$$U_{HB1}(k) = \begin{cases} 0, & k = 0,\cdots,199, \\ U(k), & k = 200,\cdots,239, \\ U(k + k_{start} - 240), & k = 240,\cdots,319, \end{cases} \tag{1}$$

where k_{start} is the start frequency bin discussed later. The operation in (1) performs an implicit resampling from 12.8 to 16 kHz with a DCT length extended to 320 samples. Components from 5 kHz up are provided to enable smooth shaping of transition band. The 5-6 kHz band in $U_{HB1}(k)$ is simply copied from $U(k)$, k = 200,…,239, without modification, because in this band the original spectrum is maintained to avoid introducing distortions when the high-band is added to the decoded low-band signal. This method also allows for high-pass filtering to be implemented in the DCT domain (see section III.C). The 6-8 kHz band in $U_{HB1}(k)$ is adaptively copied from $U(k)$, based on the start frequency bin k_{start}. The value of k_{start} is adaptively determined from the line spectrum frequency (LSF) parameters (in Hz) decoded in the low-band. The difference (delta), $d_{LSF}(k)$, between adjacent LSF frequencies is calculated and the index of the minimum difference is searched, given that this corresponds to an energy peak of the low-band spectral envelope:

$$k_{\min} = \underset{k \in [2, M2)}{\operatorname{argmin}} \left(d_{LSF}(k) \cdot \max(1 - LSF(k) \cdot W,\ 0.001) \right), \quad (2)$$

where $M2$ and W depend on bit rate [26]. The start frequency bin k_{start} is calculated as:

$$k_{start} = \left\lfloor \frac{0.5 \cdot (LSF(k_{\min}) + LSF(k_{\min} - 1)) \cdot 40}{1000} - 40 \right\rfloor \quad (3)$$

and k_{start} is limited to the [40,160] range. In order to mitigate frequent inter-frame variations of the start frequency bin, the start frequency bin of current frame is set to the one of previous frame under certain conditions according to several different parameters [26].

B. Extraction of Tonal and Ambiance Components and Adaptive Re-Combination

Tonal and ambiance components are extracted in the 6-8 kHz band, to be able to control the tonality level of the high-band. The ambiance $U_{HBA}(k)$ (in absolute value) corresponds to the local average of the magnitude spectrum $|U_{HB1}(j+240)|$ over a sliding window of 15 bins. Tonal components are defined as the residual signal

$$y(k) = |U_{HB1}(k+240)| - U_{HBA}(k) \quad (4)$$

satisfying the detection criterion $y(k) > 0$, $k = 0, \ldots, L-1$, with $L = 80$.

The extracted tonal and ambiance components are then adaptively re-mixed. The combined signal is obtained using absolute values as:

$$y'(k) = \begin{cases} \Gamma y(k) + U_{HBA}(k)/\Gamma, & y(k) > 0, \\ y(k) + U_{HBA}(k)/\Gamma, & y(k) \le 0, \end{cases} \quad k = 0 \ldots L-1, \quad (5)$$

where the factor controlling the ambiance

$$\Gamma = \beta \frac{ener_{HB} - ener_{tonal}}{ener_{HB} - \beta ener_{tonal}}. \quad (6)$$

In (6), β is the adaptive multiplicative factor and $ener_{HB}$ and $ener_{tonal}$ are respectively the energy of the excitation $U_{HB1}(k)$ and tonal components $y(k)$. The value of β is derived from several signal parameters [26]. The signs of $U_{HB1}(k)$ are then applied to $y'(k)$. Finally the scaling with the adaptive attenuation factor is applied to restore the overall energy $ener_{HB}$ and obtain the combined high-band signal $U_{HB2}(k)$.

