[Downloaded from www.aece.ro on Thursday, July 03, 2025 at 19:28:02 (UTC) by 172.70.80.243. Redistribution subject to AECE license or copyright.]

Gaussian Source Coding using a Simple Switched Quantization Algorithm and Variable Length Codewords

Zoran PERIC[1], Goran PETKOVIC[2], Bojan DENIC[1], Aleksandar STANIMIROVIC[1], Vladimir DESPOTOVIC[3], Leonid STOIMENOV[1]
[1]*University of Nis, Faculty of Electronic Engineering, Aleksandra Medvedeva 14, 18000 Nis, Serbia*
[2]*College of Applied Studies, Filipa Filipovica 20,17500 Vranje, Serbia*
[3]*University of Luxembourg, Department of Computer Science, Avenue de la Fonte 6, L-4364 Esch-sur-Alzette, Luxembourg*
zoran.peric@elfak.ni.ac.rs

***Abstract*—This paper introduces an algorithm based on switched scalar quantization utilizing a novel μ-law quantization model (optimized in terms of minimal distortion) and variable length codewords, for high-quality encoding of the signals modeled by Gaussian distribution. The implemented μ-law quantizer represents an improvement of the standard μ-law quantizer in terms of bit rate, at the same time providing the equal signal quality. The main concept of the algorithm is to divide the range of the input signal variances into a certain number of sub-ranges, and to design the optimal quantizer for each sub-range. The signal is processed frame-by-frame, and for each frame the best performing quantizer is chosen, where the estimated frame variance is used as the switching criterion. Theoretical results indicate that the proposed algorithm achieves performance comparable to the standard μ-law quantizer, enabling the compression of about 0.5 bit/sample. The simulation results are provided to confirm the correctness of the proposed model.**

***Index Terms*—Gaussian distribution, quantization, source coding, signal processing algorithms, signal to noise ratio.**

I. INTRODUCTION

Scalar quantization is a process where real-valued samples generated by the input source are mapped to the representative levels, which form the quantization codebook [1−4]. Hence, the primary goal in the design of scalar (non-uniform) quantizer is to define the adequate codebook for the particular problem under consideration. During the design process, the information source to be quantized is often modeled by a probability density function, and the codebook is specified according to a performance criterion. If the criterion is the minimal mean-squared error (MSE) distortion, in order to design the optimal codebook for the given input distribution one can use the iterative Lloyd-Max's algorithm [1−5]. However, if the codebook size is large, the former becomes impractical and as an alternative the companding quantization has been developed.

Companding quantization is a low-complex non-iterative method for the non-uniform scalar quantizer design. The structure encompasses compressor, uniform quantizer and expander connected in cascade [1−4]. Various companding models have been proposed in the literature [6−10], suitable for particular applications. Depending on the employed compression function they can be classified into the optimal compandors [6], [7] or logarithmic compandors [8−10]. The main advantage of the optimal compandor is its ability to provide the signal quality (measured by SQNR (Signal to Quantization Noise Ratio)) close to the optimal Lloyd-Max's quantizer; however its robustness remains limited, i.e. it does not provide constant SQNR over the broad variance range. On the other hand, logarithmic compandor is robust at the expense of lower maximal SQNR.

In view of practical implementation, the logarithmic compandors, due to the robustness feature, are a more adequate choice for quantization of time-varying signals (e.g. speech or audio) than the optimal compandors or Lloyd-Max's quantizers. Specifically, the logarithmic compandors use μ-law and A-law compression functions. In addition, logarithmic quantization has been applied in various research fields, including speech coding [11], image compression [12], OFDM (Orthogonal Frequency Division Multiplexing) systems [13], multi-agent systems [14], analog-to-digital converters [15] or neural networks [16].

