# Performance Evaluation of ACELP CODECs against the Compression Ratio and the Change in the Spoken Language and Accent

**Michael N. Micheal[1]   Nagy W. Messiha[2]**

[1]*Faculty of Engineering, Benha University, Cairo, Egypt, michaelnasief@yahoo.com*

[2]*Faculty of Electronic Engineering, Minofia University, Cairo, Egypt, dr.nagy_wadie@hotmail.com*

**Hala A. Mansour[3]**

[3]*Faculty of Engineering, Benha University, Cairo, Egypt, hala.mansour@gmail.com*

## ABSTRACT

*Speech coding is the process of digitizing the voice in a few bits as possible. Parametric Coding is the encoding technique at which the parameters of the speech signal will be extracted frame by frame during the encoding process. Code Excited Linear Prediction "CELP" is one major type of that category of CODECs. The comparison between the CODECs is based upon the band width, the speech quality and the associated CODEC delay. It has been proven that the quality is inversely proportional to the compression ratio. But what will happen if the speech is compressed more and at the same time the spoken language changed? This paper is made to answer about this question. In this paper an extensive testing is done on ITU G.723.1 (5.3 kbps) and 3GPP Adaptive Multi Rate "AMR" (12.2 kbps). The test is made using 3 groups of speech samples (English, Arabic and Cairo accent) with different speakers' gender. The evaluation method used is the ITU PESQ algorithm. It has been proved that the speech quality dramatically decreases as the compression ratio increases, which will be rather lower and unstable for Arabic and Cairo accent over that of the English language.*

*Keywords: performance, coders, accents, Arabic, Cairo accent, G.723.1, AMR, PESQ.*

## 1. Introduction

Most of the available speech CODECs today are always sampled at 8000 sample per second. The comparison between the CODECs is based on many factors including bit rate, the speech quality, complexity, memory used and the associated CODEC delay. The relation between the quality and the band width is inversely proportional. Also the relation between the speech quality and the delay is inversely proportional [1]. The following table summarizes the comparison between two of the most used CODECs (i.e. G.723.1 and AMR) while the complex relation between the CODEC quality, bandwidth and coding delay for many other CODECs is given in the figure (1).

Table 1: The Characteristics of the G.723.1 and AMR CODECs

| CODEC | G.723.1 | AMR |
|---|---|---|
| **Bit Rate** | 5.3/6.4 Kbps | 12.2 to 4.75 Kbps |
| **Delay** | 30 ms | 20 ms |
| **Frame Length** | 20/24 bytes | 20 bytes |

Another two parameters with less significantly related to the CODEC complexity are the DSP (Digital Signal Processor) MIPS (Mega Instruction per Second) and RAM (Random Access Memory). The DSP MIPS, the digital signal processor mega instructions per second is the minimum speed of processor to be used for that CODEC. While the RAM is the size of the minimum required random access memory to hold data during the processing. Both are directly related to the CODEC production cost.

The previous work in this field shows that, problems will occur when the speaker uses a language or accent that contains phonemes that inherent to the speaker mother language and not contained in English language [2], [3]. Also, I.S. Burnett and J.J. Parrry worked on "The Effects of Accent and Language on Low Rate Speech Coders" [4].
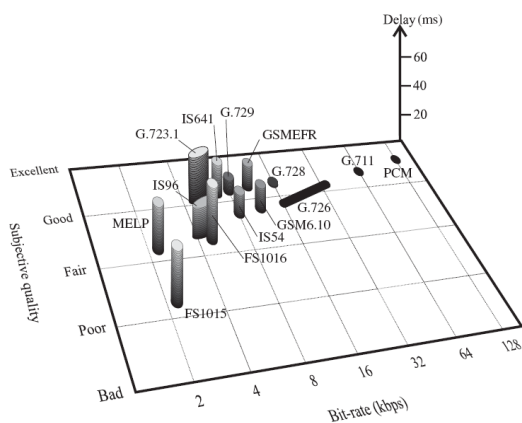


Fig. 1: Performance comparison between some standardized coders [1].

They studied the problems with accented speech and languages with low rate coders and concluded that "when the coders are used for heavily accented English or other language, significant performance degradation is noted". While, Mohamed Itani and Sarunas Paulikas were published two papers [5]-[6] on the "Influence of Language on CELP CODECs performance".. These papers which are about the influence of language on LPC performance work only on Speexs and AMR Coders and with very small

number of speech samples to get the results for the English, Arabic and Lithuanian.