C. Filtering in DCT Domain and Inverse DCT

Unlike in original AMR-WB BWE, filtering of the synthesized high-band is performed in the DCT (excitation) domain. The associated circular convolution was verified to have no adverse impact, given that this processing is performed in the excitation domain and filter spectral shapes are forced to be smooth. The excitation is de-emphasized using the frequency response of the filter $1/\left(1 - 0.68 z^{-1}\right)$ over the 6-8 kHz frequency range. This de-emphasis operation is used to revert the pre-emphasis and be consistent with the low-band signal (in the 0-6.4 kHz band), which is useful for the subsequent energy estimation and adjustment.

The excitation is also band-pass filtered in DCT domain with cut-off frequencies at 6 kHz and 7-7.8 kHz. The variable higher cut-off is motivated by the fact that adding too much bandwidth above 7 kHz may not be desirable at lowest low bit rates (6.6, 8.85 kbit/s) because the low-band quality is limited and typically degrades quality compared to limiting BWE to 7 kHz. For higher bit rates the 7.8 kHz upper limit has proven empirically to be the best trade-off between more presence and less artifacts.

Finally, the resulting high-band excitation $U_{HB3}(k)$ is transformed by inverse 320-point DCT to obtain the time-domain signal $u_{HB}(n)$ of the current frame.

D. Gain Computation and Scaling of Excitation

The signal $u_{HB}(n)$, $n = 0, \cdots, 320$, is split in subframes of 5 ms and scaled to keep the same subframe to frame energy ratio as the decoded low-band signal $u(n)$. This scaling is refined by estimating a correction factor equalizing the low-band and high-band LPC spectral envelopes in a cross-over region centered around 6.4 kHz. At 6.4 kHz the shape of the low-band LPC filter, $1/\hat{A}(z)$, is typically rolling off too quickly. Therefore, the frequency of 6 kHz was chosen as the frequency at which the high and low-band envelopes are matched – the frequency of 6 kHz is also seen as the cross-over region lower limit. In each 5 ms subframe, the frequency response of the LPC filter $1/\hat{A}(z)$ in the low-band and the LPC filter $1/A_{HB}(z)$ in the high-band are computed at the frequency of 6 kHz. The ratio of frequency responses at 6 kHz provides an estimated gain correction to be used to align the level of LPC spectral envelopes in two different bands. This principle was further adjusted using 2nd order LPC filters to optimize the correction factor estimation, in particular to avoid over-estimation. The optimized correction factor is actually estimated based on the frequency response ratio at 6 kHz and also an estimated tilt, voicing and several other signal parameters [26] to avoid over-estimating the subframe gain in the high-band.

At 23.85 kbit/s, to be able to use the gain correction from the bitstream, the excitation is converted into a time-domain signal similar to that of the AMR-WB high-band coding, by energy matching. The decoded 4-bit index in each subframe is applied to the excitation, which is further level adjusted by a subframe factor, in particular based on the de-emphasis filter characteristics and the tilt information of the low-band signal.

E. LPC Synthesis Filtering and Final Synthesis

In a way similar to the original AMR-WB BWE, the scaled excitation in the high-band is filtered in each 5 ms subframe by $A_{HB}(z) = \hat{A}(z/\gamma_{HB})$, where $\gamma_{HB} = 0.9$ at 6.6 kbit/s and 0.6 from 8.85 to 23.85 kbit/s, to obtain the decoded time-domain high-band signal at 16 kHz. The high-band signal is further delayed to compensate for the low-band resampling delay. As the EVS AMR-WB IO modes support output signals sampled at 16/32/48 kHz, interpolation filters are used to resample the high-band signal to the output sampling rate. Finally the high-band time-domain signal is added to the low-band time-domain signal to produce the final synthesis at the output sampling rate.

IV. Evaluation

The AMR-WB IO modes of the EVS codec have been formally tested with ITU-T P.800 [18] ACR and DCR subjective tests during the EVS selection and characterization phases. More details can be found in the EVS characterization report in [20].