The emphasis in this paper is on logarithmic μ-law quantization and its application in an algorithm based on switched quantization, for encoding a Gaussian source used to model real signals such as speech [1], [17], OFDM signal [18] or weights in neural networks [19]. The main idea is to develop a novel model of μ-law quantizer optimized in the sense of minimal MSE distortion, able to outperform the standard μ-law quantizer (with the same number of quantization levels) in terms of bit rate, while preserving the same signal quality. This is achieved by dividing the support region of the quantizer into two sub-regions having different number of levels, which are encoded using codewords of different length. The implementation of the novel model into the switched quantization algorithm enabled us to obtain the system equivalent to variable length encoder; note that variable length coding is widely used in reducing the bit rate [1−3]. Note also that switched quantization technique is commonly employed for improving the performance of the single quantizer in a wide dynamic range, which is important for processing of time-varying signals [1], [20−22]. Hence, with this algorithm we are able to efficiently process different Gaussian inputs at smaller bit

This work has been supported by the Ministry of Education, Science and Technological Development of the Republic of Serbia and by the Science Found of the Republic of Serbia (Grant No. 6527104, AI-Com-in-AI).

Digital Object Identifier 10.4316/AECE.2020.04002

1582-7445 © 2020 AECE

[Downloaded from www.aece.ro on Thursday, July 03, 2025 at 19:28:02 (UTC) by 172.70.80.243. Redistribution subject to AECE license or copyright.]

rates. A comprehensive theoretical analysis is provided and supported by simulation to verify the developed theory.

The remaining of the paper is organized as follows: in Section 2 we describe the design of the classical and the proposed μ-law quantizer. Section 3 introduces and analyses the application of proposed μ-law quantizer in switched quantization algorithm. In Section 4 the results of simulation are presented. Finally, concluding remarks are provided in Section 5.

II. Logarithmic Scalar Quantization−Design Methods

Logarithmic scalar quantization is commonly realized using a cascade connection of compressor, uniform quantizer, and expander [1−4]. The non-uniform quantization can be achieved by compression of the input signal x using the compressor with a nonlinear characteristic $c(\cdot)$, followed by quantization of the compressed signal $c(x)$ using a uniform quantizer, and finally by expanding the quantized values of the compressed signal using the nonlinear inverse compression characteristic $c^{-1}(\cdot)$.

A. Design of Standard μ-Law Quantizer

In logarithmic μ-law companding quantization a compression function is characterized by the following expression [1−4]:

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

where $x_{\max}$ is the upper limit of the compressor range and also defines the upper threshold of quantizer support (granular) region and μ is the compression factor. The granular region of the quantizer is defined in the range $[-x_{\max}, x_{\max}]$, while the overload region is defined in the range $(-\infty, -x_{\max}) \cup (x_{\max}, \infty)$.

Let us denote by N the number of levels of the non-uniform μ-law quantizer, by $x_{u,i}$ the decision thresholds and by $y_{u,i}$ the representative levels of the corresponding uniform quantizer, defined as:

$$x_{u,i} = -x_{\max} + i\Delta,\ i = 0,1,...,N, \tag{2}$$

$$y_{u,i} = -x_{\max} + \left(i-\frac{1}{2}\right)\Delta,\ i = 1,...,N, \tag{3}$$

where $\Delta = 2 \cdot x_{\max} / N$ is the parameter known as step size [1−4]. Then, the parameters (i.e. decision thresholds x_i and representative levels y_i) of the equivalent non-uniform quantizer can be obtained from the conditions:

$$c(x_i) = x_{u,i} \Rightarrow x_i = c^{-1}(x_{u,i}), \tag{4}$$

$$c(y_i) = y_{u,i} \Rightarrow y_i = c^{-1}(y_{u,i}), \tag{5}$$

which gives:

$$x_i = \frac{x_{\max}}{\mu}\left((1+\mu)^{\frac{x_{u,i}\operatorname{sgn}(x_{u,i})}{x_{\max}}} - 1\right)\operatorname{sgn}(x_{u,i}),\ i = 0,...,N, \tag{6}$$

$$y_i = \frac{x_{\max}}{\mu}\left((1+\mu)^{\frac{y_{u,i}\operatorname{sgn}(y_{u,i})}{x_{\max}}} - 1\right)\operatorname{sgn}(y_{u,i}),\ i = 1,...,N, \tag{7}$$

by which the μ-law quantizer is completely defined.

