ENGINEERING RESEARCH JOURNAL (ERJ)

Vol. 1, No. 41 July. 2019, pp. 80-88

Journal Homepage: www.feng.bu.edu.eg

EVS LTE Coder's Performance under Different Frame Loss Conditions

Christina Gamal, Michael N. Mikhael and Hala A. Mansour

Faculaty of Engineering, Benha University, Cairo,

Abstract. The quality of speech is essential in real-time communications. There are many aspects that can affect the quality of speech; one of them is the error concealment techniques that are included in the codec itself which have a significant role to recover the lost frames due to error transmission. The codec for Enhanced Voice Services (EVS) was standardized by 3GPP for LTE network. EVS has several error concealment techniques to deal with different coding modes achieving higher robustness against error transmission. This paper presents the performance evaluation of EVS codec at bit-rate 8 kbps through losing and zero padding frames from one frame to forty frames with Arabic and English languages compared with AMR-NB codec at 7.95 kbps. The testing evaluation was done using PESQ objective method. The test results show that EVS codec achieves better quality than AMR codec in case of zero padding frames by 8% and 18%, where the frames are conducted continuously and intermittently, respectively. On the other hand, by losing frames both coders are relatively close to each other by 2% in terms of quality.

Keywords: EVS, AMR, PESQ, Frame losses.

1. INTRODUCTION

EVS codec was evaluated under channel error conditions (FER 3% and 6%) at bit-rate 9.6 kbps in narrow band against AMR-NB codec at 12.2 kbps, the used database included clean speech for Finnish language and North American English language, noisy speech for Swedish language and French language, and Mixed content/music for Danish language and Latin American Spanish language, all the results emphasized the improved robustness of EVS against transmission errors, as compared to AMR [1].

Anssi, Antti, and Henri showed that EVS codec at bit-rate 8 kbps is better than AMR-NB at 12.2 kbps, at every frame erasure rate which is 0%, 5%, 10%, 20% and 30%, although at low FER rates the difference is not significant, using subjective and objective tests [2].

Jérémie Lecomte and others, tested the TCX time domain concealment techniques in EVS codec under clean and impaired channel conditions (6% FER), for wideband at bit-rates 9.6 kbps and 24.4 kbps, against AMR-WB/G.718 IO at bit-rates 12.65 kbps and 23.85 kbps for noisy speech under impaired channel conditions. The used database included an English male, an English female and a German male with different speech effects. The subjective listening tests showed that EVS outperforms AMR-WB/G.718 IO for clean and noisy channel, as well as EVS with 6% packet loss, competes with the clean channel AMR-WB/G.718 IO at bit-rates around 24 kbps [3]. Also, EVS codec at bit-rate 7.2 and 8 kbps was compared with AMR-WB at 12.65 and 15.85 kbps for mixed content and music with 6% frame loss, using hybrid concealment technique that mixes time domain coding with frequency domain coding, the results showed that EVS is better than AMR-WB [4]. EVS codec was tested against lost frames with 3% and 6% FER compared to AMR-WB/G.718 IO for wideband clean speech, AMR-WB and G.722.1 for WB mixed and music and G.722.1C and G.719 for super wideband (SWB) clean speech. The testing was done using the subjective mean opinion score to evaluate the performance of EVS which showed that EVS outperforms the reference coders [5].

This paper presents the performance evaluation of EVS codec in the narrow band at bit-rate 8 kbps, compared with AMR-NB at bit-rate 7.95 kbps under different packet loss conditions by losing frames or inserting zeros instead of frame data content. The evaluated results were done using PESQ objective test. The frames were conducted continuously or discontinuously. The database used in the experiment includes clean speech for Arabic and English language, with different genders for each language (i.e. American English accent and Saudi Arabia accent).

1.1.1 AMR Frame Description

AMR codec has 8 different bit-rates: 12.2 kbps, 10.2 kbps, 7.95 kbps, 7.40 kbps, 6.70 kbps, 5.90 kbps, 5.15 kbps and 4.75 kbps, which are corresponded to modes 7 through 0, respectively. The switching between different bit-rates depends on the channel error conditions, in case of a busy channel; the existing bit-rate is changed to the lower bit-rate. Not only the channel conditions determine the bit-rate but also the AMR source which is based on the theoretic content and characteristic of the speech signal [6].

The input speech file to the AMR encoder is 16-bit binary data sampled at 8 kHz without header's content. The AMR encoder generates frames each with size 20ms; the number of bytes in each frame depends on the mode of the frame as shown in the table (1). The first byte of each encoded frame indicates the mode and the quality of the frame as shown in the table (2).

TABLE 1: Number of Bytes in each Frame [7]

Mode	**Bit Rate**	**Number of bytes in each frame**
0	4.75 kbps	13
1	5.15 kbps	14
2	5.90 kbps	16
3	6.70 kbps	18
4	7.40 kbps	20
5	7.95 kbps	21
6	10.2 kbps	27
7	12.2 kbps	32
8	SID	6
15	No data	1

TABLE 2: First Byte Representation [7]

Bit 0	Bit 1	Bit 2	Bit 3	Bit 4	Bit 5	Bit 6	Bit 7
Pad	Mode				Quality	Pad	Pad

From table (2), Bit 5 indicates the quality of the frame; if the bit value is logic 0 this means a bad frame and logic 1 means a good frame, while bit 0, bit 6, and bit 7 are not used bits [7].

