International Journal of Engineering Sciences & Emerging Technologies, August 2013.
ISSN: 2231 – 6604 Volume 6, Issue 1, pp: 1-8 ©IJESET

THE EFFECT OF WEIGHT FACTOR ON THE PERFORMANCE OF G.729A SPEECH CODER

M. A. Rashed[1], T. A. El-Garf[1], I. F. Tarrad[2], H. A. Almotaafy[1]

[1] Department of Electrical & Computer Engineering, Higher Technological Institute, 10th of Ramadan City, Egypt.
[2] Department of Electronics & Communications Engineering, Al-Azhar University, Cairo, Egypt.
mohammed.adel@windowslive.com
talaatelgarf@yahoo.com
tarradif@gmail.com
Prdr_hosny_aaam@yahoo.com

ABSTRACT

G.729A (CS-ACELP Annex A) is a high quality low bandwidth codec at 8 kbit/s with low complexity. G.729A is a bit stream interoperable with G.729; i.e., speech coded with G.729A can be decoded with G.729, and vice versa. Several algorithmic changes have been introduced into G.729 which resulted in 50% drop in its complexity. The Aim of this paper is to study the effect of weight factor on the performance of G.729A coder synthesized speech. This coder is implemented in MATLAB with many different values of weight factor. The effect of weight factor on the recovered speech quality has been investigated and studied. Objective analysis on weighted speech has been carried out in order to observe its overall performance. The performance measures prove that at certain values of weight factor, the speech quality is better than at the value listed in the standard.

KEYWORDS: *Speech Coding, CS-ACELP, G.729, G.729A, Analysis-by-synthesis, Weighted speech, Speech synthesis.*

I. INTRODUCTION

ITU-T has standardized G.729 as the standard speech-coding algorithm for VoIP, DSVD (Digital Simultaneous Voice over Data) and multimedia applications [1-3]. This ITU-T recommendation is based on Conjugate Structure-Adaptive Code Excited Linear Prediction (CS-ACELP) algorithm, operating at a bit-rate of 8 Kbps for discrete speech samples sampled at a rate of 8000 samples per second. Later, it has led to the design of various annexes of standard G.729 with variable bit rate and enhanced speech quality. The G.729A is a low complexity continuous data transmission scheme for DSVD applications [3-5,13].

In this paper, we have focused on the CS-ACELP codec G.729A (8kbps) and the effect of different values of weight factor γ on the speech quality. The rest of the paper is organized as follows. The description of G.729A (Principles of Encoder and Decode) and the algorithmic changes with G729 standard will be presented in Section 2. Proposed Modification in Weight Factor of G.729A will be explained in Section 3. Objective Analysis which used in quality measurements will be presented in Section 4. Section 5 will show the experimental results. The conclusion will be given in the last section.