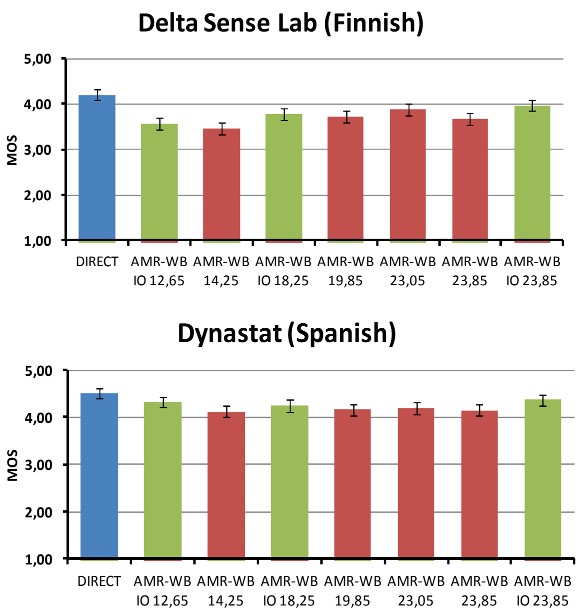

Figure 3. Clean speech quality of EVS AMR-WB IO (ACR) [20].

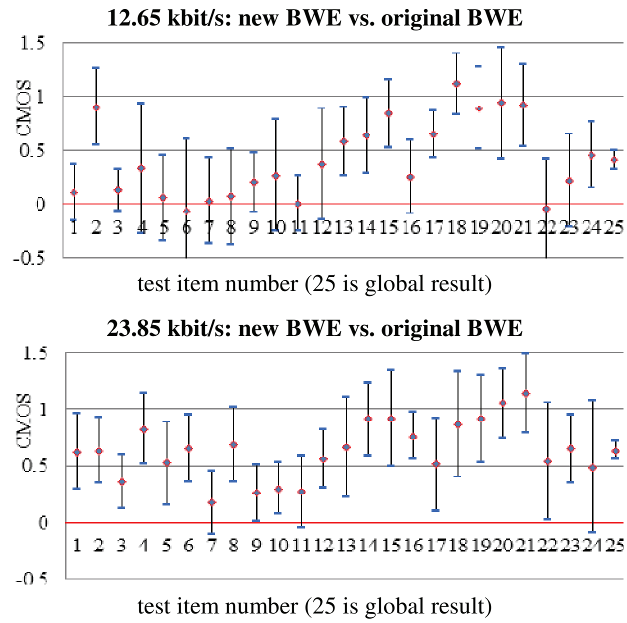

Figure 4. Ref/A/B assessment of enhanced BWE.

Fig. 3 provides an extract from official 3GPP EVS selection test results for AMR-WB IO in clean speech for conditions at -26 dBov [20]. The experiment was conducted in two languages (Finnish in Delta SenseLab, Spanish in Dynastat) using clean speech with 6 talkers and 5 sentence pairs per talker and 32 naïve listeners [19]. Sennheiser HD 280 Pro headphones were used for diotic listening. In Fig. 3, not all of the bit rates of AMR-WB and AMR-WB IO are listed because the overall number of test conditions had to be limited to fit into one ACR test. Furthermore, the bit rates of the two codecs do not always match because AMR-WB IO was required to perform not worse than the original AMR-WB operating at a higher bit rate to ensure enhanced performance [21]. It is important to note that all conditions (apart from the original speech denoted 'DIRECT') were coded using the same encoder (original AMR-WB encoder) and the decoder was either original AMR-WB (red bars in Fig. 3) or EVS AMR-WB IO (green bars in Fig. 3).