E.g. if the frame is encoded with 7.95 kbps and it is received by the decoder with good quality, then the first byte of the frame presents in binary as (00101100) and the number of the bytes in the frame is 21.

1.1.2 AMR Error Concealment of Lost Frames

AMR error concealment method depends on the network, the network shall set the flag (RX_TYPE) values to SPEECH_BAD or SID_BAD in which case the Bad Frame Indication (BFI) flag is set to 1 to indicate bad data frame based on lost speech or lost Silence Descriptor frame (SID), respectively. And also, the flag (RX_TYPE) was set to value SPEECH_PROBABLY_DEGRADED in which case the Potentially Degraded Frame Indication (PDFI) flag is also set to 1 based on the degraded frame. In case of lost speech, the lost frames are substituted with either a repetition or an extrapolation of the previous good speech frame(s). On the other hand, in case of losing many frames subsequently, a muting technique should be applied [8].

1.2.1 EVS Frame Description

In addition to EVS primary modes which are 11 different bit-rates and one variable bit-rate, supporting NB, WB, SWB, and FB; EVS has 9 different bit-rates (AMR-WB IO modes).

The input file to the EVS encoder is 16-bit binary data without header's content, sampled at 8 kHz, 16 kHz, 32 kHz or 48 kHz [9]. The data is read and written in 16-bit words. The output from EVS encoder is a bit-stream file in either ITU G.192 or EVS-MIME file storage format.

The format of ITU G.192 for every frame, each with size 20ms is divided into 16-bit word as shown in the table (3).

TABLE 3: The Representation of ITU G.192's Frame [10]

Sync Word	Data Len	1st Data Bit	2nd Data Bit	....	Nth Data Bit

The Sync Word is always 0x6B21.

The Data Len represents the number of 16-bit data words in the frame that differs for each bit-rate and in case of bit-rate = 8kbps, the Data Len is 0x00A0, but in the case using DTX there is no data, the Data Len is zero. Bit 0and Bit 1 is represented as 0x007F and 0x0081, respectively [10].

1.2.2 EVS Packet Loss Concealment Techniques

The error concealment techniques are first based on signal classification for frame erasure concealment (FEC) which is different from signal classification for coding mode [5].In case the last good frame before the erased frame was coded with ACELP and its classification is other than UNVOICED class, the periodic part of the excitation is constructed by repeating the low-pass filtered last pitch period of the previous frame. The random excitation which is generated randomly, filtered through a linear phase FIR high pass filter to decrease the number of noisy components during voiced segments. For rates 5.9, 7.2, 8.0, 13.2, 32 and 64 kbps as shown in figure (1), the periodic excitation and the random excitation are added together to form the total excitation which is then post-processed and filtered through an LP synthesis filter to obtain the synthesized signal. However, for rates of 9.6, 16.4 and 24.4 kbps the periodic excitation and the random excitation are first filtered separately through an LP filter then added together [3]. Then, the synthesized signal is de-emphasized and passed through an adaptive post-processing for enhancing the formant and harmonic structure of the signal. The signal is then up-sampled to the output selected rate. If the last good frame is UNVOICED class, only the innovation (random) excitation is used and the periodic part of the excitation is not generated.

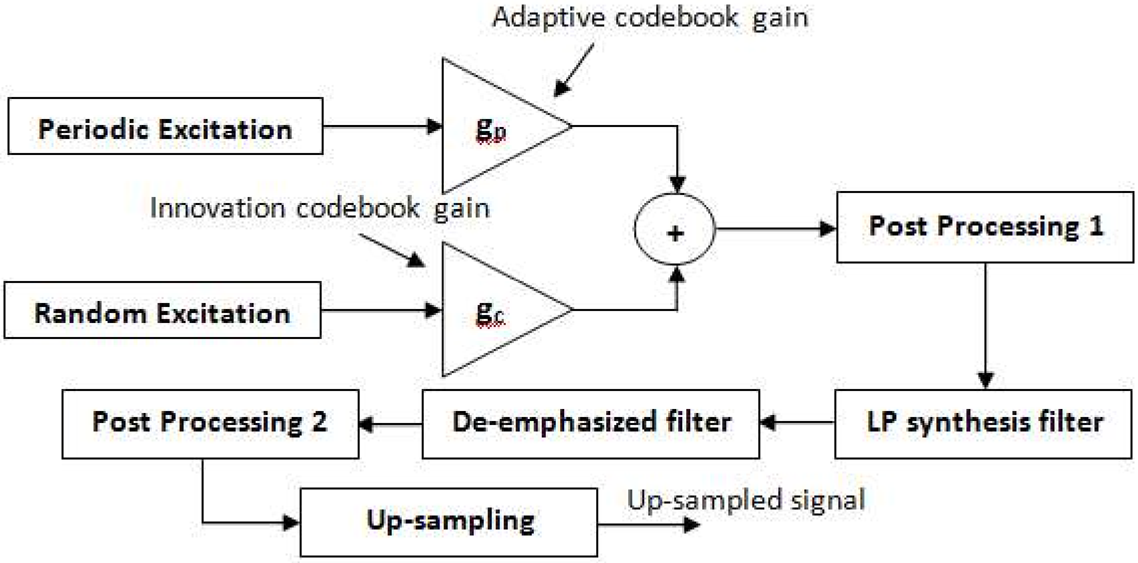


Fig 1: Block Diagram of Packet Loss Concealment with ACELP.

