Modified BTC Algorithm for Signal Quantization

Stefan Tomic, Zoran Peric, Jelena Nikolic
Faculty of Electronic Engineering at University of Niš
Aleksandra Medvedeva 14, 18000 Niš
Serbia
{stefan@elfak.rs}

ABSTRACT: *Extensive research is carried out in signal coding to achieve wideband speech signals and in this area we took the modified block truncation coding for study. The modified image coding is done to change the original BTC algorithm for processing black and white images where we preserve the basic principles. We found that the modified algorithm addresses the speech signal and images where the signal quantization is implemented. The experimentation revealed that the algorithm is effective and lead to the wider implementation of the algorithm for speech signals.*

Keywords: Adaptive Coding, Algorithm, BTC, Speech Signal Coding, Quantization

Received: 20 August 2020, Revised 15 December 2020, Accepted 22 December 2020

DOI: 10.6025/stj/2021/10/1/1-6

1. Introduction

Speech signal is one of the most important signals in nature, which digital form we meet every day. From the time of just using landline telephones, we came to an area where voice can easily be transmitted through the internet, almost instantaneously, while having a great variety of methods to do so. As modern telecommunication networks deal with a great amount of data daily, it is highly important that they use optimal amount of storage space and transmission bandwidth. This makes signal compression highly important in every telecommunications system which handles speech signals [1], [2]. Properly compressed speech signals need to comply with certain quality standards, while demand an optimal storage space. Quality is determined by the compression algorithm applied and the chosen bit rate [1], [2]. Efficient usage of available bit rate is crucial for obtaining high quality output signal, which can be easier to store and send through communication system. This gives a great importance to the research in the field of signal compression, as technologies constantly evolve and new challenges are faced.

Due to the great number of researches on this topic, there are a lot of speech signal coding algorithms in the literature. As multimedia data transmission plays important part in modern signal processing, we can find some similar coding techniques applied in audio and video signal coding. This has inspired the authors to consider implementing image coding algorithm in speech signal coding. The application of an image coding algorithm in speech signal coding represents a novelty and by applying certain modifications to the original Block Truncation Coding (BTC) algorithm [3], we obtain a novel speech signal coding algorithm. The original BTC algorithm is based on the input signal decomposition into nonoverlapping blocks of input signal samples (frames) [3]. Along with the input signal samples, the algorithm uses the mean value and standard deviation, as statistical parameters of the input signal frame in the coding process. As these parameters are also required for decoding, they also need to be transmitted through the channel. In this case, mean value and standard deviation are representing the side information, which provides higher quality of the output signal, but also increase the bit rate. Since the side information is added per frame, the amount of side information depends on the frame size. The basic BTC algorithm implements uniform quantization, while in this paper we choose more robust quantizers, suitable for wideband speech signal coding.

It has been shown in [4] that the BTC algorithm can be successfully applied in narrowband speech signal coding. Also, the similar approach has been utilized in [5], where the modified BTC algorithm has been applied in audio signal coding. Unlike in [4] and [5], in this paper we introduce the usage of sub-frames for higher adaptation of the algorithm to the statistics of the wideband speech signal. In addition, we implement changes in the quantization process, which we describe in Section 2. The modified algorithm is applied in coding of wideband speech signal, with sampling frequency of 16 KHz. The rest of the paper is organized as follows. Section 2 describes the modified BTC algorithm design. The performance analysis of the algorithm is presented in Section 3, while Section 4 is dedicated to conclusions.

2. Modified BTC Algorithm

As mentioned, the original BTC algorithm is developed for black and white image compression [3]. It is based on decomposing the input signal into frames, which are processed individually. For each frame, the mean value and standard deviation are calculated and used in the quantization process. This gives us three variables in total, the input signal samples, and the two aforementioned statistical parameters which describe the frame. The basic BTC algorithm implements three uniform quantizers, with the total bit rate of 2 bits per sample [3]. This is possible due to the algorithm design and the fact that black and white image pixel value is defined in the interval from 0 to 255. As wideband speech signal sample amplitude can have unpredictable value, we implement more robust quantization methods, while applying higher bit rates. Additionally, we narrow the range of the input signal samples, by subtracting the mean value of the subframe from each subframe sample. In such a way, difference signal frame is formed, which is more suitable for coding, due to its lower amplitude dynamics. In difference to the original algorithm, in the modified BTC algorithm, each input signal frame is divided into subframes.

