ITC 3/48
Journal of Information Technology and Control
Vol. 48 / No. 3 / 2019
pp. 454-463
DOI 10.5755/j01.itc.48.3.21685

Simple Speech Transform Coding Scheme Using Forward Adaptive Quantization for Discrete Input Signal

Received 2018/09/19 | Accepted after revision 2019/08/02

http://dx.doi.org/10.5755/j01.itc.48.3.21685

Simple Speech Transform Coding Scheme Using Forward Adaptive Quantization for Discrete Input Signal

Zoran Perić, Milan Tančić, Nikola Simić

Faculty of Electronic Engineering, University of Niš, Department of Telecommunications, Serbia, Aleksandra Medvedeva 14, 18000, Niš, Serbia; e-mails: zoran.peric@elfak.ni.ac.rs, milan_tancic@yahoo.com, nikola.simic@elfak.ni.ac.rs

Vladimir Despotović

University of Belgrade, Technical Faculty in Bor, Vojske Jugoslavije 12, 19210 Bor, Serbia; e-mail: vdespotovic@tfbor.bg.ac.rs

Corresponding author: milan_tancic@yahoo.com

The speech coding scheme based on the simple transform coding and forward adaptive quantization for discrete input signal processing is proposed in this paper. The quasi-logarithmic quantizer is applied for discretization of continuous input signal, i.e. for preparing discrete input. The application of forward adaptation based on the input signal variance provides more efficient bandwidth usage, whereas utilization of transform coding provides sub-sequences with more predictable signal characteristics that ensure higher quality of signal reconstruction at the receiving end. In order to provide additional compression, transform coding precedes adaptive quantization. As an objective measure of system performance, signal-to-quantization-noise ratio is used. System performance is discussed for two typical cases. In the first case, it was considered that the information about continuous signal variance is available, whereas the second case considers system performance estimation when only the information about discretized signal variance is present, which means that there is a loss of input signal information. The main goal of such performance estimation comparison of the proposed speech signal coding model is to explore what is the objectivity of performance if the information about a continuous source is absent, which is a common phenomenon in digital systems. The advantages of the proposed coding scheme are demonstrated by comparing the performance of the reconstructed signal with other similar exiting speech signal coding systems.

KEYWORDS: Speech coding, Transform coding, Forward adaptive quantization, Quasi-logarithmic quantizer.

1. Introduction

Quantization represents the process of mapping the range of signal amplitude values, which can be continuous and infinite in general, into a set of discrete values and it represents a core method exploited in signal processing algorithms. It is an indispensable part of "lossy" signal compression algorithms, which may incorporate additional coding techniques to manipulate the presentation of a signal in digital domain. A constant need for solutions of less complexity, which would require the usage of lower bit-rates but keeping the high quality of reconstructed signal, is a demanding challenge with the rapid growth of information systems [10], [4], [11], [31], [28]. Quantization is the process of preparing a signal in digital domain and making it suitable for processing by a computer or any digital circuit [9]. Considering the growing interest in man-machine communication, speech and voice recognition is considered as important [37], [23-24], [36], [7], [16], [25]. Recent research and applications, which exploit neural networks, commonly incorporate quantization of weight coefficients and activation functions. It is shown that scalar quantization is suitable for such task as well as for speech recognition algorithms using deep convolutional neural networks [35], [13]. This paper proposes a speech signal coding scheme with scalar quantization implemented, where every sample of input signal is processed separately using scalar quantizers [6]. Such an approach is less complex than a vector quantization based approach, which processes input signal samples grouped into the vectors (arrays) and where vectors are quantized at once [35], [20].

For non-stationary signals coding, such as speech signal, the usage of adaptive schemes increases quality of reconstructed signal in the manner of adapting quantizer design to the varying statistics of an input signal frame (mean value or variance typically) to achieve better performances. More efficient bandwidth usage of the original samples can be provided by including transform coding algorithms into the coding scheme [14-15], which is a part of several standards in the field of high quality wideband speech/audio coding [19]. Some of the most well-known transforms are discrete wavelet transform (DWT) [17], [34], [12], discrete cosine transform (DCT) [8], [26], Karhunen-Loeve transform (KLT) [2-3] and Hadamard transform [33].

In this paper, we analyze and propose a transform-based adaptive coding scheme based on forward adaptive quantization in the case of discrete input signal. The coding process presented in the proposed scheme can be observed as a two stage coding. Firstly, the continuous signal is discretized using a quasi-logarithmic quantizer Q_0, which forms amplitude limited discrete signal that is further coded in the stage two. The amplitude limitation of discretized signal causes the absence of overload distortion, which is disregarded in the processing of further step. In the second stage, discretized signal is encoded using a simple transform coding that decomposes wideband speech signal into sub-sequences with narrower bandwidth range and it is adapted using a forward adaptation technique. The aim of such transform-based coding step is to provide additional compression before adaptive quantization [32], [30]. After that, speech signal is divided into frames whose size is 240 samples, which are further adapted to the variance framewise [5], [27], [21]. Since the forward adaptation is used, the variance is quantized using the log-uniform quantizer and this information is transferred to the receiver [18].