For consecutive frame loss, the periodic and the innovation excitations are faded towards zero and Comfort Noise Generation (CNG) excitation respectively by changing the codebook gains as in equation (1) [11].

$$(1) g^{[m]} = \alpha g^{[m-1]} + (1-\alpha). g^{target}.$$

Where

- $g^{[m]}$ is the gain of the current frame,
- $g^{[m-1]}$ is the gain of the previous frame,
- g^{target} is the target gain which is zero for periodic excitation and CNG level for random excitation,
- α is the fading factor (e.g. for VOICED class, α is 1 with maximally three consecutive lost frames and 0.4 for more than 3).

The faded output signal for the periodic or the random excitation in each sub-frame is computed as in equation (2).

$$Faded\ signal[i] = \left[g^{[m-1]} - \left(\frac{i}{L_{frame}}\right)\left(g^{[m-1]} - g^{[m]}\right)\right].signal[i]$$

$$i = 0, \ldots., L_{frame} - 1. \qquad (2)$$

Where L_{frame} is the length of the frame and signal[i] is the input signal.

For a long burst of losses with many consecutive lost frames, a muting strategy is applied [11].

1.3 Objective Evaluation Method

The objective evaluation methods are divided into two types, intrusive and non-intrusive. The intrusive method is one that compares original speech with degraded speech, although this method is not used for measuring the speech quality for real-time communication. The second method is the non-intrusive method that measures the quality based only on the degraded speech, so it is suitable for real-time quality measurement [12].

The Perceptual Evaluation of Speech Quality (PESQ) (ITU-T P.862) is the intrusive objective evaluation method that is used to measure end-to-end speech quality [13].

The value of PESQ MOS is between -0.5 and 4.5 which differs from the subjective test value that is between 1 for bad quality and 5 for excellent quality. Equation (3) recalculates PESQ MOS to be suitable with subjective test values so the range will be from 1 to 4.55 [12].

$$y = 0.999 + \frac{4.999-0.999}{1+e^{-1.4945x+4.6607}} \qquad (3)$$

Where y is the MOS value that is matched with the subjective test (ITU-T P.862.1) and x is the PESQ MOS value. PESQ has the ability to measure the quality of the speech in case of single frame losses compared to the subjective test that has difficulty to recognize the effect of frame losses [14].

3. Experimental Explanation

The used speech files in the experiment are Arabic and English languages with the following specifications: 16 bits-Raw, mono, with different time duration ranges from 2 to 6 seconds and sampled at 8000 samples per second. The speech files were passed through the encoder, and then frame loss and zero padding methods were performed separately for each speech file. The degraded encoded frames were then decoded. This experiment was done with EVS and AMR coders using the objective evaluation PESQ p.862 to evaluate the quality of original and degraded samples. The bit-rate used with EVS and AMR coders is 8 kbps and 7.95 kbps respectively. Figure (2) shows the block diagram of the experiment.

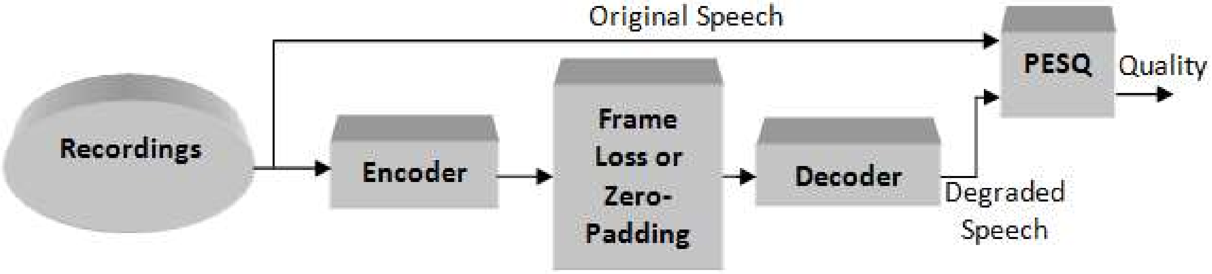


Fig 2: Experiment Block Diagram.

4. RESULTS

In the first part of the experiment, the zero-padding method was performed by inserting zeroes instead of frame(s) content continuously from one frame to forty frames; this was done with 20 different sentences for each speaker; one male and one female in each language. Then, the average quality was calculated for all 20 sentences. This method was then repeated intermittently instead of continuously. The next part of this experiment was the frame loss method which was performed by dropping frames and assembling the rest with the same procedures as the zero-padding method, except that the number of sentences for each speaker was 5 different sentences. Figures (3) and (4) show the average quality against the number of zero padding frames and frame losses, respectively with AMR and EVS coders for an American male.