Unlike with the original BTC algorithm [3], we design three different quantizers to be applied to different variables of the input signal. Figure 1 presents the block diagram of our modified BTC algorithm, where quantizers are denoted by Encoder 1-3 and Decoder 1-3. Difference signal defined by the frame length of *Mf* samples for *L* number of subframes, each containing *Msf* samples is calculated as:

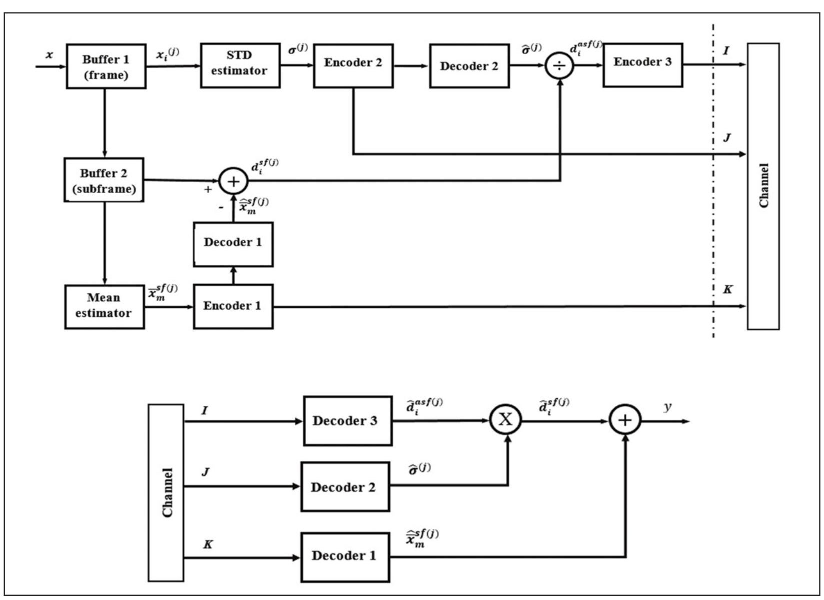

Figure 1. Modified BTC algorithm: encoder (above) and decoder (below)

$$d_i^{sf(j)} = x_i^{(j)} - \hat{\hat{x}}_m^{sf(j)}, i = 1,2,\dots,M_{sf}, j = 1,2,\dots,L\ , \quad (1)$$

where $x_i^{(j)}$ denotes the ith input signal sample, of the jth subframe, while $\hat{\hat{x}}_m^{sf(j)}$ is the quantized mean value of the jth subframe. Mean value is calculated for each subframe and it is quantized by applying the quasilogarithmic μ law quantizer [6], denoted by Encoder 1 and Decoder 1. The thus defined difference signal frame is brought to the input of Encoder 3, along with the standard deviation of the frame. Standard deviation is calculated for each frame and it is quantized with the application of the log-uniform quantizer [7] denoted by Encoder 2 and Decoder 2. Standard deviation of the frame is used in adaptive quantization of the subframes that the frame consists of, which outputs adaptive difference signal subframe denoted as $\hat{d}_i^{asf\,(j)}$. The outputs of the three defined encoders denoted by I, J and K are sent through the channel as binary information. The received information needs to be reconstructed into usable form, which represents the original input signal. Decoded output signal sample is obtained by adding the quantized mean value of the subframe to the adaptive difference signal subframe samples:

$$y_i^{(j)} = \hat{\hat{x}}_m^{sf(j)} + \hat{d}_i^{asf\,(j)}, i = 1,2,\dots,M_{sf}, j = 1,2,\dots,L\ . \quad (2)$$