The proposed coding scheme shows suitability for speech signal coding as it provides 4.9 – 7.8 [dB] of gain comparing to the results which provide the usage of quasi-logarithmic quantizer in the second stage [29], and up to 10 [dB] of gain comparing to the results which provide the usage of uniform quantizer in the first and optimal comandor in the second stage [22].

The remaining of the paper is organized as follows: in Section 2, a detailed description of the proposed quantizer model is provided. Next, numerical results and discussion are presented in Section 3. In the end, concluding remarks are summed up in Section 4.

2. Quantizer Design for Discrete Input Signal

The proposed coding scheme is shown in Fig. 1. It can be seen that the encoder is composed of a quasi-logarithmic quantizer Q_0, which is exploited for discretization of continuous input speech signal, then buffer, variance estimator, transform encoder and forward-adaptive quasi-logarithmic quantizers AQ_1 and AQ_2. The purpose of quasi-logarithmic quantizer Q_0 (R_0=8 [bits/sample], μ=255) is to obtain discrete samples of continuous input speech signal that is further fed into the buffer. Although a robust quasi-logarithmic quantizer is used, significant errors may occur for higher values of variance due to range mismatch, resulting a huge difference of estimated performance between the cases where the information about continuous signal is available and when the information about discrete signal only is achievable. If the wide range of variances is observed, it is not enough to exploit only one robust quantizer Q_0, but it is necessary to use two quantizers in pre-processing, where one will cover processing of lower bands range while another will be used for processing signals of upper band range variances. On the other hand, if uniform quantizers are exploited, which are known as not robust, system performance would be much worse. Such an analysis has not been done before, and it is therefore considered as a significant step forward in the field.

After discretizing continuous signal, discretized signal is fed to the buffer which is used to divide signal into frames. Consequently, further signal processing is not anymore sample-based but frame-based. Next, simple transforms t_1 and t_2 are applied to form two sub-sequences y_1 and y_2, creating two independent signals with more predictable characteristics. These transforms are defined similarly as Hadamard transform for a group of two samples and their form is [32], [30]:

$$(t_1): y_1 = \frac{x_n + x_{n+1}}{2}, \tag{1}$$

$$(t_2): y_2 = \frac{x_n - x_{n+1}}{2}, \tag{2}$$

where x_n and x_{n+1} are neighboring samples of the input signal, while y_1 and y_2 represent samples of transformed signal. These transformed signals have variances σ_{12} and σ_{22}, respectively, whose values depend on the discrete input signal variance σ_x^2 and correlation coefficient ρ [30]:

$$\sigma_1^2 = \frac{\sigma_x^2}{2}(1+\rho), \tag{3}$$

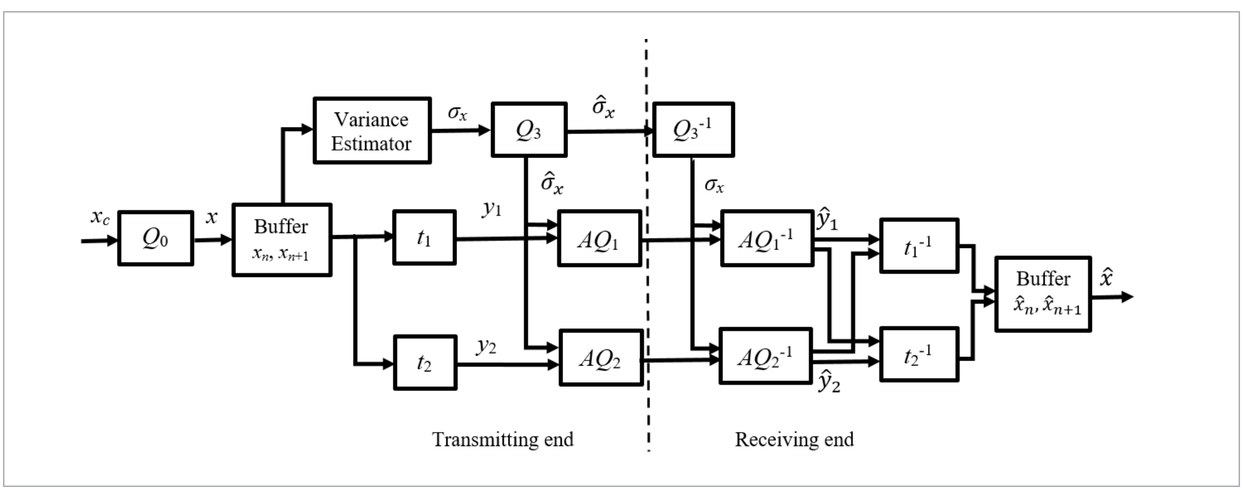
$$\sigma_2^2 = \frac{\sigma_x^2}{2}(1-\rho). \tag{4}$$

Such transformed sequences are independent and are further encoded separately using quantizers with forward adaptation applied (AQ_1 and AQ_2).