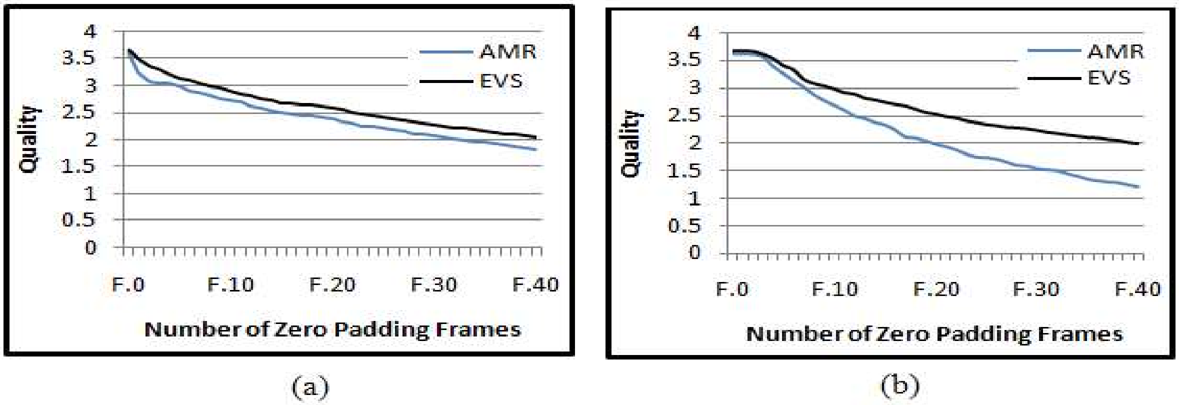


Fig 3: Average Quality of EVS and AMR codec, in case
(a) Continuous and (b) Discontinuous Zero Padding for an American Male.

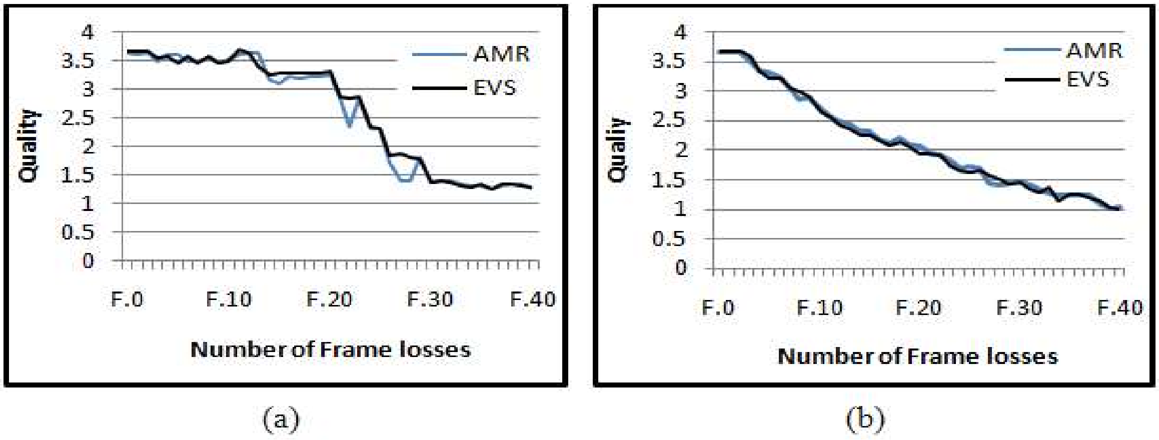


Fig 4: Average Quality of EVS and AMR codec, in case
(a) Continuous and (b) Discontinuous Frame Loss for an American Male.

Figures (5) and (6) show the average quality against the number of zero padding frames and frame losses, respectively with both coders for an American female.

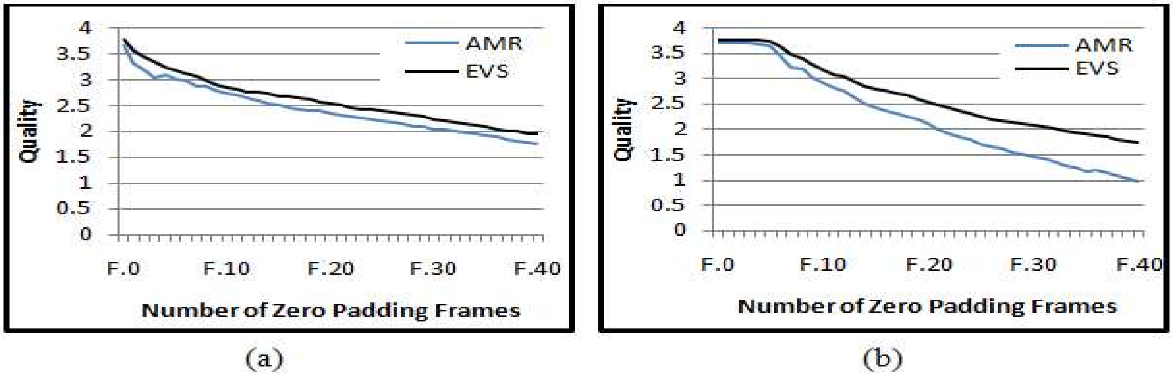


Fig 5: Average Quality of EVS and AMR codec, in case
(a) Continuous and (b) Discontinuous Zero Padding for an American Female.