This procedure is performed for all subframes of the input signal, which gives us the reconstructed signal. As we have the original input signal and its reconstructed form, we can inspect the objective quality of the output signal. For this purpose, we use a well-known measure, named signal to quantization noise ratio (SQNR) [1], [2], [8], [9]. SQNR is calculated for a certain bit rate used. As our algorithm implements frames and subframes, the total bit rate is influenced by the frame and subframe size and the bit rate used in individual quantizers. Accordingly, the total bit rate used in the modified BTC algorithm is defined by:

$$R = r_3 + \frac{r_1}{M_{sf}} + \frac{r_2}{M_f}\ , \quad (3)$$

where r_3 denote the bit rate used for quantization of the difference signal frame, r_1 and r_2 represent the bit rates used for quantizing the mean value and standard deviation, respectively, while M_{sf} and M_f denote the number of samples which subframes and frames contain, respectively.

Along with the changes in the algorithm design, our algorithm implements significant modifications in the quantization process. Quantization is a significant step in obtaining a digital representation of the signal, and it has high impact on the output signal quality. Quantizer can be defined as a series of encoder and decoder [9]. It performs mapping of the unknown input signal amplitudes, into the group of allowed signal amplitudes. For defining a quantizer one need to determine the support region of the quantizer, which is defined by the minimum and maximum support region thresholds as [x_{min}, x_{max}]. Depending on the quantizer type and implemented bit rate, the decision thresholds which divide the support region are chosen. Based on which decision threshold is defined, the input speech signal sample is mapped to the corresponding representative. In the quantization design process, speech signal can be successfully modelled with a memoryless Laplacian source, with the mean value equal to zero, as it is assumed in this paper. Laplacian source of variance σ^2 and mean value equal to zero is defined by [1], [2], [9]:

$$p(x) = \frac{1}{\sqrt{2\sigma^2}} \exp\left\{-\frac{|x|\sqrt{2}}{\sigma}\right\}. \quad (4)$$

As above mentioned, the modified BTC algorithm implements three different quantizers, where each of them is dedicated to a certain input signal parameter. The input signal is divided into frames, while each frame is divided into subframes. Mean value is calculated for each subframe, while the standard deviation is calculated just for the main frames. As standard deviation does not change as fast as the mean value, we can use the standard deviation of the main frame to perform the adaptive quantization of the subframe and reduce the bit rate.

Mean value of the subframe is quantized with the application of quasilogarithmic quantizer, that is the logarithmic quantizer defined with the μ compression law. This quantizer performs signal compression by applying a compressor function defined by [6], [9]:

$$c_\mu(x) = \frac{x_{\max}}{\ln(1+\mu)} \ln\left(1 + \mu\frac{|x|}{x_{\max}}\right)\operatorname{sgn}(x), \quad |x| \le x_{\max}\ , \quad (5)$$

where μ represents the compression factor. We implemented a compression factor equal to 255, to ensure robustness in wider area of possible subframe mean values. The bit rate used for mean value of the subframe amounts to 5, so that the side information added for the mean amounts to 5 bits per subframe. The designed quantizer implements optimal support limit, determined for the quasilogarithmic quantizer with 32 quantization levels designed for the Laplacian source of unit variance.

Unlike the mean value, the standard deviation is calculated just for main frames as it assumed that it will properly correspond to each subframe in the main frame. Standard deviation of the input signal frame is quantized with the application of the log-uniform quantizer, constructed for N_g quantization levels and for the support region [20 $log\ \sigma_{min}$, 20 $log\ \sigma_{max}$] [7]:

$$20\log\left(\hat{\sigma}^{(j)} = \hat{\sigma}_k^{(j)}\right) = 20\log_{10}\left(\sigma_{\min}\right) + \frac{2k-1}{2}\Delta^{\mathrm{lu}}, \tag{6}$$

where $k = 1,\ldots, N_g$, while a single quant width is defined by:

$$\Delta^{\mathrm{lu}} = \frac{20\log_{10}\left(\dfrac{\sigma_{\max}}{\sigma_{\min}}\right)}{N_g}. \tag{7}$$

For quantizing standard deviation of the frame, we have designed a log-uniform quantizer with 16 representational levels, which use 4 bits per input signal frame. The support region of this quantizer is given by the range of [-20dB, 20dB].