The quantization rule is simple: $Q(x) = y_i$, if $x_{i-1} < x \le x_i$, $i = 1,\ldots, N$; Q denotes the quantization operation and x is the input signal value. Throughout this paper, we assume that the signal at the input of the quantizer is modeled by a memoryless Gaussian source with zero-mean having the probability density function (pdf) [1−4]:

$$p(x,\sigma) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{x^2}{2\sigma^2}\right), \tag{8}$$

where σ^2 is the variance of the signal. Recall that due to the symmetry of a given pdf, the quantizer will be symmetric as well, implying $x_{-i} = x_i$, $i = 0,\ldots, N/2$, $y_{-i} = y_i$, $i = 1,\ldots, N/2$.

Two parameters that characterize the scalar quantizer are the bit rate and MSE distortion [1−4]. For this specific case, the representative levels are encoded using fixed-length code words. If N is a power of 2, then the bit rate is determined as $R = \log_2 N$ bit/sample; else the bit rate is specified as $R = \lceil \log_2 N \rceil$ bit/sample where $\lceil x \rceil$ is the nearest integer higher than x.

MSE distortion D (the expected mean-square error between the original and the quantized signal) can be represented as the sum of the distortions inserted in granular (D_g) and overload (D_{ov}) parts, i.e. $D = D_g + D_{ov}$. Since we consider a high-resolution quantization (N is large), then the Bennett's integral can be used to estimate the granular distortion [1]:

$$D_g(\sigma) = \frac{x_{\max}^2}{3N^2} \int_{-x_{\max}}^{x_{\max}} \frac{p(x,\sigma)}{\left[\frac{\partial c(x)}{\partial x}\right]^2} dx, \tag{9}$$

whereas overload distortion can be estimated using [1]:

$$D_{ov}(\sigma) = 2\int_{x_{\max}}^{\infty} (x - x_{\max})^2\, p(x,\sigma)\, dx. \tag{10}$$

For the Gaussian pdf in (8) and compression function in (1), the following expressions can be obtained for components of the MSE distortion [21], [22]:

$$D_g(\sigma) = \sigma^2\left[\frac{\ln^2(1+\mu)}{3N^2}\left(1+\frac{c^2}{\mu^2}+\frac{2c}{\mu}\sqrt{\frac{2}{\pi}}\right)\right], \tag{11}$$

$$D_{ov}(\sigma) = \sigma^2\left[\left(1+c^2\right)\left(1-erf\left(\frac{c}{\sqrt{2}}\right)\right) - \sqrt{\frac{2}{\pi}}c\exp\left(-\frac{c^2}{2}\right)\right], \tag{12}$$

where $c = x_{\max} / \sigma$ is the loading factor of the μ-law quantizer and $erf(\cdot)$ is the error function.

Based on distortion, SQNR (objective measure of the quantized signal quality) can be evaluated as [1−4]:

$$\mathrm{SQNR}(\sigma) = 10\log_{10}\left(\frac{\sigma^2}{D(\sigma)}\right). \tag{13}$$

As eqs. (11)−(13) show, for a given σ, the performance of the quasi-logarithmic quantizer is highly dependent on c (i.e. $x_{\max}$) and μ. Let us recall that a proper selection of parameters of the scalar quantizer is of great significance, especially of the upper support region threshold [23], [24]. These two parameters are determined according to:

$$\frac{\partial D}{\partial c} = 0 \Rightarrow c = c_{opt},\quad \frac{\partial D}{\partial \mu} = 0 \Rightarrow \mu = \mu_{opt}, \tag{14}$$

i.e. such that minimal MSE distortion is ensured.