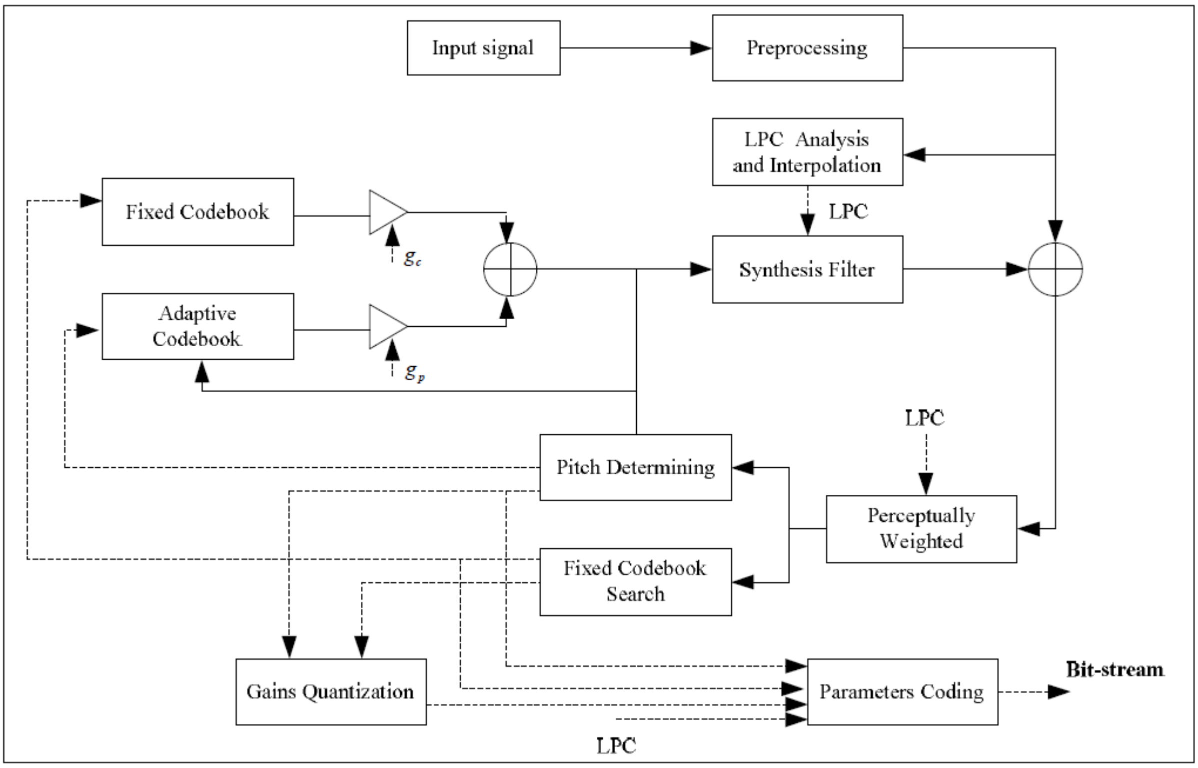

Figure 1. Principle of the CS-ACELP encoder in G.729A

II. DESCRIPTION OF G.729A

The general description of the coding/decoding algorithm of G.729A is similar to that of G.729 [1, 2, and 13]. The same conjugate- structure algebraic code-excited linear-predictive (CS-ACELP) coding concept is used. The coder operates on speech frames of 10 ms corresponding to 80 samples at a sampling rate of 8000 samples/s. For every 10 ms frame, the speech signal is analyzed to extract the parameters of the CELP model (linear prediction filter coefficients, adaptive and fixed codebook indices and gains). These parameters are encoded and transmitted. At the decoder, these parameters are used to retrieve the excitation and synthesis filter parameters. The speech is reconstructed by filtering this excitation through the short-term synthesis filter. The long-term or pitch synthesis filter is implemented using the so-called adaptive codebook approach. After computing the reconstructed speech, it is further enhanced by a postfilter. The encoding and decoding principles are further explained.

2.1. Encoder

The encoding principle is shown in Figure 1. The input signal is high-pass filtered and scaled in the pre-processing block. The pre-processed signal serves as the input signal for all subsequent analysis. The 10th-order linear prediction (LP) analysis is done once per 10 ms frame to compute coefficients of the LP filter l/A(z). These coefficients are converted to line spectrum pairs (LSPs) and quantized using predictive two-stage vector quantization (VQ) with 18 bits. The excitation signal is chosen by an analysis-by-synthesis search procedure. In this procedure, the error between the original and reconstructed speech is minimized according to a perceptually weighted distortion measure. This is done by filtering the error signal with a perceptual weighting filter, whose coefficients, unlike G.729, are derived from the quantized LP filter. A weighting filter of the form $W(z) = \hat{A}(z)/\hat{A}(z/\gamma)$ is used, where $\hat{A}(z)$ is the quantized version of A(z) [3,5].
The excitation parameters (fixed and adaptive codebook parameters) are determined for subframes of 5 ms (40 samples) each. The quantized LP filter coefficients are used for the second subframe, while interpolated LP filter coefficients are used in the first subframe. An open-loop pitch delay is estimated once per 10 ms frame based on the perceptually weighted and low-pass-filtered speech signal. Then the following operations are repeated for each subframe. The target signal x(n) is computed by

filtering the LP residual through the weighted synthesis filter l/ Â (z/y). The initial states of this filter are updated by computing the weighted error signal at the end of the subframe. This is equivalent to the common approach of subtracting the zero-input response of the weighted synthesis filter from the weighted speech signal. The impulse response h(n) of the weighted synthesis filter is computed. Closed-loop pitch analysis is then done (to find the adaptive-codebook delay and gain), using the target x(n) and impulse response h(n), by searching around the value of the open-loop pitch delay. A fractional pitch delay with 1/3 resolution is used. The pitch delay is encoded with 8 bits in the first subframe and differentially encoded with 5 bits in the second subframe. The target signal x(n) is updated by subtracting the (filtered) adaptive-codebook contribution, and this new target, x`(n), is used in the fixed codebook search to find the optimum excitation. An algebraic codebook with 17 bits is used for the fixed codebook excitation. The gains of the adaptive and fixed codebook contributions are vector quantized with 7 bits (with moving-average prediction applied to the fixed-codebook gain). Finally, the filter memories are updated using the determined excitation signal.