The third variable in the designed algorithm is difference signal frame, which carries the most important information about the input signal. Difference signal is quantized by applying forward adaptive quasilogarithmic quantizer, designed for smaller value of compression factor. We chose compression factor equal to 80, as we are aware of the fact that a smaller value is a proper choice for adaptive quantization. Quantization is performed in two phases. Firstly, we obtain the fixed represents of the difference signal frame, by applying compression function defined by Eq. (5). Fixed represents are multiplied by the quantized value of standard deviation, which gives us adaptive represents of the difference signal frame.

3. Experimental Results and Analyses

This section describes the implementation of the designed algorithm in real wideband speech signal coding and its objective performance evaluation. When applying quantization, we introduce an irreversible error for each input signal sample. These errors can be summed up into a mean squared error, which form a measure called signal distortion. In the case of using real input signal defined by:

$$D = \frac{1}{S}\sum_{n=1}^{S}\left(x_n - y_n\right)^2, \tag{8}$$

where x_n and y_n represent the original and quantized input signal samples, respectively, while S represents the total number of the input signal samples. Signal distortion determines the objective quality measure used in this paper, named SQNR, defined by [9]:

$$\mathrm{SQNR[dB]} = 10\log_{10}\left(\frac{\sigma^2}{D}\right). \tag{9}$$

Input signals used in the experiments are male and female speech signals, sampled at 16 KHz. Table I presents the experimental results of implementing the modified BTC algorithm to a male speech signal, sampled at 16 KHz. Frame size is equal to 320 samples, while subframe size varies in the range from 5 to 320 samples. As we buffer 320 samples, of the input speech signal sampled at 16 kHz, this introduce delay of 0.02 seconds into the speech signal transmission process. In the case when frame and subframe have the same size, we practically do not implement subframe in the algorithm. This special case is presented to show the benefits of including subframes. Additionally, we compare the results with Pulse Code Modulation (PCM) [6], a widely implemented coding standard, typically used for comparison. In this case, PCM is designed for a fixed bit rate, equal to 7 bits per sample, as it does not implement frames and subframes, so we do not introduce the side information.

By observing Table 1, one can notice that the modified BTC algorithm provides gain in SQNR for all observed parameters, when compared to PCM. In the case when the subframes are not implemented, this gain amounts to around 4.5 dB. To analyze the influence of subframes, we can observe the case when subframe size is equal to 10 samples. In this case bit rate is increased for 0.5 bits per sample. As it is known that one additional bit increases the SQNR for approximately 6 dB [9], increase of 0.5 bits per sample, should increase the SQNR for around 3 dB. By comparing the values from the Table 1, we see that the SQNR of the modified BTC algorithm is increased for around 4.4 dB. This means that we obtain the gain in SQNR of around 1.4 dB, just by implementing subframes into the algorithm. By applying the same logic to the PCM, and estimating its SQNR for 7.51 bits per sample, we can conclude that the modified BTC algorithm provides gain in SQNR of around 5.9 dB, for the case of subframe size is equal to 10 samples: 37.05dB - (28.13 dB + 3 dB).