In order to achieve higher quality of reconstructed signal, adaptive quantization based on the variance of the signal has been performed in the next step [15], [27]. Firstly, frames that consist of M samples of input signal are being loaded into the buffer. The variance of

Figure 1

The proposed speech coding scheme for discrete input signal



M samples in a frame is calculated by using variance estimator, then quantized by using log-uniform quantizer Q_3 and after that it is transmitted to the receiving end. It should be noted that log-uniform quantizer is actually uniform quantizer designed in logarithmic domain which is described in details in [15], [27]. Next, variance quantized this way is sent to adaptive quantizers AQ_1 and AQ_2 for support range determination for each frame using the following expressions:

$$x_{\max(AQ_1(i))} = x_{\mathrm{mL}} \cdot \hat{\sigma} \cdot \sqrt{\left(\frac{1+\rho}{2}\right)}, \tag{5}$$

$$x_{\max(AQ_2(i))} = x_{\mathrm{mL}} \cdot \hat{\sigma} \cdot \sqrt{\left(\frac{1-\rho}{2}\right)}, \tag{6}$$

where $\hat{\sigma}$ represents quantized standard deviation, ρ is correlation coefficient of the input signal whereas x_{mL} is initial support range value of quantizer designed for Laplacian source with the unit variance, compression factor μ and N quantization levels [30], [1].

Quasi-logarithmic quantizers used in the proposed model (AQ_1, AQ_2, and Q_0) are designed using the μ-logarithmic compression law, whose compression function is defined as [10], [27]:

$$c(x) = \frac{x_{\max}}{\ln(1+\mu)} \ln\left(1 + \mu \frac{|x|}{x_{\max}}\right) \mathrm{sgn}(x), \quad |x| \le x_{\max}, \tag{7}$$

where $x_{\max}$ is the support range of the quantizer, whereas μ is the compression factor. According to the μ-logarithmic compression function, decision thresholds x_i' and representation levels y_i' are obtained [10], [27].

Commonly, a logarithmic quantizer is suitable to use for middle and high bit-rates ($N \ge 8$) [10], [27]. As the quasi-logarithmic quantizers are exploited for both signals, y_1 and y_2, the quantizers AQ_1 and AQ_2 are designed for bit-rates defined as [27]:

$$R_t = R + \frac{1}{2}\log_2 \frac{\sigma_t^2}{\prod_{k=1}^{M} \left(\sigma_k^2\right)^{\frac{1}{M}}}. \tag{8}$$

Equation (8) represents general expression for obtaining optimal bit-rate where M is the total number of branches (independent sub-sequences), whereas t can take values 1 or 2 for the proposed model (branches 1 and 2), while k presents counter through the branches. Consequently, the optimal values of R_1 and R_2 (AQ_1 and AQ_2, respectively) for the proposed coding scheme could be defined as:

$$R_1 = R + \frac{1}{4}\log_2 \frac{1+\rho}{1-\rho}, \tag{9}$$

$$R_2 = R + \frac{1}{4}\log_2 \frac{1-\rho}{1+\rho}. \tag{10}$$

Among other common objective performance measures, in this paper we observe the signal-to-quantization-noise ratio (SQNR) which is calculated for the proposed coding scheme by using a modified model that have been proposed in [29]. Estimation is performed such that the input signal is divided firstly into frames of M samples in order to calculate signal dynamics, by determining the highest and lowest variance of a frame as:

$$B[\mathrm{dB}] = 10 \cdot \log\left(\frac{\sigma_{\max}^2(M)}{\sigma_{\min}^2(M)}\right), \tag{11}$$

where $\sigma_{\max}^2$ and $\sigma_{\min}^2$ represent highest and lowest signal variances, respectively, whereas B denotes signal dynamics. In order to provide an adequate comparison, referent variance is equal to 0 [dB], as it is a common case, and it is chosen at the half of the dynamic range, limited by $\sigma^2_{\max}$ and $\sigma^2_{\min}$. The whole range B=60 [dB] (-30 to 30 [dB]) is divided into the segments with 2 [dB] step size, whereas input signal is divided into frames of M samples as for variance determination. Next, for each frame level L_i is calculated. It denotes in which segment the observed frame is located. For each segment of the range, the number of frame appearances for the observed variance is counted, and the mean SQNR value of all frames that are located in a certain segment after that is calculated. Equation (12) shows SQNR for a single frame (i^{th}) which is located in the j^{th} segment. SQNR of each frame that is located in the segment presents logarithmic ratio of signal variance for a certain frame and its distortion, and it can be calculated as [29]:

$$SQNR\left(\sigma^2_{y(i)}\right)[dB] = 10 \cdot \log\left(\frac{\sigma^2_{x(i)}}{D\left(\sigma^2_{y(i)}\right)}\right)[dB], \quad (12)$$

$$D\left(\sigma^2_{y(i)}\right) = \sum_{k=1}^{M}\left(x_k - \hat{x}_k\right)^2, \quad (13)$$

where $\sigma^2_{x(i)}$ represents signal variance of the i^{th} frame in the j^{th} segment, while $D\left(\sigma^2_{y(i)}\right)$ represents its distortion. Furthermore, x_k and $\hat{x}_k$ are real and rounded values of the k^{th} sample in the i^{th} frame, while M is the number of samples in the frame. SQNR (L_j) represents the average SQNR for the j^{th} segment, and it is defined as:

$$SQNR\left(L_j\right) = \frac{1}{M_i}\sum_{m=1}^{M_i} SQNR\left(\sigma^2_{y(i)}\right)[dB], \quad (14)$$

where level L_i is calculated as 10 times of logarithmic ratio of frame variance and referent variance at 0 [dB], M_i is the number of frames allocated in the segment L_j, while the average value of SQNR for the whole dynamic range is calculated as:

$$SQNR_{avg} = \frac{1}{L}\sum_{j=1}^{L} SQNR\left(L_j\right)[dB]. \quad (15)$$

For the purpose of accuracy consideration of the performance measure model, it has been shown the results of the standard average quality measure $SQNR_{avg(st)}$ in the wide range of variances, which do not incorporate segmentation of dynamic range. It is defined as [29]:

$$SQNR_{avg(st)}[dB] = \frac{1}{F}\sum_{l=1}^{F} SQNR(l)[dB], \quad (16)$$

where F is the total number of frames in which the signal is divided, SQNR(l). It should be noted that in the experimental analysis, M=240 samples per frame have been used, whereas the recorded benchmark test signal consists of about 4500 frames.

3. Numerical Results and Analysis

This section provides an analysis of numerical results obtained by theoretical considerations and by performing an experiment. As it has been mentioned previously, the experiment was performed for sets of bit-rates ($R_1 \in \{5, 6, 7\}$ bits/sample and $R_2 \in \{3, 4, 5\}$ [bits/sample]). Since we consider forward adaptive quantization, it is necessary to transmit additional information about signal variance to the receiving end, which is causing additional bits for encoding the variance of a frame. The variance is encoded using the log-uniform quantizer designed for low and middle bit-rates ($R_3 \in \{4, 5\}$ [bits/sample]), whereas the frame size M is equal to 240 samples.

Figure 2 shows theoretically obtained SQNR in the wide range of signal variances for quasi-logarithmic quantizer, without included transform coding ($R \in \{4, 5, 6\}$ [bits/sample], $\mu \in \{20, 255\}$) designed for continuous input signal in the case of bit-rate of the log-uniform quantizer R_3=4 and R_3=5 [bit/sample]. It can easily be seen that robustness can be increased by increasing R_3, as SQNR is higher for lower compression factor value.

Next, Table 1 shows the average SQNR for the same system from Figure 2 (theoretically obtained), with

Figure 2

SQNR in wide range of input signal variances in the case of $R \in \{4, 5, 6\}$ [bits/sample] and different μ and R_3 (R_3 = 4 – red line, R_3 = 5 – green line) values

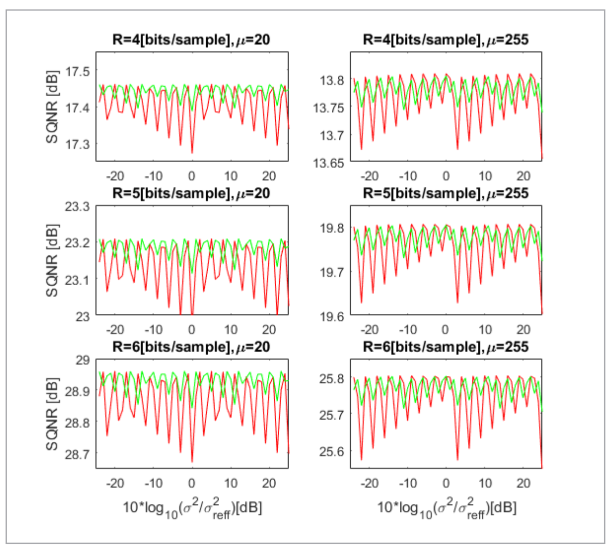


Table 1

Theoretically obtained SQNR for quasi-logarithmic quantizer with and without forward adaptation

R_{av} [bits/ sample]	R_3=4 [bits/sample]		R_3=5 [bits/sample]		Non-adaptive logarithmic quantizer	
	SQNR [dB] μ=255	SQNR [dB] μ=20	SQNR [dB] μ=255	SQNR [dB] μ=20	SQNR [dB] μ=255	SQNR [dB] μ=20
4	13.8025	17.4037	13.8080	17.4406	11.2086	11.7467
5	19.7970	23.1321	19.8045	23.1819	15.7951	15.7195
6	25.7909	28.8026	25.8009	28.9296	20.2153	19.6486

additional case of non-adaptive quasi-logarithmic quantizer, due to comparison and observing suitability of applied adaptation in the proposed coding scheme. It can be noticed that adaptive solutions provide from 2.55 [dB] up to 9.2 [dB] higher SQNR, which can be considered as a significant improvement.