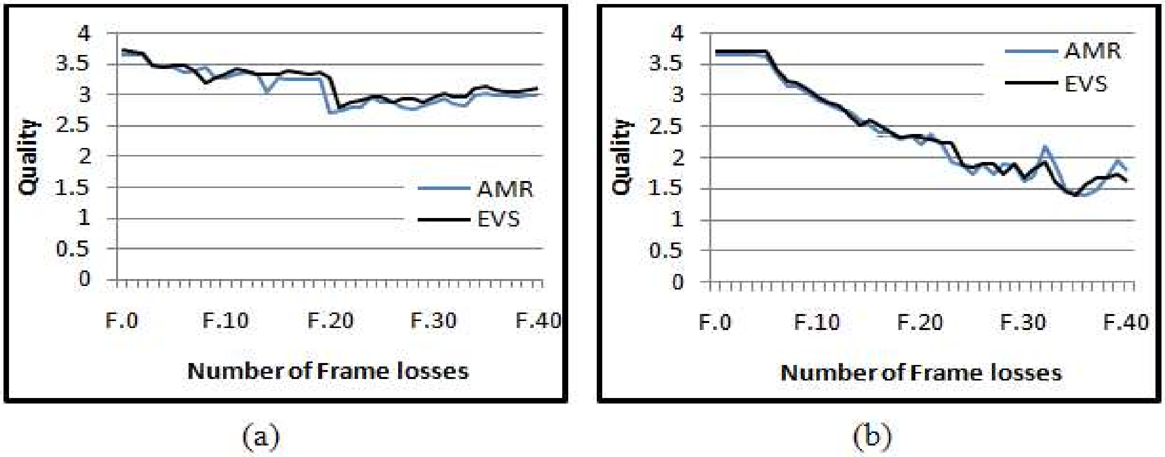


Fig 6: Average Quality of EVS and AMR codec, in case
(a) Continuous and (b) Discontinuous Frame Loss for an American Female.

Figures (7) and (8) show the average quality against the number of zero padding frames and frame losses, respectively with both coders for an Arabic male.

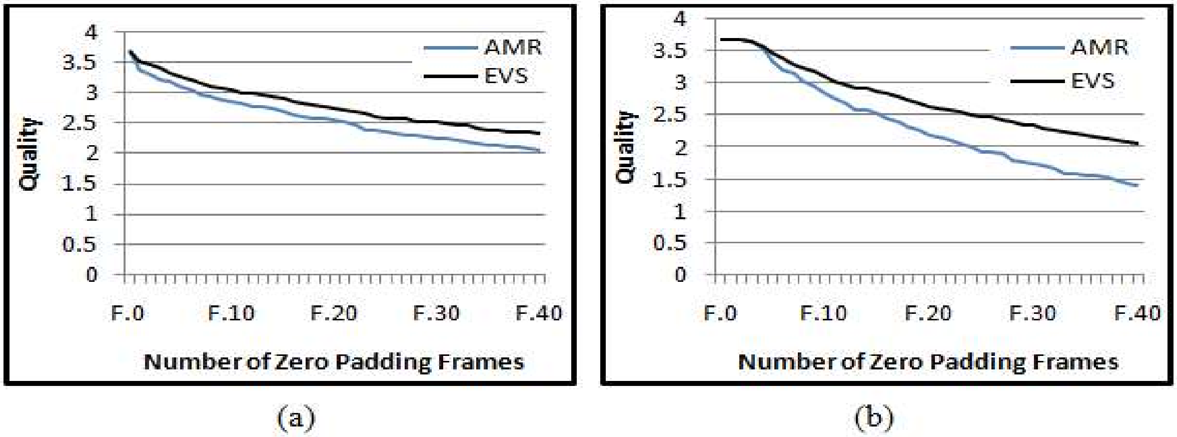


Fig 7: Average Quality of EVS and AMR codec, in case
(a) Continuous and (b) Discontinuous Zero Padding for an Arabic Male.