Results in Fig. 3 confirm earlier observations that AMR-WB shows degraded performance at 23.85 kbit/s in clean speech and the results also demonstrate that the performance at this bitrate has been improved in AMR-WB IO. They indicate that EVS AMR-WB IO provides consistent improvement compared to AMR-WB operating at the next higher bit rate. It is important to note that the results in Fig. 3 capture the overall quality of the EVS AMR-WB IO modes. The quality improvement of the AMR-WB IO modes comes from the new BWE as well as enhancements to the low-band decoding, e.g. formant sharpening, dynamic normalization, and improved post-processing [3]. The average effect of the low-band post-processing is quite limited in the case of clean speech at nominal level (-26 dBov). However, for mixed content and music, the frequency-domain BWE method of the AMR-WB IO benefits significantly from low-band enhancements, especially the inter-tone noise reduction [3], since the high-band excitation is generated based on the low-band signal after post-processing. It is also worth noting that AMR-WB IO provides slightly more audio bandwidth (up to 7.8 kHz) than AMR-WB (up to 7 kHz) which can also provide a slight advantage in quality.

To isolate the performance contribution of enhanced BWE, a Ref/A/B test with P.800 CCR grading scale [27] was conducted; where codec "A" was the EVS codec, while Codec "B" was the same codec but replacing the BWE by the original white-noise excitation based BWE used in AMR-WB and G.718 AMR-WB IO. The test was conducted in the Huawei Lab at two bit rates (12.65 and 23.85 kbit/s) with 8 expert listeners. It should be noted that at 23.85 kbit/s the original BWE did not use subframe gains (i.e. 0.8 kbit/s side information) in a similar manner to G.718 AMR-WB IO. In total 24 test samples were used with 12 speech samples in Mandarin Chinese (6 clean and 6 noisy) and 12 mixed content/music samples (6 mixed content and 6 music). The Ref/A/B test results are presented in Fig. 4 with the test item number in the x-axis (first 12 are speech, next 12 are music/mixed content) and the Comparative Mean Opinion Score (CMOS) value in y-axis. The last item (#25) corresponds to the average score. In Fig. 4, a positive score indicates a quality advantage for the new BWE. It can be seen that at both 12.65 kbit/s and 23.85 kbit/s quality is improved due to the enhanced EVS AMR-WB IO BWE. The improvement is more pronounced for mixed/music items and it is greatest at 23.85 kbit/s.

The AMR-WB IO BWE method is around 1.2 WMOPS more complex and adds about 0.2 kWords of ROM and 1.5 kWords of RAM compared with the original AMR-WB BWE method. AMR-WB IO BWE has in principle no extra delay compared to the low-band decoding, since all FIR filtering steps are replaced by DCT processing. However, in EVS, the BWE output is delayed to be time-synchronized with the low-band output.

V. Conclusion

In this paper the enhanced AMR-WB BWE method used in the EVS codec has been described. Quality is improved in particular thanks to the new excitation generation and processing in DCT domain (instead of white noise generation), refined subframe gain computation, and a slightly extended bandwidth up to 7.8 kHz, that allowed increasing the effect of bandwidth extension while limiting the artifacts coming from over-estimation of energy.

REFERENCES

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

[2] ITU-T G.722.2, Wideband coding of speech at around 16 kbit/s using Adaptive Multi-Rate Wideband (AMR-WB), July 2003.

[3] Martin Dietz et al., "Overview of the EVS Codec Architecture," in Proc. ICASSP, pp. 5698-5702, Apr. 2015.

[4] X. Maitre, "7 kHz audio coding within 64 kbit/s," IEEE Journal on Selected Areas in Communications, vol. 6, no. 2, Feb. 1988.

[5] J. Makhoul and M. Berouti, "High-frequency regeneration in speech coding systems," in Proc. ICASSP, pp. 428-431, vol. 1, Apr. 1979.

[6] P. Jax and P. Vary, "Feature selection for improved bandwidth extension of speech signals," in Proc. ICASSP, pp. 697-700, May 2004.

[7] A. McCree, "A 14 kb/s wideband speech coder with a parametric highband model," in Proc. ICASSP, pp. 428-43, June 2000.

[8] B. Geiser, P. Jax, P. Vary, H. Taddei, S. Schandl, M. Gartner, C. Guillaume, and S. Ragot, "Time-domain Bandwidth Extension for Hierarchical Speech and Audio Coding in ITU-T Rec. G.729.1," IEEE Transactions on Speech, Audio and Language Processing, vol. 15, no. 8, pp. 2496-2509, Nov. 2007.