## 2. Code Excited Linear Prediction.

The code excited linear prediction which is a speech coder algorithm was first proposed in 1985,. It is the basic principle of other CODEC algorithms like ACELP "Algebraic CELP" and CS-CELP "Conjugate Structure CELP". The CELP is based on three main ideas [7] as shown in figure (2):

- Linear prediction model (LP).
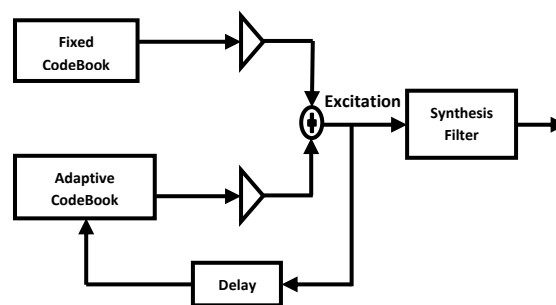- The use of the adaptive and fixed codebook to excite the LP model.
- Closed loop search.



Fig. 2: The CELP model of speech synthesis [8].

The main principle of the CELP is the "Analysis by Synthesis". "Meaning that the encoding (analysis) is performed by perceptually optimizing the decoded (synthesis) signal in a closed loop"[8]. Some methods use a comparison between many streams and choose the best one. The ACELP uses a specific algebraic structure for its codebook.

## 3. The PESQ Algorithm.

For an original speech signal X(t) which is passed through a communication system and the output from that system Y(t), the PESQ algorithm compares X(t) with Y(t) and evaluates the equivalent MOS (Mean Opinion Score) as in figure (2).
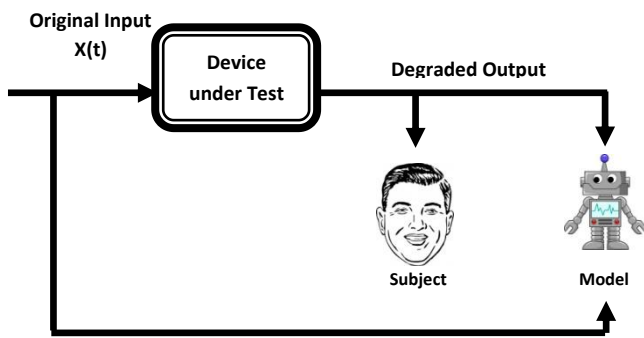
Fig. 3: PESQ basic philosophy.

The first step of the PESQ algorithm is to use a series of delays between the original signal and the degraded one to get the actual time delay between both. This step is called the time alignment algorithm. The algorithm can handle delay changes during both silence and active speech. Based on the set of delays that are found the PESQ compares the original signal with the delayed one using a computer model as illustrated in the figure (3) [9].

A computer model replaces the subject (human). This model actually consists of two models. The first one is the perceptual model that responsible for extracting the speech parameters and the second one is the cognitive one that makes the actual judgement. These models are used to compare the input and the output of the device under test which will be the G.723.1 and AMR CODECs which are considered in our case.
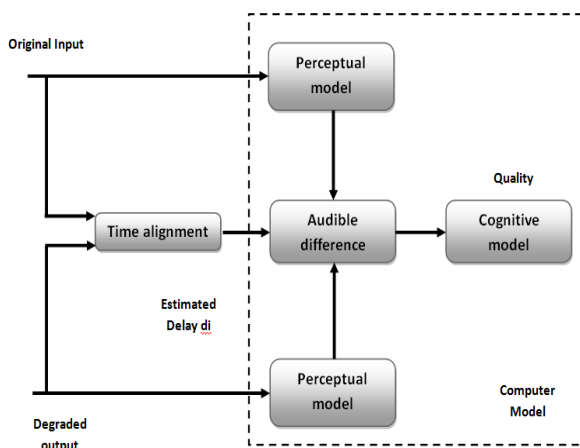


Fig. 4: the Basic block diagram of the PESQ

algorithm

## 4. Sample of the Most Used CODECs.

### A. AMR "Adaptive Multi Rate".

**Adaptive Multi-Rate (AMR)** is a speech coding developed as the standard speech codec by 3GPP for the 2$^{nd}$ and the 3ed mobile generations [10]. The principle of the AMR CODEC is to use similar algorithms of many CODECs with different data rates. The communication network is responsible for the selection between these rates according to the channel robustness. For highly noise channel the lower rate will be used in order to increase the number of redundancy bits, so maintain the connection with acceptable speech quality. For good channel (i.e. free of errors) AMR full rate at 12.2 kbps will be used giving the best speech quality. The following table shows the different AMR rates.