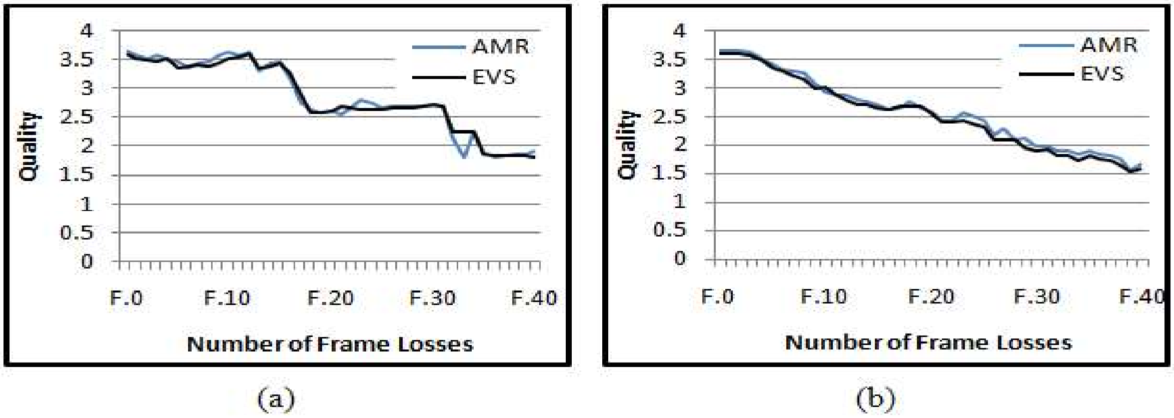


Fig 8: Average Quality of EVS and AMR codec, in case
(a) Continuous and (b) Discontinuous Frame Loss for an Arabic Male.

Figures (9) and (10) show the average quality against the number of zero padding frames and frame losses, respectively with both coders for an Arabic female.

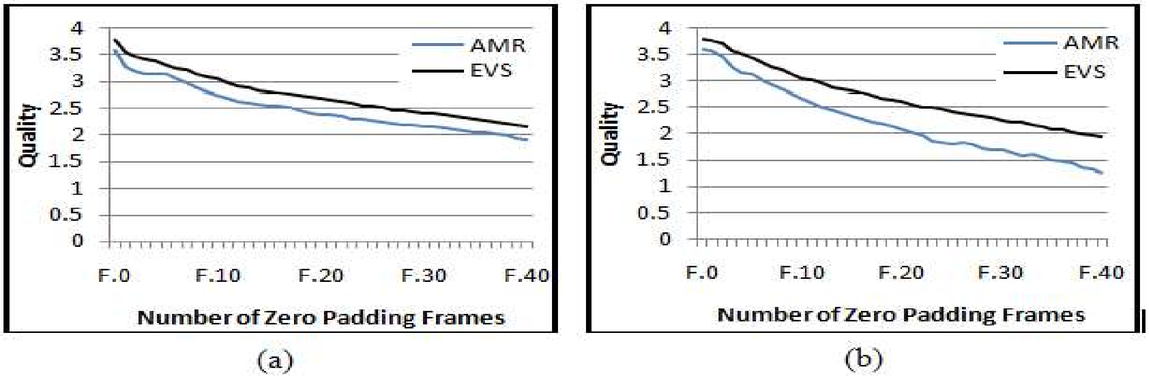


Fig 9: Average Quality of EVS and AMR codec, in case
(a) Continuous and (b) Discontinuous Zero Padding for an Arabic Female.

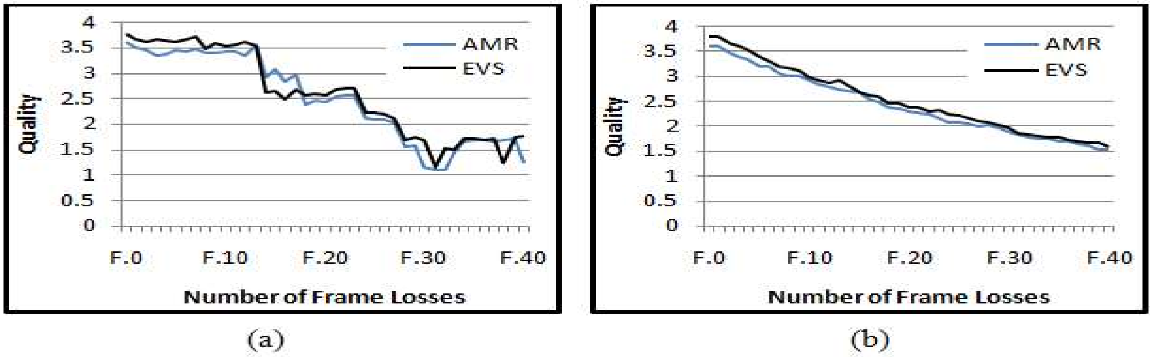


Fig 10: Average Quality of EVS and AMR codec, in case
(a) Continuous and (b) Discontinuous Frame Loss for an Arabic Female.

As shown from figures (3), (5), (7), and (9) the quality decreases with the increasing number of zero padding frames for both coders, and EVS is better than AMR for every point of zero padding frames. The deviation of quality for AMR codec increases with the increasing number of discontinuous zero padding frames, while EVS maintains its quality in case of discontinuous compared to continuous zero padding.

As shown from figures (4), (6), (8), and (10) the average qualities for both coders are relatively close to each other in case of continuous and discontinuous frame losses along with the number of frame losses axis. The average quality in case of discontinuous frame loss decreases sharply with the increasing number of frame losses.

Figures (11) to (14) show the allover average quality and standard deviation in case of inserting zeros or losing frames, where the frames were conducted continuously or intermittently for an American male, American female, Arabic male, and Arabic female, respectively. The calculated standard deviation represents the stability of the codec; where the stability increases by decreasing standard deviation and vice versa.