The theoretical comparison to the results in paper [29] shows that the proposed speech signal coding scheme provides over 3 [dB] higher average SQNR for the same bit-rate.

The rest of this section will be dedicated to discussion of the experimental results. As it was mentioned previously, the experiment is performed for wideband speech signal whose variance is σ_x^2=0.0021 (sampled at 16 kHz) using the proposed coding scheme shown in Figure 1.

Quasi-logarithmic quantizer Q_0, applied for input signal discretization, is designed for the bit-rate R_0= 8 [bits/sample] while compression factor value is μ=255 for all cases. Tables 2 and 3 show SQNR values obtained by using the proposed coding scheme for various values of the average bit-rate: $R_{av} = R+R_v$, where $R = (R_1+R_2)/2$ and $R_v = R_3 /M$, represent the required number of bits per sample to transmit signal variance for the observed frame. Firstly, Table 2 shows SQNR values obtained by using the proposed coding scheme in the case when quantizer AQ_1 is designed for R_1=5 [bits/sample], while quantizer AQ_2 is designed for R_2=3 [bits/sample], whereas the signal variance is quantized using log-uniform quantizer designed for R_3=4 [bits/sample].

The main reason for choosing these bit-rates for quantizers AQ_1 and AQ_2 is convenient comparison with the results obtained using the coding scheme without transform coding and forward adaptation included [22], in order to demonstrate the influence of these techniques on coding scheme, i.e. to show the coding gain that these techniques provide. By comparing the results from Table 2 with the performance from [29], it can be seen that the proposed coding scheme provides 4.9–5.9 [dB] higher SQNR in the case of μ=255 and 5.8–7.8 [dB] higher SQNR in the case of μ=20, for different bit-rates.

The differences between SQNR results obtained using Equations (15) and (16) that can be noticed are the consequences of different way of measuring SQNR. Equation (16) applies standard model for coding quality measuring that midranges all frames in the dynamic variance range while the second model for quality measuring, applied in this paper (Equation (15)), midranges SQNR values for all frames in every segment of dynamic range and after that midranges these SQNR values providing the more accurate average SQNR for dynamic range. It can be noticed that the coding quality gain can be increased by increasing bit-rate of the log-uniform quantizer.

Table 2

Experimental results for the proposed coding scheme in the case of R_3 = 4 [bits/sample]

			R_{av} [bits/sample]		
μ	Input	SQNR [dB]	4.01667	5.01667	6.01667
20	Cont.	$SQNR_{avg}$	20.0062	25.6692	31.1689
	Disc.	$SQNR_{avg}$	20.7713	27.3206	32.0262
	Cont.	$SQNR_{avg(st)}$	20.7398	26.4888	32.2254
	Disc.	$SQNR_{avg(st)}$	21.5754	27.7954	33.1885
255	Cont.	$SQNR_{avg}$	16.9248	22.9970	28.9011
	Disc.	$SQNR_{avg}$	18.3151	24.5445	30.4459
	Cont.	$SQNR_{avg(st)}$	17.3997	23.3876	29.4358
	Disc.	$SQNR_{avg(st)}$	19.2437	24.6927	30.8136

Table 3

Experimental results for the proposed coding scheme in the case of $R_3 = 5$ [bits/sample]

			R_{av} [bits/sample]		
μ	Input	SQNR [dB]	4.02083	5.02083	6.02083
20	Cont.	$SQNR_{avg}$	20.1395	25.7187	31.4030
	Disc.	$SQNR_{avg}$	20.9472	27. 7113	32.9634
	Cont.	$SQNR_{avg(st)}$	20.8485	26.6353	32.3656
	Disc.	$SQNR_{avg(st)}$	21.6146	28.0653	33.9725
255	Cont.	$SQNR_{avg}$	16.9693	23. 0142	29.2311
	Disc.	$SQNR_{avg}$	18.4096	24. 8170	30.9516
	Cont.	$SQNR_{avg(st)}$	17.4584	23.5475	29.5184
	Disc.	$SQNR_{avg(st)}$	19.3170	25.3108	31.3613