Table 2: AMR Different Rates [10]

| CODEC | Bit Rate |
|---|---|
| AMR Full Rate | 12.2 kbps |
| AMR 10.2 | 10.2 Kbps |
| AMR 7.95 | 7.95 kbps |
| AMR 6.7 | 6.7 Kbps |
| AMR 5.9 | 5.9 Kbps |
| AMR 5.15 | 5.15 Kbps |
| AMR NB | 4.75 kbps |

The AMR CODEC is an algebraic code excited linear prediction (ACELP) CODEC. It uses 10th order short term linear prediction filters and both algebraic and adaptive codebooks. Figure (4) shows the block diagram of the AMR coder.
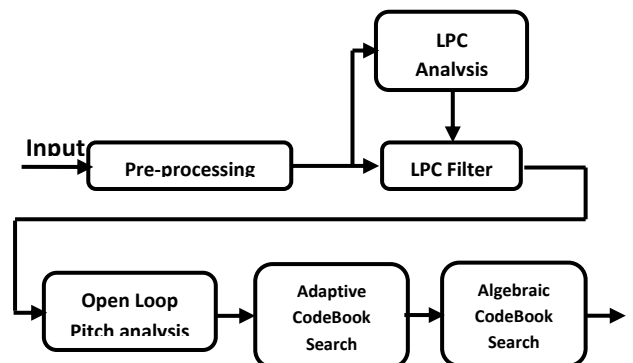


Fig. 5: the Basic block diagram of the AMR Coder

[11]

The first stage is the pre-processing stage it consists of a high pass filter and a down scaling to reduce the system complexity. Second stage is the 10th order linear prediction analysis it is done twice per frame for the full rate at 12.2 Kbps. Another parameter to be extracted is the pitch of the sound; this is done using the open loop pitch estimation. The open loop pitch estimation is performed each half frame. The final two stages consist of the adaptive and algebraic codebooks.

For AMR rates lower than 12.2 Kbps algorithms will be the same except some minor computations. For example for the AMR 5.15 Kbps the linear prediction analysis will be done once per frame not twice [11].

### B. The G.723.1 CODEC

The ITU G.723.1 is one of the most popular LPC CODECs. It can work in either one of two modes, the MPC-MLQ (Multipulse LPC with Maximum Likelihood quantization) with a rate of 6.3 Kbps and the ACELP (Algebraic code excited linear prediction) with 5.3 Kbps. Each frame contains 240 samples for 30 ms duration [12] [13]. Figure (5) shows the Coder block diagram while the later one illustrates the Decoder.
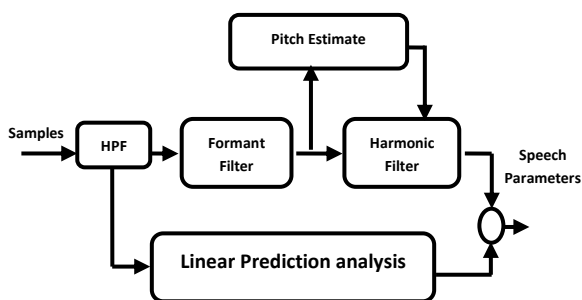


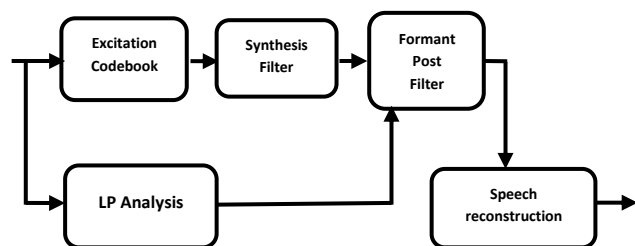Fig. 6: (a) G.723.1 Encoder.



Fig. 6: (b) G.723.1 Decoder.

The coder operates on two time scales (i.e. frame scale 30 ms, 240 samples and sub-frame scale 7.5 ms, 60 samples). The coding process starts from the high pass filter (HPF) to remove the DC component. Then the Linear Prediction analysis is done on each sub-frame this creates 4 groups of LP coefficients. The output from the HPF is also fed to the formant filter to extract the filter coefficients of the speech formants. The output from the formant filter is used to estimate the pitch of the sound. Both of the outputs from the formant filter and the pitch estimation will be used to get the harmonics associated with the speech using the harmonic weighting filter.