2.2. Decoder

The decoder principle is shown in Figure 2. First, the parameter indices are extracted from the received bitstream. These indices are decoded to obtain the coder parameters corresponding to a 10 ms speech frame. These parameters are the LSP coefficients, the two fractional pitch delays, the two fixed codebook vectors, and the two sets of adaptive and fixed codebook gains. The LSP coefficients are interpolated and converted to LP filter coefficients for each subframe. Then, for each 5 ms subframe, the following steps are done:

- The excitation is constructed by adding the adaptive and fixed codebook vectors scaled by their respective gains.
- The speech is reconstructed by filtering the excitation through the LP synthesis filter.
- The reconstructed speech signal is passed through a post-processing stage. This includes an adaptive postfilter based on the long-term and short-term synthesis filters, followed by a high-pass filter and scaling operation.

2.3. Description of Algorithmic Changes to G.729

The LP analysis and quantization procedures as well as the joint quantization of the adaptive and fixed codebook gains are the same as G.729 [1,3]. The major algorithmic changes to G.729 are summarized below [1,4]:

- The perceptual weighting filter uses the quantized LP filter parameters and it is given by $W(z) = \hat{A}(z)/\hat{A}(z/\gamma)$ with a fixed value of $\gamma = 0.75$.
- Open-loop pitch analysis is simplified by using decimation while computing the correlations of the weighted speech.
- Computation of the impulse response of the weighted synthesis filter $W(z)/\hat{A}(z)$, computation of the target signal, and updating the filter states are simplified since $W(z)/\hat{A}(z)$ is reduced to $1/\hat{A}(z/\gamma)$.
- The search of the adaptive codebook is simplified. The search maximizes the correlation between the past excitation and the backward filtered target signal (the energy of filtered past excitation is not considered).
- The search of the fixed algebraic codebook is simplified. Instead of the nested-loop focused search, an iterative depth-first tree search approach is used.
- At the decoder, the harmonic postfilter is simplified by using only integer delays.

III. PROPOSED MODIFICATIONS IN WEIGHT FACTOR OF G.729

Unlike G.729, the perceptual weighting filter is based on the quantized LP filter coefficients âi, and is given by:

$$W(z) = \frac{\hat{A}(z)}{\hat{A}(z/\gamma)} \tag{1}$$

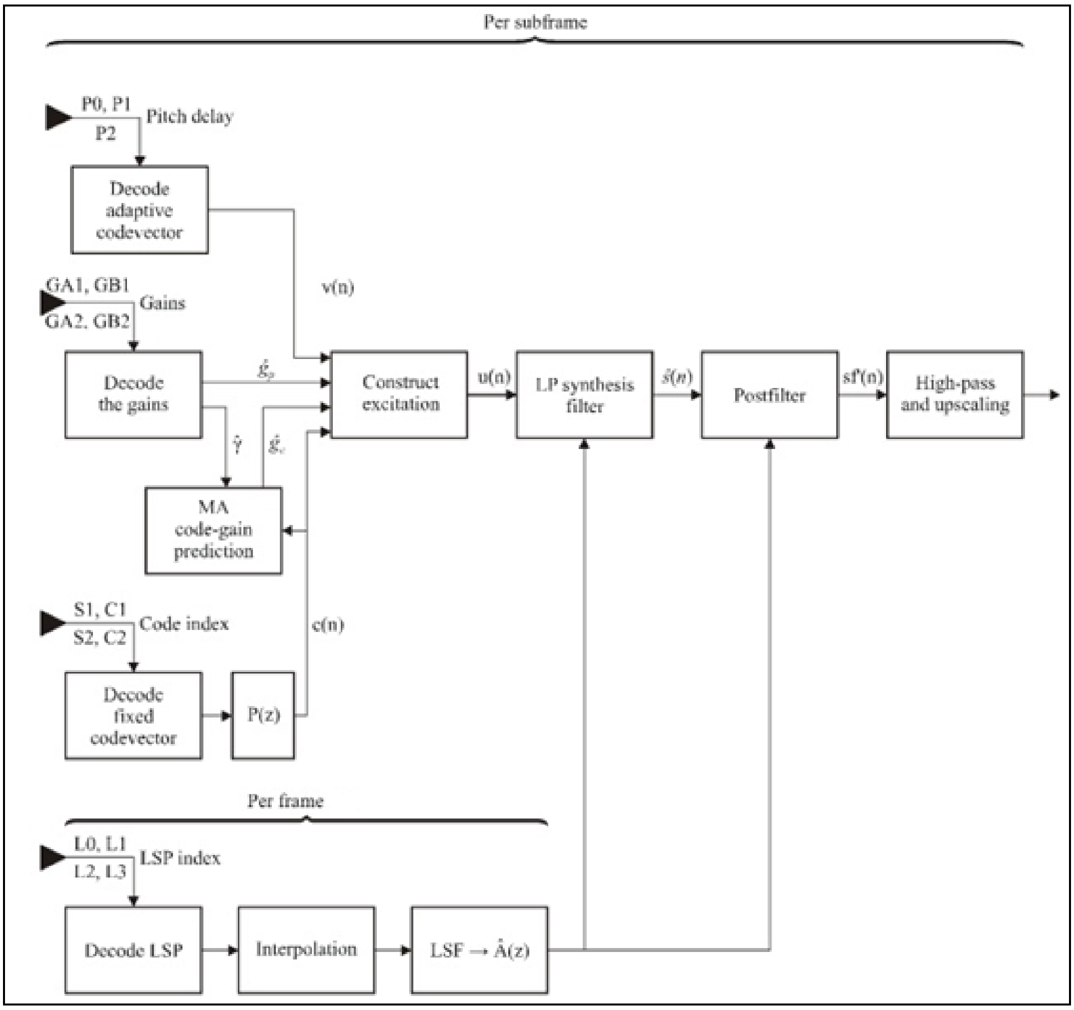

Figure 2. Principle of the CS-ACELP decoder in G.729A

with $\gamma = 0.75$. This simplifies weighting filters to $W(z)/\hat{A}(z) = 1/\hat{A}(z/\gamma)$, which reduces the number of filtering operations for computing the impulse response and the target signal and for updating the filter states. Note that the value of γ is fixed to 0.75 and the procedure for the adaptation of the factors of the perceptual weighting filter described in G.729 is not used in G.729A [1, 15].
The weighted speech signal is not used for computing the target signal since an alternative approach is used. However, the weighted speech signal (low-pass filtered) is used to compute the open-loop pitch estimate. The low-pass filtered weighted speech is found by filtering the speech signal s(n) through the filter $\hat{A}(z)/[\hat{A}(z/\gamma)(1 - 0.7z^{-1})]$. First, the coefficients of the filter $A'(z) = \hat{A}(z/\gamma)(1 - 0.7z^{-1})$ are computed, then the low-pass filtered weighted speech in a subframe is computed by:

$$S_w(n) = r(n) - \sum_{i=1}^{10} a_i' s_w(n-i), \quad n = 0,\ldots,39 \tag{2}$$

where r(n) is the LP residual signal given by:

$$r(n) = s(n) + \sum_{i=1}^{10} \hat{a}_i s(n-i), \quad n = 0,\ldots,39 \tag{3}$$

The signal $S_w(n)$ is used to find an estimation of the pitch delay in the speech frame. The simplification of the weighting filter resulted in some quality degradation in cases of input signals with flat response. In fact, the adaptation of the weighting factors was introduced in G.729 to improve the performance for such signals.
In this paper we measure the quality of reconstructed speech with variation in γ at the encoder in the range of 0.4 to 0.95 with fixed step 0.05. The proposed measurements in several weight factor values

proved that at certain values of weight factor, the speech quality is better than at the value listed in standard as shown in section 5.

IV. OBJECTIVE ANALYSIS

To evaluate the performance of speech quality, the different types of Objective Analysis have been carried out in this paper. Objective Analysis has been categorized into spectral, perceptual and composite measures.