Table 3 shows the results for SQNR obtained using both models in the case of log-uniform quantizer's bit-rate R_3=5 [bits/sample]. By comparing the results presented in Table 3 with the corresponding one shown in Table 2, it can be seen that by increasing the total average bit-rate of the proposed scheme by 0.00416 [bit/sample] (due to exploiting additional 1 bit for variance coding), the coding quality increases up to 0.9 [dB], depending of compression factor μ and total average bit-rate. Next, it can also be noticed by observing Table 2 that the value of compression factor μ=20 provides higher SQNR comparing to the case when μ=255. Furthermore, it can easily be seen that the proposed coding scheme estimates 1–3 [dB] higher SQNR in the case when there is available information only about variance of discretized signal, for all cases of R and μ, comparing to the case when variance of continuous signal is available. Similar remarks can be made after observing Table 3. This means that the amplitude limitation of the discrete signal provides higher coding quality than the case of continuous signal coding, because of the huge loss of information in the quantizing process of the continuous signal.

Thus, in this paper, we have shown the numerical results obtained using Equation (15) for a wide range of input signal variances and different bit-rates in Figures 3 and 4. Both cases are shown: when the information about continuous signal variance is present as well as the case when the information about discrete variance is available only.

In Fig. 3, it is presented SQNR in the case of total average bit-rate of the coding scheme R_{av}=4.01667 [bits/sample] (R_1=5, R_2=3, R_3=4) and R_{av}=4.02083 [bits/sample] (R_1=5, R_2=3, R_3=5), whereas in Fig. 4, it is presented SQNR in the case of coding scheme total average bit-rates of R_{av}=5.01667 [bits/sample] (R_1=6, R_2=4, R_3=4) and R_{av}=5. 02083 [bits/sample] (R_1=6, R_2=4, R_3=5).

Figure 3

SQNR in wide range of input signal variances in the case of different μ and R_3 and coding scheme total average bit-rate $R_{av} \approx 4$ [bits/sample]

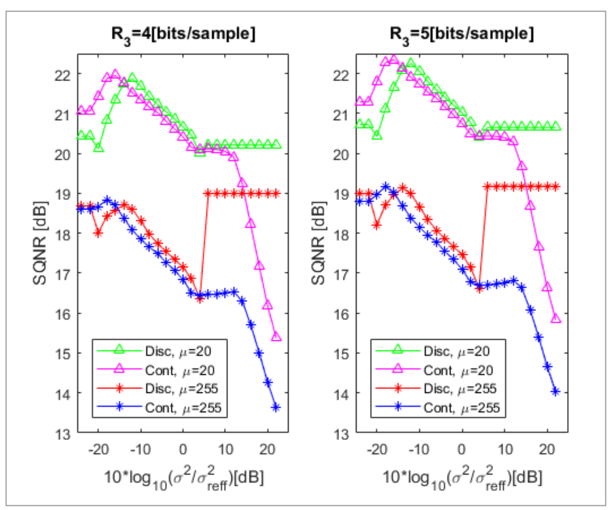
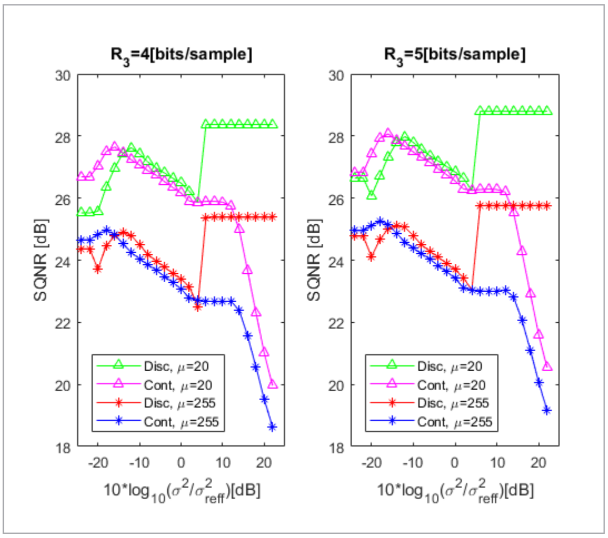


Figure 4

SQNR in wide range of input signal variances in the case of different μ and R_3 and coding scheme total average bit-rate $R_{av} \approx 5$ [bits/sample]



By observing both figures, it can be noticed that estimation of SQNR for the proposed coding scheme in the case when the information about discrete input signal is available only, provides approximately the same SQNR estimation for lower signal variances. However, estimation in the case when input signal variances are higher than 5 [dB], shows unexpected performance improvement, which appears due to support range mismatch of quantizer Q_0 [29].

4. Summary and Conclusions

In this paper, we have presented the coding scheme for discrete input signal. We have experimentally demonstrated its suitability for speech signal coding by analysing performances in the cases of middle bit-rates. It was demonstrated that application of transform coding and forward adaptation significantly increase the coding quality and provide quality robustness in the wide range of input signal variances, which shows its potential for application to speech coding algorithms. Furthermore, quantizer design and deep analysis of performance measures are other main contributions of the paper.