The decoder on the other hand must retrieve the original speech. It also works on frame by frame basis. This will be done in four main steps. The speech parameters received from the coder will fed to both the LP analysis to reproduce the filters from the coefficients that have been sent after decoding them. The excitation code book is searched to generate the excitation signal. This signal then is passed through the synthesis filter whose output is input to the formant post filter. The final gain scaling to maintain the energy level of the original signal will be adapted in the final step of the speech reconstruction.

## 5. The Test Strategy and the Results Analysis.

### A. The Test Strategy

The test strategy can be summarized as the following. First we saved 10 voice files for English, Arabic and Cairo accent each 10 seconds long [with the same person]. In the second step we use the same files to generate the output of the AMR Full rate 12.2 Kbps using standard Ericson tool and ACELP G.723.1 implemented using MATLAB. While the third step is the using of the PESQ algorithm to estimate the quality, and the final step is to tabulate the results and conclude the effects.

### B. Results analysis

We use more than 200 of speech files that are recorded at the same testing environment for this experiment. As stated in the test strategy every speech file is passed through a coder and decoder of the same CODEC (i.e. AMR or G.723.1) and

tested for quality using the PESQ algorithm. In this case the experiment will evaluate the effect of the CODEC only since there is no channel included.

The primary point for the selection of the AMR and the G.723.1 for out experiment is that, both are lying on the same principle of ACELP, however AMR is a 3GPP standard while the G.723.1 is ITU standard. Another important point is that the AMR used here is the full rate version at 12.2 Kbps while G.723.1 is used at 5.3 Kbps, so in this way we can see the effect of the bit rate reduction on the speech quality.

In order to estimate the effect of the change of the spoken language or accent we select English, Arabic and Cairo accent to compare the PESQ values from them and conclude the results. The output results of PESQ score versus speech sample from AMR 12.2 Kbps and G.723.1 5.3 Kbps for English, Arabic and Cairo accent are shown in the figures and tables below.

Table 3: PESQ Values for English with G.723.1 and AMR CODECs

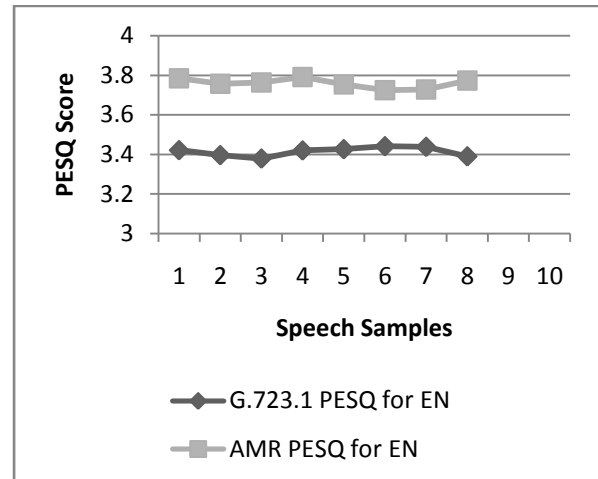| G.723.1 PESQ for EN | AMR PESQ for EN |
|---|---|
| 3.395 | 3.7799 |
| 3.4025 | 3.8234 |
| 3.4209 | 3.7834 |
| 3.396 | 3.7567 |
| 3.3787 | 3.7626 |
| 3.4194 | 3.7903 |
| 3.4259 | 3.7533 |
| 3.4407 | 3.724 |
| 3.4376 | 3.7276 |
| 3.3897 | 3.7732 |