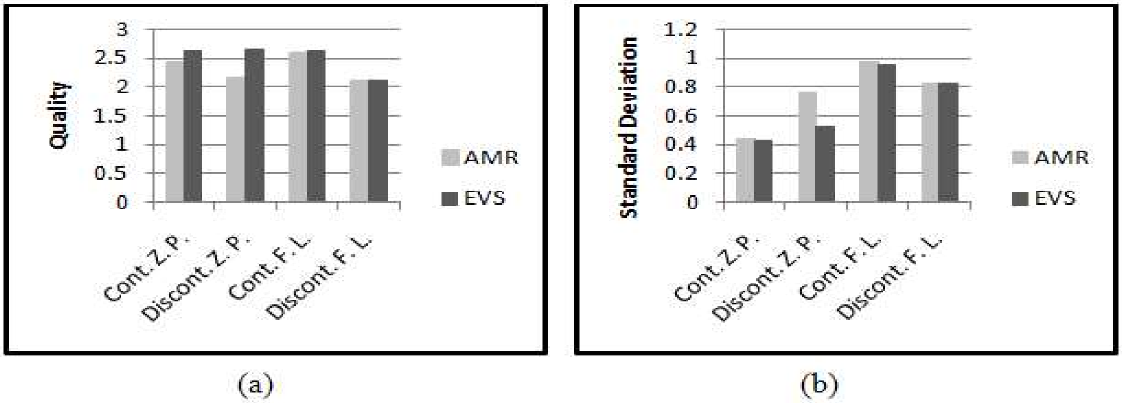
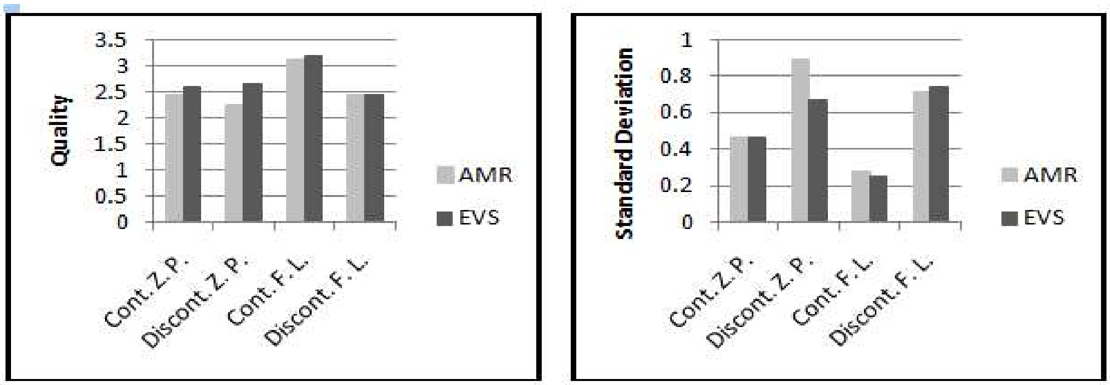
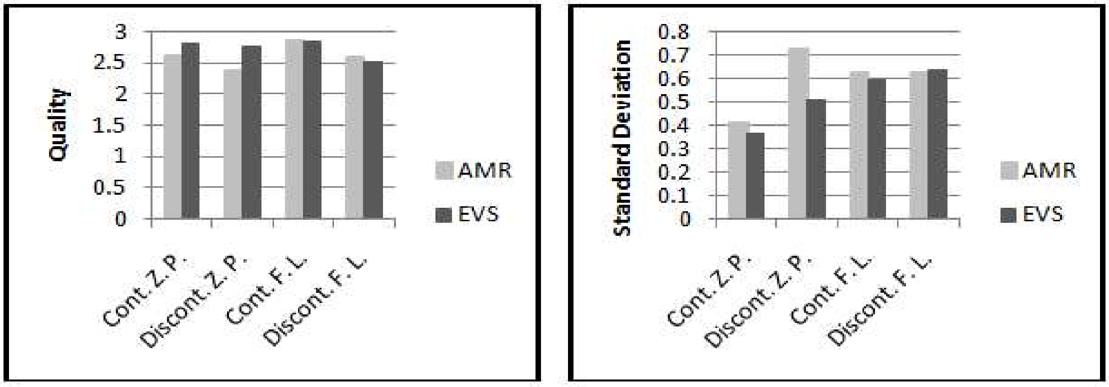


Fig 11: Compared EVS with AMR codec, in terms of (a) Average Quality and (b) Standard Deviation for an American Male.