4.1. Perceptual based Objective Analysis

Perceptual Evaluation of Speech Quality (PESQ) is the most important parameter for performing perceptual based analysis. In comparison with other objective measures, the PESQ measure is the most complex to compute and is the one recommended by ITU-T P.862 for speech quality assessment of narrow-band speech codecs [6, 7, 14]. PESQ score is computed as a linear combination of the average disturbance value D_{ind} and the average asymmetrical disturbance values A_{ind} as follows:

$$PESQ = a_0 + a_1 D_{ind} + a_2 A_{ind} \tag{4}$$

Where $a_0 = 4.5$, $a_1 = -0.1$, $a_2 = -0.0309$. A higher value of PESQ indicates better quality of speech.

4.2. Spectral based Objective Analysis

The Weighted Spectral Slope (WSS) is evaluated in this category of Objective Analysis. The WSS distance measure is a direct spectral distance measure. It is based on comparison of smoothed spectra from the clean and distorted speech samples. The smoothed spectra can be obtained from either LP analysis, Cepstrum liftering (a term coined for filtering in the Cepstrum domain), or filter bank analysis. A lower value of WSS indicates better quality of speech. The WSS measure computed as:

$$WSS = \frac{1}{M}\sum_{m=0}^{M-1}\frac{\sum_{j=1}^{10} W(j,m)(S_c(j,m)-S_d(j,m))^2}{\sum_{j=1}^{K} W(j,m)} \tag{5}$$

Where W(j, m) is the weight computed, S_c(j, m) and S_d(j, m) are the spectral slopes of the m^{th} frame of enhanced speech signals in the j^{th} frequency band, K is the total number of critical bands and M is the number of the data segments. WSS has been studied extensively in recent years, and has enjoyed wide acceptance [8,14].

4.3. Composite Measure

As conventional objective measures are not sufficient to provide high correlations in terms of speech/noise distortion and overall speech quality, it is hence necessary to combine different objective measures in order to produce Composite measure [9]. Composite objective measures are the linear combination of four basic objective measures such as Segmental SNR (SNRseg), Weighted Slope Spectral distance (WSS), Perceptual Evaluation of Speech Quality (PESQ) and Log Likelihood Ratio (LLR) to form a new and more accurate measure. These measures highly correlate with speech/noise distortions and overall quality. In this paper, we have chosen a composite measure for signal distortion (Csig), a composite measure for noise distortion (Cbak), and a composite measure for overall speech quality (Covl). A higher value of composite measures indicates better quality of speech. These values are obtained by linearly combining the existing objective measures by the following relations [10,11]:

$$Csig = 3.093 - 1.029LLR + 0.603PESQ - 0.009WSS \tag{6}$$

$$Cbak = 1.634 + 0.478PESQ - 0.007WSS + 0.063segSNR \tag{7}$$

$$Covl = 1.594 + 0.805PESQ - 0.512LLR - 0.007WSS \tag{8}$$

V. RESULTS

Here, G.729A is implemented in MATLAB and performance of reconstructed speech of several γ values is evaluated using different Objective measures. For the purpose of Objective analysis, two different wave files have been chosen. Each Wave file is sampled at 8 kHz and coded by 16 bit linear PCM then coded by G729A with variation in γ at the encoder in the range of 0.4 to 0.95 with fixed step 0.05. Then we measure the quality for each synthesized speech for each γ.

Table 1 demonstrates the result analysis and comparison between speech qualities of different γ. Figure 3 demonstrates the comparison of PESQ score. Figure 4 demonstrates the comparison of WSS score.