Fig 7: PESQ Curves for English with G.723.1 and AMR CODECs

Table 4: PESQ Values for Arabic with G.723.1 and AMR CODECs

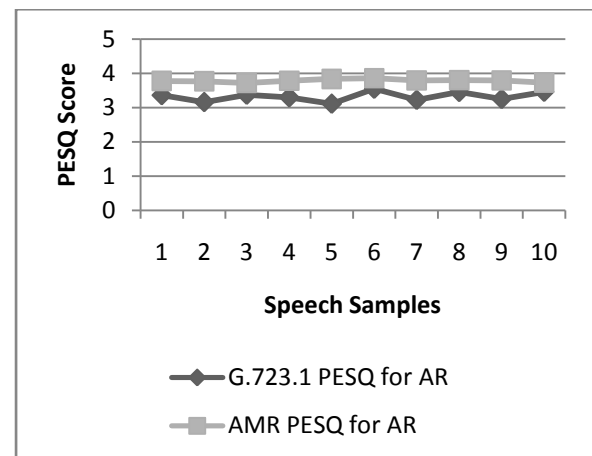| G.723.1 PESQ for AR | AMR PESQ for AR |
|---|---|
| 3.3583 | 3.7745 |
| 3.1569 | 3.7658 |
| 3.3704 | 3.7179 |
| 3.2962 | 3.7814 |
| 3.1093 | 3.8311 |
| 3.5436 | 3.8534 |
| 3.221 | 3.7897 |
| 3.4499 | 3.7994 |
| 3.2482 | 3.7894 |
| 3.4556 | 3.7301 |



Fig 8: PESQ Curves for Arabic with G.723.1 and AMR CODECs

Table 5: PESQ Values for Cairo accent with G.723.1 and AMR CODECs

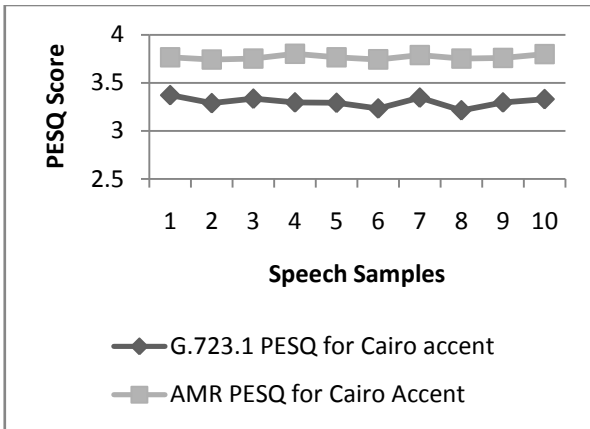| G.723.1 PESQ for Cairo Accent | AMR PESQ for Cairo Accent |
|---|---|
| 3.3715 | 3.7663 |
| 3.2888 | 3.7434 |
| 3.335 | 3.7544 |
| 3.2966 | 3.803 |
| 3.2924 | 3.7673 |
| 3.2338 | 3.7446 |
| 3.346 | 3.7902 |
| 3.215 | 3.7534 |
| 3.2963 | 3.7609 |
| 3.3306 | 3.7992 |



Fig 9: PESQ Curves for Cairo accent with G.723.1 and AMR CODECs

Table 6: PESQ Values for Cairo accent, English with G.723.1 and English with AMR CODECs

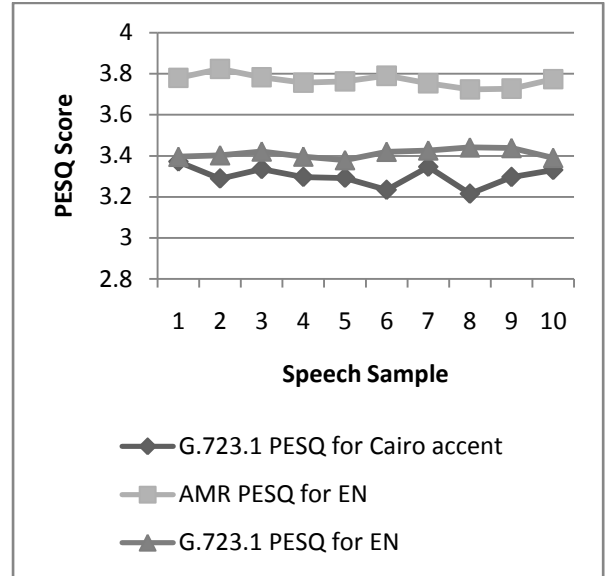| G.723.1 PESQ for Cairo accent | AMR PESQ for EN | G.723.1 PESQ for EN |
|---|---|---|
| 3.3715 | 3.7799 | 3.395 |
| 3.2888 | 3.8234 | 3.4025 |
| 3.335 | 3.7834 | 3.4209 |
| 3.2966 | 3.7567 | 3.396 |
| 3.2924 | 3.7626 | 3.3787 |
| 3.2338 | 3.7903 | 3.4194 |
| 3.346 | 3.7533 | 3.4259 |
| 3.215 | 3.724 | 3.4407 |
| 3.2963 | 3.7276 | 3.4376 |
| 3.3306 | 3.7732 | 3.3897 |



Fig 10: PESQ Curves for Cairo accent, English with G.723.1 and English with AMR CODECs

For better estimation of the language or accent change we will use the standard deviation statistic value. The standard deviation shows how much dispersion exists from the mean value. Low standard deviations indicate the data values are very close to the average which means better signal stability and vice versa. Table (7) shows the standard deviation for the tabulated above samples.