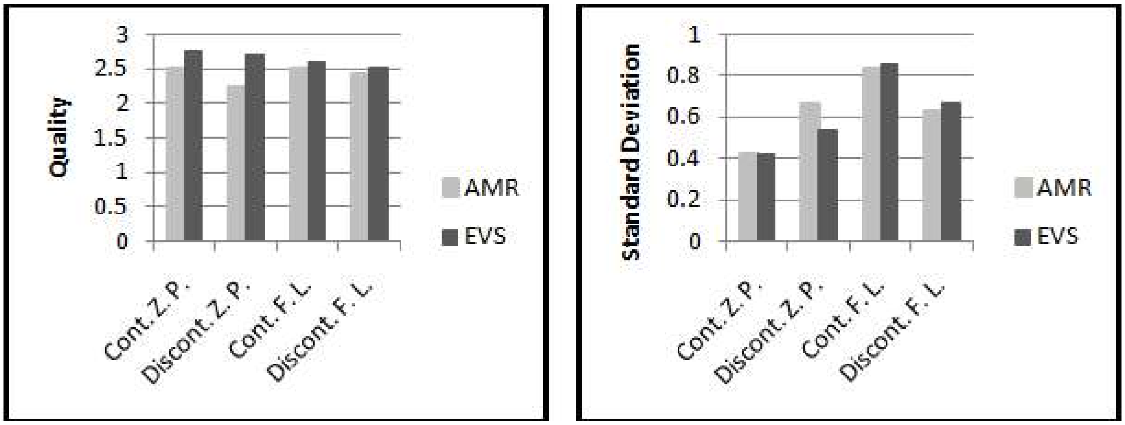
(a) (b)

Fig 12: Compared EVS with AMR codec, in terms of (a) Average Quality and (b) Standard Deviation for an American Female.



(a) (b)

Fig 13: Compared EVS with AMR codec, in terms of
(a) Average Quality and (b) Standard Deviation for an Arabic Male.



(a) (b)

Fig 14: Compared EVS with AMR codec, in terms of (a) Average Quality and (b) Standard Deviation for an Arabic Female.

Figures (11), (12), (13), and (14) show that in case of continuous zero padding frames; EVS codec is slightly better than AMR by 8%, and both coders are relatively equal in stability with all speech samples except with an Arabic male EVS which is more stable than AMR by 11%.

In case of discontinuous zero padding, EVS is better than AMR in terms of average quality and stability by 18% and 30%, respectively. On the other hand, in case of continuous or discontinuous frame loss, both coders are relatively close to each other by 2% and 4% in terms of average quality and stability, respectively.

5. CONCLUSION

By testing the performance of EVS codec compared with AMR codec under errors through losing frames or inserting zeros instead of frame content continuously and intermittently, it was shown that EVS outperform AMR codec in terms of quality in case of zero padding frame. In case of discontinuous zero padding, AMR codec deviated in quality by increasing the number of zero padding frames more than EVS, which proves that EVS achieves better performance in quality than AMR. On the other hand, in case of losing frames, both coders are relatively close to each other, in terms of average quality and stability.

REFERENCES

[1] 3GPP, TR 26.952, “Codec for Enhanced Voice Services (EVS); Performance characterization (Release 13),” 2016.

[2] Anssi Ramo, Antti Kurittu, and Henri Toukomaa, “EVS Channel Aware Mode Robustness to Frame Erasures,” Interspeech 2016, At San Francisco, CA, USA, September 2016.

[3] 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.

[4] T Vaillancourt, V Malenovsky, and R salami, “ADVANCES IN LOW BITRATE TIME-FREQUENCY CODING,” in Acoustics, Speech, and Signal Processing (ICASSP), 2015.

[5] Jérémie Lecomte, Tommy Vaillancourt, and Stefan Bruhn, "PACKET-LOSS CONCEALMENT TECHNOLOGY ADVANCES IN EVS," in Acoustics, Speech, and Signal Processing (ICASSP), 2015.

[6] Chapter 12; AMR speech codecs: operation and performance, 12 Aug. 2009.

[7] Technical Documentation; Experimental AMR Codec; Codec Pro, Version 1.0, Revision A, Nov. 2013.

[8] 3GPP Spec., Adaptive Multi-Rate (AMR) speech codec; Error concealment of lost frames, TS 26.091, v.14.0.0, April. 2017.

[9] 3GPP TR 26.952, Codec for Enhanced Voice Services (EVS); Performance characterization, version 13.1.0 Release 13, April 2016.

[10] 3GPP Spec., Codec for Enhanced Voice Services (EVS); ANSI C code (fixed-point), TS 26.442, v.14.0.0, Mar. 2017.

[11] 3GPP Spec., Codec for Enhanced Voice Services (EVS); Error concealment of lost packets, TS 26.447, v.14.1.0, Jul. 2017.

[12] Zdenek Becvar, Michal Vondra and Lukas Novak (2011). "Assessment of Speech Quality in VoIP", VoIP Technologies, Dr. Shigeru Kashihara (Ed.), ISBN: 978-953-307-549-5, InTech, Available from http://www.intechopen.com/books/voip-technologies/assessment-of-speech-quality-in-voip.

[13] “The PESQ Algorithm as the Solution for Speech Quality Evaluation on 2.5G and 3G Networks”, Technical paper. Ascom Comp., 2009.

[14] “Predicting Performance of PESQ in Case of Single Frame Losses”, Prague, CZ, June 2004 MESAQIN 2004 -Measurement of Speech and Audio Quality in Networks.