As it is a common case that the information about continuous signal is not available in digitization systems and that system performance should be estimated using the information about discrete signal variance, it has been analyzed performance estimation in both cases. It has been shown that the proposed coding scheme provides excellent estimation in the case of input signal variances lower than 5 [dB] for various values of system parameters. However, the model overrates the performance for higher variances in the case when the information about continuous signal is not available, due to support range mismatch of quantizer Q_0, which will be considered in the future research.

Acknowledgments

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

References

1. Aleksić, D., Perić, Z., Nikolić, J. Support Region Determination of Quasi-Logarithmic Quantizer for Laplacian Source. Przeglad Elektrotevhniczny, 2012, 88 (7A), 130-132.
2. Aristov, V. Distance Localization of the Moving Object by Applying the Karhunen-Loeve Transform to Ultra-Wideband Impulse Signals. Automatic Control and Computer Sciences, 2017, 51(5), 294-300. https://doi.org/10.3103/S0146411617050029
3. Bavirisetti, D. P., Dhuli, R. Fusion of Infrared and Visible Sensor Images Based on Anisotropic Diffusion and Karhunen-Loeve Transform. IEEE Sensors Journal, 2016, 17(1), 203-209 https://doi.org/10.1109/JSEN.2015.2478655
4. Biswas, A., Sahu, P. K., Bhowmick, A., Chandra, M. Speech Recognition Using ERB-like Admissible Wavelet Packet Decomposition based on Perceptual Sub-Band Weighting. IETE Journal of Research, 2015, 62(2), 129-139. https://doi.org/10.1080/03772063.2015.1056844
5. Chu, W. C. Speech Coding Algorithms. Foundation and Evolution of Standardized Coders, John Wiley & Sons, New Jersey, 2003, Chapters 5-6, 143-183. https://doi.org/10.1002/0471668850
6. Farias, R. C., Brossier, J-M. Scalar Quantization for Estimation: From an Asimptotic Design to a Practical Solution. IEEE Transactions on Signal Processing, 2014, 62(11), 2860-2870. https://doi.org/10.1109/TSP.2014.2318140