Table 7: The Standard deviation values for the PESQ outputs for different languages.

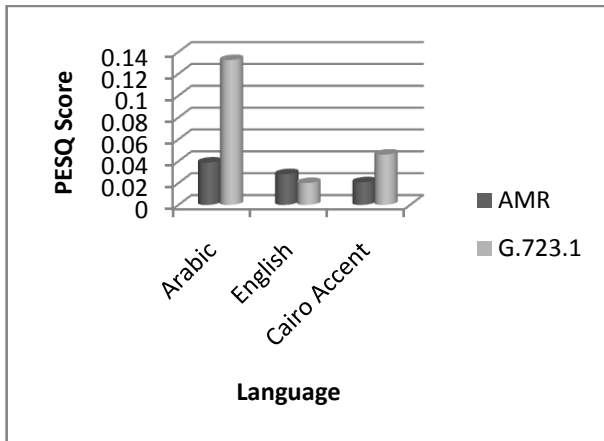| | AMR | G.723.1 |
|---|---|---|
| Arabic | 0.03877 | 0.13258 |
| English | 0.028125 | 0.020073 |
| Cairo Accent | 0.020725 | 0.046046 |

Fig 11: The Standard deviation values for the PESQ outputs for different languages.

## 6. The Conclusion.

In this paper two issues have been proved. The first one is the great reduction in the speech quality as the bit rate reduced or in other word when the speech is compressed more, which is not a new conclusion for the CODECs researchers, however the concerned point is the dramatically effect of changing the language from English to either Arabic or Cairo accent.That effect appears both in PESQ scores and the change in the standard deviation which means less stability of the coder performance with language other than English when using the lower rate CODECs like G.723.1. So the final conclusion will be that the more compression of the speech using ACELP CODECS the less the speech quality, the more compression with language or accents other than English the less the quality and more instability of the CODEC performance.

## REFERENCES

1.  WAI C. CHU," *Speech Coding Algorithms Foundation and Evolution of Standardized Coders*", 2003 by John Wiley & Sons, Inc.

2.  J.H.L. Hansen, L.M. Arslan, "*Foreign Accent Classification Using Source Generator Based Prosodic Features*",
Proc. Int. Conf. Acoust. Speech Sign. Process. Detroit, pp. 836-839, 1995.

3.  J.J. Parry, "*Accent Classification for Speech Coding*", Honours thesis, The University of Wollongong, 1995.

4.  I. S. Burnett and J.J.Parry ," *On the Effects of Accent and Language on Low Rate Speech Coders* ".

5.  Mohamad Itani, Šarūnas Paulikas, "Influence *of Language on CELP CODECS Performance*", ISSN 1392 – 124X INFORMATION TECHNOLOGY AND CONTROL, 2008, Vol.37, No.2.

6.  Mohamad Itani, Šarūnas Paulikas, "LITHUANIAN SPEECH RECORDS DATABASE FOR VOICE CODECS QUALITY ASSESSMEENT", ISSN 1392 – 124X INFORMATION TECHNOLOGY AND CONTROL, 2010, Vol.39, No.1

7.  B.S. Atal, "*The History of Linear Prediction*," *IEEE Signal Processing Magazine*, vol. 23, no. 2, March 2006, pp. 154–161.

8.  M. R. Schroeder and B. S. Atal, "*Code-excited linear prediction (CELP): high-quality speech at very low bit rates*", in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*

9.  ITU – T Recommendation for PESQ P.862 -1993.

10. 3GPP. "3GPP TS 26.090 - Mandatory Speech Codec speech processing functions; Adaptive Multi-Rate (AMR) speech codec; Transcoding functions". Retrieved 2010-07-21.

11. 3GPP (2008-12-11) 3GPP TS 26.073 - AMR speech Codec, Retrieved 2009-09-08.

12. ITU – T Recommendation G.723.1 -2009.

13. P. Kabal, ITU-T G.723.1 Speech Coder: "*A Matlab Implementation, Department of Electrical & Computer Engineering McGill University*", 2009.