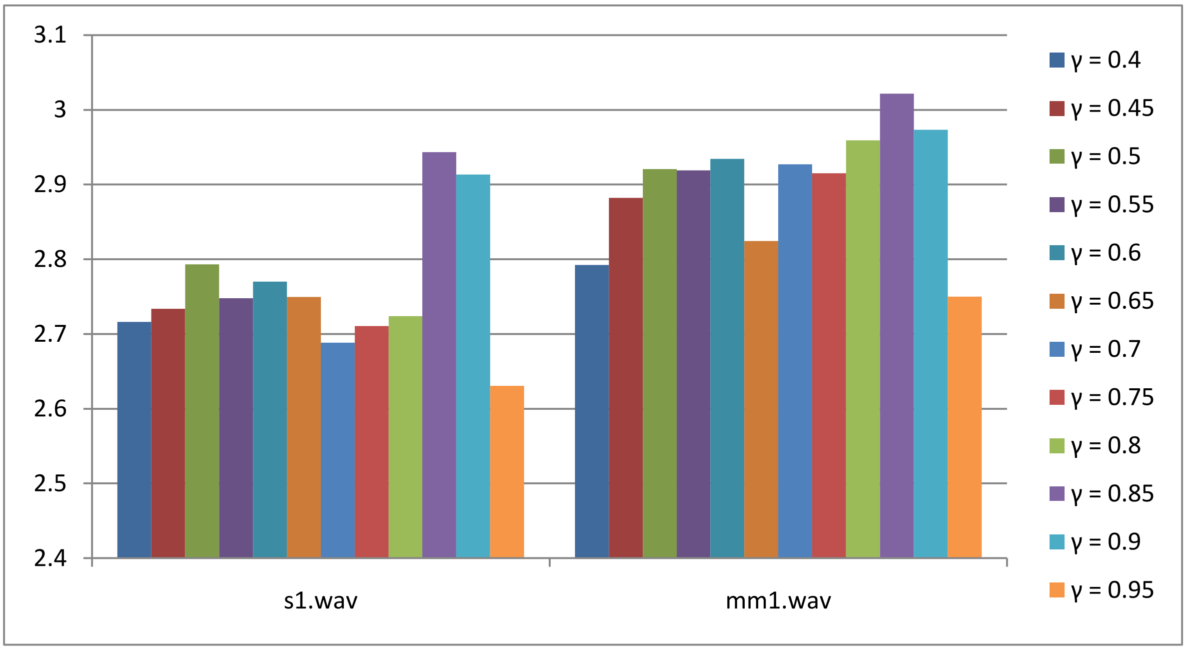

Figure 3. PESQ scores comparison between speech qualities of different γ

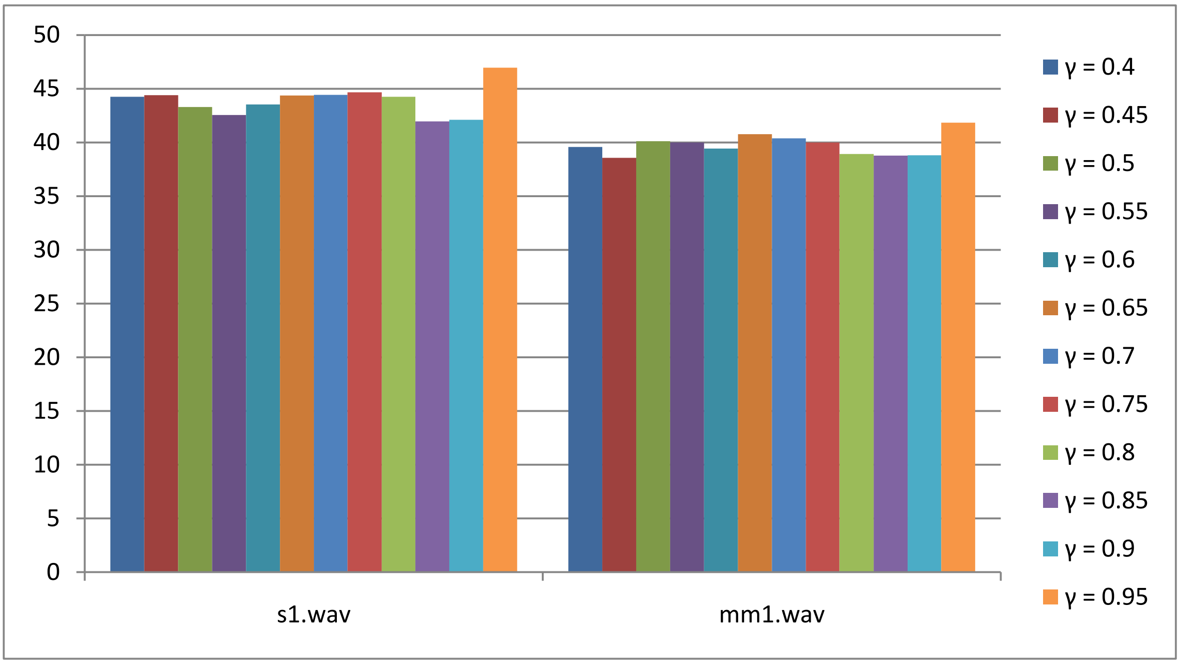

Figure 4. WSS scores comparison between speech qualities of different γ

Table 1. Comparison between speech qualities of different γ

Wave file	γ	PESQ	WSS	Composite Measures		
				Csig	Cbak	Covl
s1.wav	0.40	2.7163	44.2559	3.8256	2.5163	3.2186
	0.45	2.7337	44.4008	3.8153	2.5180	3.2219
	0.50	2.7931	43.2970	3.8662	2.5500	3.2800
	0.55	2.7477	42.5624	3.8363	2.5335	3.2440
	0.60	2.7702	43.5426	3.8505	2.5381	3.2599
	0.65	2.7497	44.3866	3.8086	2.5202	3.2266
	0.70	2.6885	44.4235	3.7837	2.4897	3.1832
	0.75	2.7108	44.6765	3.8008	2.5012	3.2024
	0.80	2.7239	44.2592	3.7928	2.5118	3.2061
	0.85	2.9433	41.9766	3.9469	2.6189	3.3993
	0.90	2.9135	42.1166	3.9151	2.5995	3.3681
	0.95	2.6306	46.9710	3.7693	2.4472	3.1405
mm1.wav	0.40	2.7922	39.5731	4.0027	2.5917	3.3568
	0.45	2.8823	38.5782	4.0660	2.6437	3.4363
	0.50	2.9206	40.1307	4.0855	2.6501	3.4614
	0.55	2.9191	39.9936	4.0857	2.6500	3.4612
	0.60	2.9342	39.4269	4.0842	2.6594	3.4695
	0.65	2.8245	40.7751	4.0037	2.6021	3.3712
	0.70	2.9270	40.3961	4.0739	2.6514	3.4582
	0.75	2.9150	39.9912	4.0783	2.6505	3.4554
	0.80	2.9592	38.9182	4.1149	2.6789	3.4986
	0.85	3.0215	38.7743	4.1499	2.7132	3.5478
	0.90	2.9732	38.8012	4.1294	2.6935	3.5279
	0.95	2.7498	41.8431	3.9854	2.5546	3.3211

VI. CONCLUSION