Frame size	Subframe size	Bit rate	$SQNR_{BTC}$ [dB]	$SQNR_{PCM}$ [dB]
320	5	8.0125	39.7409	28.13
320	10	7.5125	37.0506	28.13
320	20	7.2625	34.5204	28.13
320	40	7.1375	33.2855	28.13
320	320	7.0281	32.666	28.13

Table 1. SQNR Obtained for Male Speech Signal

Frame size	Subframe size	Bit rate	$SQNR_{BTC}$ [dB]	$SQNR_{PCM}$ [dB]
320	5	8.0125	41.2025	29.60
320	10	7.5125	38.2895	29.60
320	20	7.2625	35.4304	29.60
320	40	7.1375	33.8531	29.60
320	320	7.0281	33.2202	29.60

Table 2. SQNR Obtained for Female Speech Signal

Table 2 presents the performance of the modified BTC algorithm and PCM, when both are applied to female speech signal, sampled at 16 KHz. By applying the same principles as for Table 1, the modified BTC algorithm without implementing the subframes provides gain in SQNR of around 3.6 dB, compared to PCM. When we implement subframes, consisting of 10 input signal samples, the practical gain in SQNR amounts to around 5.7 dB, compared to the PCM, for the same bit rate equal to 7.51 bits per sample. Again, to analyze the influence of the subframes, we compare the performance of the modified BTC algorithm, when we do not implement subframes, to the case when subframe consists of 10 input signal samples. By comparing the results of the modified BTC algorithm with and without implementing subframes, we conclude that when applied to the female speech signal, implementation of subframes provides gain in SQNR equal to 2 dB.

4. Conclusion

In this paper, we have presented the modified BTC algorithm and its application in wideband speech signal coding. The basic principles of the original black and white image coding algorithm has been preserved, while the main modifications have been performed in the quantization process. Additionally, we have introduced the implementation of subframes, which, as shown, improve objective output signal quality, while do not significantly increase the complexity of the algorithm. By applying the algorithm in real wideband speech signal coding, we have shown that the modified BTC algorithm provides gain in SQNR ranging from 3.6 up to 5.9 dB, when compared to PCM. Furthermore, by introducing the subframes, the gain in SQNR increase in the range from 1.4 to 2 dB, while complexity of the algorithm does not significantly increase. By observing the numerical results, we can conclude that the proposed modified BTC algorithm can be successfully applied in wideband speech signal coding. The proposed algorithm use simple quantization techniques, while providing high quality output speech signal. This leaves a great space for possible improvements, which are left to the future research.

Acknowledgement

This work is supported by Serbian Ministry of Education, Science and Technologic Development (Project TR 32035).

References

[1] Chu, W. C. (2003). Speech Coding Algorithms: Foundation and Evolution of Standardized Coders, New Jersey, John Wiley & Sons, p. 143-183.

[2] Hanzo, L., Somerville, C., Woodard, J. (2007). Voice and Audio Compression for Wireless Communications, London, John Wiley & Sons, Chapter 1, p. 1-28.

[3] Delp, E. J., Mithcell, O. R. (1979). Image Compression using Block Bruncation Coding, IEEE Trans. Commun., 27 (9) 1335-1342, doi: 10.1109/tcom.1979.1094560

[4] Periæ, Z., Tomiæ, S., Tanèiæ, M. (2015). High Quality Speech Signal Coding with the Application of BTC Algorithm, TELSIKS'15, Conference Proceedings, p. 23-26, Niš, Serbia.

[5] Tomic, S., Peric, Z., Nikolic, J. (2016). Modified BTC Algorithm for Audio Signal Coding, *Advances in Electrical and Computer Engineering*, 16 (4), p. 31-38.

[6] ITU-T, Recommendation G.711, Pulse Code Modulation (PCM) of Voice Frequencies, *International Telecommunication Union*.

[7] Nikoliæ, J., Periæ, Z. (2008). Lloyd-Max's Algorithm Implementation in Speech Coding Algorithm Based on Forward Adaptive Technique, *Informatica*, 19 (2) 255-270.

[8] Kondoz, A. (2004). Digital Speech: Coding for Low Bit Rate Communication Systems, John Wiley & Sons, Ch. 2–3, p. 5-55.

[9] Jayant, N. S., Noll, P. (1984). Digital Coding of Waveforms, New Jersey, Prentice Hall, Chapter 5, p. 221-251.