7. Fayek, H. M., Lech, M. Cavedon, L. Evaluating Deep Learning Architectures for Speech Emotion Recognition. Neural Networks, 2017, 92, 60-68. https://doi.org/10.1016/j.neunet.2017.02.013
8. Fracastoro, G., Fosson, S. M., Magli, E. Steerable Discrete Cosine Transform. IEEE Transactions on Image Processing, 2017, 26 (1), 303-314. https://doi.org/10.1109/TIP.2016.2623489
9. Gray, R. M., Neuhoff, D. L. Quantization. IEEE Transactions on Information Theory, 1998, 44(6), 2325-2383. https://doi.org/10.1109/18.720541
10. Jayant N. S., Noll, P. Digital Coding of Waveforms. New Jersey, Prentice Hall, Chapter 5, 1984, 221-251.
11. Kadyan, V., Mantri, A., Aggarwal, R. K. Refinement of HMM Model Parameters for Punjabi Automatic Speech Recognition. (PASR) System, IETE Journal of Research, 2018, 64 (5), 673-688. https://doi.org/10.1080/03772063.2017.1369370
12. Khalifa, N., Filali, R. L., Benrejeb, M. A Fast Selective Image Encryption Using Discrete Wavelet Transform and Chaotic Systems Synchronization. Information Technology and Control, 2016, 45(3), 235-242. https://doi.org/10.5755/j01.itc.45.3.12650
13. Li, L., Wang, D., Chen, Y., Shi, Y., Tang, Z., Zheng, T. F. Deep Factorization for Speech Signal. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Calgary, Canada, 2018, Electronic ISSN: 2379-190X. https://doi.org/10.1109/ICASSP.2018.8462169
14. Malvar, H. S., Staelin, D. H. The LOT: Transform Coding Without Blocking Effects. IEEE Transactions on Acoustics, Speech, and Signal Processing, 1989, 37(4), 553-559. https://doi.org/10.1109/29.17536
15. Mariani, J. Language and Speech Processing. John Wiley & Sons, 2013.
16. Mirsamadi, S., Hansen, J. H. L. Multi-Domain Adversarial Training of Neural Network Acoustic Models for Distant Speech Recognition. Speech Communication, 2019, 106, 21-30. https://doi.org/10.1016/j.specom.2018.10.010
17. Nematollahi, M. A., Al-Haddad, S. A. R., Zarafshan, F. Blind Digital Speech Watermarking Based on Eigen-value Quantization in DWT. Journal of King Saud University – Computer and Information Sciences, 2015, 27(1), 58-67. https://doi.org/10.1016/j.jksuci.2014.03.012
18. Nikolić, J., Perić, Z. Lloyd-Max's Algorithm Implementation in Speech Coding Algorithm Based on Forward Adaptive Technique. Informatica, 2008, 19(2), 255-270
19. Ogunfunmi, T., Togneri, R., Narasimha, M. S. Speech and Audio Processing for Coding. Enhancement and Recognition. Springer, 2014. https://doi.org/10.1007/978-1-4939-1456-2
20. Patel, K. K., Fowler, M. L. Vector Quantizer Design for Speech Signal Compression. 7th International Conference on Cloud Computing, Data Science & Engineering – Confluence, 2017, Noida-India. https://doi.org/10.1109/CONFLUENCE.2017.7943249
21. Perić, Z., Aleksić, D., Stefanović, M., Nikolić, J. New Approach to Support Region Determination of the μ-law Quantizer. Elektronika ir Elektrotechnika, 2013, 19(8), 111-114. https://doi.org/10.5755/j01.eee.19.8.5406
22. Peric, Z., Simic, N., Savic, M. Analysis and Design of Two Stage Mismatch Quantizer for Laplacian Source. Elektronika ir Elektrotechnika, 2015, 21(3), 49-53. https://doi.org/10.5755/j01.eee.21.3.10380
23. Polap, D. Model of Identity Verification Support System Based on Voice and Image Samples. Journal of Universal Computer Science, 2018, 24(4), 460-474.
24. Polap, D. Neuro-Heuristic Voice Recognition. Proceedings of the Federated Conference on Computer Science and Information Systems, 11-14 September 2016, 487-490. https://doi.org/10.15439/2016F128
25. Polap, D. Human-Machine Interaction in Intelligent Technologies Using the Augmented Reality. Information Technology and Control, 2018, 47(4), 691-703. https://doi.org/10.5755/j01.itc.47.4.21602
26. Raid, A. M., Khedr, W. M., El-dosuky, M. A., Ahmed, W. Jpeg Image Compression Using Discrete Cosine Transform – A Survey. International Journal of Computer Science & Engineering Survey (IJCSES), 2014, 5(2), 39-47. https://doi.org/10.5121/ijcses.2014.5204
27. Sayood, K. Introduction to Data Compression, 3rd ed, 2005. https://doi.org/10.1016/B978-012620862-7/50006-7
28. Sinha, P. Speech Processing in Embedded Systems. Springer, 2009. https://doi.org/10.1007/978-0-387-75581-6
29. Tancic, M., Peric, Z., Simic, N., Tomic, S. Performace of Quasi-Logarithmic Quantizer for Discrete Input Signal. Information Technology and Control, 2017, 46(3), 395-402. https://doi.org/10.5755/j01.itc.46.3.16197
30. Tancic, M., Peric, Z., Tomic, S., Simic, N. Speech Signal Coding Using Forward Adaptive Quantization and Simple Transform Coding. Elektronika ir Elektrotechnika, 2016, 22(3), 74-77. https://doi.org/10.5755/j01.eie.22.3.15318

31. Thennattil, J. J., Mary L. Phonetic Engine for Continuous Speech in Malayalam. IETE Journal of Research, 2016, 62(5), 679-685. https://doi.org/10.1080/03772063.2016.1162673

32. Tomic, S., Peric, Z., Tancic, M., Nikolic, J. Backward Adaptive and Quasi-Logarithmic Quantizer for Sub-band Coding of Audio. Information Technology and Control, 2018, 47(1), 131-139. https://doi.org/10.5755/j01.itc.47.1.16190

33. Tu, J., Xu, J., Zhao, X. Research of Speech Amplitude Distribution Based on Hadamard Transformation. Xinlin, H. (Ed.) Machine Learning and Intelligent Communications. MLICOM 2016. Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering, Vol. 183, Springer. https://doi.org/10.1007/978-3-319-52730-7_27

34. Vaidelienė, G., Valantinas, J., Ražanskas, P. On the Use of Discrete Wavelets in Implementing Defect Detection System for Texture Images. Information Technology and Control, 2016, 45(2), 214-222. https://doi.org/10.5755/j01.itc.45.2.12654

35. Wang, Y., Li, J., Gong, Y. Small-Footprint High-Performance Deep Neural Network-Based Speech Recognition Using Split-VQ. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2015 ISBN: 978-1-4673-6997-8. https://doi.org/10.1109/ICASSP.2015.7178919

36. Xiao, X., Watanabe, S., Erdogan, H., Lu, L., Hershey, J., Seltzer, M. L., Chen, G., Zhang, Y., Mandel, M., Yu, D. Deep Beamforming Networks for Multi-Channel Speech Recognition. Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 20-25 March 2016, Shanghai, China, 5745-5749. https://doi.org/10.1109/ICASSP.2016.7472778

37. Zhang, Y., Chan, W., Jaitly, N. Very Deep Convolutional Networks for End-to-End Speech Recognition. Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 5-9 March 2017, New Orleans, LA, USA, 4845-4849. https://doi.org/10.1109/ICASSP.2017.7953077