The 8kbit/s speech coding algorithm G.729A realized high-quality and low-delay speech coding. Adopting MATLAB to realize the algorithm is faster than with traditional programming languages such as C code. Furthermore, using MATLAB is a method which is of much more benefit to analysis and optimization for research of G.729A speech coding algorithm [12]. Through this paper we studied the design of G.729 and G.729A speech coding and investigate the effect of weight factor on the recovered speech quality. We showed that it is possible to achieve better quality and meet the standard requirements with very little modifications. As can be seen from Table 1, all measurements offer satisfactory results and it can be observed that at certain values of γ especially at γ=0.85, the speech quality is the highest in the range of 0.4 to 0.95, and is better than γ value listed in the standard (γ=0.75).

REFERENCES

[1] ITU-T Rec. G.729, "Coding of speech at 8 kbit/s using Conjugate-Structure Algebraic-Code-Excited Linear-Prediction (CSACELP)," June 2012.

[2] R. Salami, "Design and description of CS-ACELP: A toll quality 8 kb/s speech coder," IEEE Trans. Speech and Audio Processing, Vol. 6, No. 2, pp. 116-130, March 1998.

[3] R. Salami, C. Laflamme, B. Bessette, and J. Adoul, "ITU-T G.729 Annex A: Reduced complexity 8kb/s CS-ACELP codec for digital simultaneous voice and data," IEEE Communication Magazine, Vol. 35, No. 9, pp. 56-63, Sept. 1997.

[4] R. Shan Zhu and Y. Xie, "ITU-T G.729A Speech Coding Algorithm and Implementation," ICALIP 2008 International Conference, pp. 455-459, July 2008.

[5] X. Tan, R. Hu, W. Tu and H. Ai, "Implementation of G.729A on Embedded SIMD Processor," Parallel Architectures, Algorithms and Programming (PAAP), 2011 Fourth International Symposium, pp. 181-184, Dec. 2011.