[9] M. Lei, L. Zexin, C. Hu, V. Eksler, S. Ragot, C. Lamblin, B. Kovesi, J. Sung M. Fukui, S. Sasaki, and Y. Hiwasaki, "G.711.1 Annex D and G.722 Annex B – New ITU-T superwideband codecs," in Proc. ICASSP, pp. 5232-5235, May 2011.

[10] B. Kovesi, S. Ragot, C. Lamblin, M. Lei, Z. Liu, and C. Hu, "Re-engineering ITU-T G.722: Low delay and complexity superwideband coding at 64kbit/s with G.722 bitstream watermarking," in Proc. ICASSP, pp. 5248-5251, May 2011.

[11] M. Schnell, R. Geiger, M. Schmidt, M. Jander, M. Multrus, G. Schuller, and J. Herre, "Enhanced MPEG-4 low delay AAC—Low bitrate high quality communication," in 122nd Convention of AES, May 2007.

[12] Y.M. Cheng, D. O'Shaughnessy, and P. Mermelstein, "Statistical recovery of wideband speech from narrowband speech," IEEE Trans. Speech and Audio Proc., vol. 2, no. 4, pp. 544-548, 1994.

[13] M. Dietz, L. Liljeryd, K. Kjorling, and O. Kunz, "Spectral Band Replication, a novel approach in audio coding," in 112th Convention of AES, Munich, 2002.

[14] C. Neukam, F. Nagel, G. Schuller, and M. Schnabel, "A MDCT based harmonic spectral bandwidth extension method," in Proc. ICASSP, pp. 566 – 570, May 2013.

[15] S. Bruhn et al., "Standardization of the new EVS Codec," in Proc. ICASSP, pp. 5703-5707, Apr. 2015.

[16] 3GPP TR 26.976, Performance characterization of the Adaptive Multi-Rate Wideband (AMR-WB) speech codec, v.12.0.0, Sept. 2014.

[17] M. Jelínek et al., "G.718: A New Embedded Speech and Audio Coding Standard with High Resilience to Error-Prone Transmission Channels," IEEE Communications Magazine, vol. 47, no. 10, pp. 117-123, Oct. 2009.

[18] ITU-T P.800, Methods for Subjective Determination of Transmission Quality, Aug. 1996.

[19] 3GPP Tdoc S4-140736, EVS Permanent Document EVS-8b: Test plans for selection phase including lab task specification, v.1.1.0, May 2014.

[20] 3GPP TR 26.952, Codec for Enhanced Voice Services (EVS); Performance characterization, v.12.2.0, June 2015.

[21] 3GPP Tdoc S4-130522, EVS Permanent Document EVS-3: EVS performance requirements, v.1.4, April 2013.

[22] V. Krishnan, V. Rajendran, A. Kandhadai, and S. Manjunath, "EVRC-Wideband: The New 3GPP2 Wideband Vocoder Standard," in Proc. ICASSP, Apr. 2007.

[23] L. Laaksonen, H. Pulakka, V. Myllyla, and P. Alku, "Development, evaluation and implementation of an artificial bandwidth extension method of telephone speech in mobile terminal," IEEE Trans. On Consumer Electronics, vol. 55, no. 2, pp. 780-787, 2009.

[24] K. Tsujino and K. Kikuiri, "Low-complexity Bandwidth Extension in MDCT domain for low-bitrate speech coding," in Proc. ICASSP, Apr. 2009.

[25] J. Zhan, K. Choo, E. Oh, "Bandwidth Extension for China AVS-M standard," in Proc. ICASSP, Apr. 2009.

[26] 3GPP TS 26.445, Codec for Enhanced Voice Services (EVS); Detailed algorithmic description, v.12.2.0, March 2015.

[27] ITU-T P.800 Annex E, Comparison Category Rating (CCR) method, Aug. 1996.