[6] A. Rix, J. Beerends, M. Hollier, and A. Hekstra, "Perceptual evaluation of speech quality (PESQ)-A new method for speech quality assessment of telephone networks and codecs," in Proc. IEEE Int. Conf. Acoust., Speech, Signal Process., vol. 2, pp. 749–752, 2001.

[7] ITU-T Rec. P. 862 "Perceptual evaluation of speech quality (PESQ), and objective method for end-to-end speech quality assessment of narrowband telephone networks and speech codecs," Feb. 2000.

[8] K.Kondo, ''Subjective Quality Measurement of Speech,'' Springer, 2012

[9] Y. Hu, P. Loizou "Evaluation of Objective Quality Measures for Speech Enhancement," IEEE Transactions on Audio, Speech and Language processing, Vol. 16, No. 1, Jan. 2008

[10] ITU-T Rec. P. 835, "Subjective test methodology for evaluating speech communication systems that include noise suppression algorithm," Nov. 2003.

[11] J. Salmela and V. Mattila, "New intrusive method for the objective quality evaluation of acoustic noise suppression in mobile communications," 16th Audio Eng. Soc. Conv., PP. 6145, May 2004.

[12] M. Banerjee, B. A. Vani, S. Madhusudhan and S. Monga, "Optimizations of ITU G.729 speech codec," IEEE 60th Vehicular Technology Conference, pp. 3913-3918, Sept. 2004.

[13] Philipos C. Loizou, "Speech Enhancement: Theory and Practice, Second Edition", CRC Press, 2013.

[14] S. Moller and R. Heusdens, "Objective Estimation of Speech Quality for Communication Systems", Proceedings of the IEEE, Issue 99, pp. 1-13, April 2013.

[15] K.E.B. Ooi, M. Lech, N.B. Allen, "Multichannel Weighted Speech Classification System for Prediction of Major Depression in Adolescents", IEEE Transactions on Biomedical Engineering, Vol. 60, pp. 497-506, Feb. 2013.

AUTHORS

M. A. Rashed received the B.Sc. with grade Excellent (Honor) in Electrical & Computer Engineering from Faculty of Engineering, at Higher Technological Institute, 10th of Ramadan City, Egypt, in 2008. He achieved 7th position on his department. Graduation Project "Secure Mobile Communication System Based on CDMA2000", with grade Excellent. Currently, he is Assistant Lecturer in Electrical and Computer Engineering department (Higher Technological Institute) at 10th of Ramadan City.

T. A. El-Garf received B.Sc. and diploma degrees in Electronics and Communications Engineering Department from Military Technical College, Cairo, Egypt in 1976 and 1990 respectively. He received M.Sc. and Ph.D degrees in Electronics and Communication Engineering Department, Faculty of Engineering, Ain Shams University, Cairo, Egypt in 1990 and 1993 respectively. Currently, he is professor of communications in Electrical Engineering department (Higher Technological Institute) at 10th of Ramadan City. He published papers in international conferences in the areas of Communications and Communication Security. His research interests are in Digital Communication Security Systems, Cipher systems and VoIP security.

I. F. Tarrad received B.Sc. and M.Sc. degrees in Electronics and Communications Engineering Department from Al-Azhar University, Cairo, Egypt in 1984 and 1989 respectively and received Ph.D degree in Electronics and Communications Engineering Department from Academy of Sciences, Budapest, Hungary in 1996. Currently, he is Lecturer of Electronics and Communications Engineering at Al-Azhar University. He published papers in international conferences and journals in the areas of Electronics and Communications. His research interests are in Communication Networks, Digital Signal Processing, Wave Digital Filters and Adaptive Signal Processing.

H. A. Almotaafy received B.Sc., diploma and M.Sc. degrees in Electronics and Communications Engineering Department from Military Technical College, Cairo, Egypt in 1974, 1977 and 1979 respectively. He received Ph.D degree in Electronics and Communication Engineering Department, National Higher School of Aeronautics and Space, Toulouse, France in 1982. Currently, he is professor of communications in Electrical Engineering department (Higher Technological Institute) at 10th of Ramadan City. He published papers in international conferences in the areas of Communications, Electronics, Radar Systems and Microwave Engineering. His research interests are in Radar systems and Microwave Engineering.