B. Design of the Proposed μ-Law Quantizer

The novel N-level μ-law quantizer model proposed in this paper is obtained by modifying the standard N-level μ-law quantizer described in Section II-A, with a goal to achieve a reduction of the bit rate, while retaining the same SQNR.

The support region of the proposed N-level μ-law

[Downloaded from www.aece.ro on Thursday, July 03, 2025 at 19:28:02 (UTC) by 172.70.80.243. Redistribution subject to AECE license or copyright.]

quantizer is divided into two intervals denoted as $I_1 = (-x_d, x_d)$ and $I_2 = (-x_{\max}, -x_d) \cup (x_d, x_{\max})$, where $x_{\max}$ is the upper support region threshold, and x_d is the boundary between these two intervals specified by the condition:

$$c(x_d) = \frac{1}{3} x_{\max}, \tag{15}$$

which gives:

$$x_d = \frac{x_{\max}}{\mu}\left(\sqrt[3]{1+\mu} - 1\right). \tag{16}$$

where $c(\cdot)$ is the compression function defined by (1).

The main idea behind this approach is to employ a different number of quantization levels in the particular interval, so that codewords of different length are produced during encoding, i.e. to obtain a system equivalent to the variable length encoder. Variable length coding (or entropy coding) is a popular method for reducing the bit rate (i.e. compression), where the shorter codewords are employed for high probable symbols (levels) and vice-versa [1−4]. Thus, within the interval I_1, $N_1 = 2^{R_1}$ levels are placed, and samples belonging to this interval are quantized to the nearest levels and encoded with R_1 bits. Similarly, $N_2 = 2^{R_2}$ levels are placed within the interval I_2, and the corresponding samples are quantized to the nearest levels in that interval and encoded with R_2 bits. Here, the following is valid: $R_1 < R_2$ which implies $N_1 < N_2$ and $N = N_1 + N_2$ (observe that N is not a power of 2).

Assuming the symmetry of the proposed N-level quantizer, only the positive part of the characteristic can be observed. Obviously, by specifying the representative levels and the decision thresholds the quantizer is uniquely defined. The parameters of the uniform quantizer in interval I_1 are defined as:

$$x_{u,i}^{I_1} = i\Delta_1,\ i = 0,1,...,N_1/2, \tag{17}$$

$$y_{u,i}^{I_1} = \left(i - \frac{1}{2}\right)\Delta_1,\ i = 1,...,N_1/2, \tag{18}$$

where $\Delta_1 = 2 \cdot x_{\max} / 3N_1$. Hence, the decision thresholds and representative levels can be calculated respectively by:

$$x_i^{I_1} = \frac{x_{\max}}{\mu}\left((1+\mu)^{\frac{x_{u,i}^{I_1}}{x_{\max}}} - 1\right),\ i = 0,1,...,N_1/2, \tag{19}$$

$$y_i^{I_1} = \frac{x_{\max}}{\mu}\left((1+\mu)^{\frac{y_{u,i}^{I_1}}{x_{\max}}} - 1\right),\ i = 1,...,N_1/2. \tag{20}$$

The corresponding uniform quantizer in interval I_2 is defined as:

$$x_{u,i}^{I_2} = \frac{x_{\max}}{3} + i\Delta_2,\ i = 0,1,...,N_2/2, \tag{21}$$

$$y_{u,i}^{I_2} = \frac{x_{\max}}{3} + \left(i - \frac{1}{2}\right)\Delta_2,\ i = 1,...,N_2/2, \tag{22}$$

where $\Delta_2 = 4 \cdot x_{\max} / 3N_2$. Therefore, the decision thresholds and representative levels are calculated respectively by:

$$x_i^{I_2} = \frac{x_{\max}}{\mu}\left((1+\mu)^{\frac{x_{u,i}^{I_2}}{x_{\max}}} - 1\right),\ i = 0,1,...,N_2/2, \tag{23}$$

$$y_i^{I_2} = \frac{x_{\max}}{\mu}\left((1+\mu)^{\frac{y_{u,i}^{I_2}}{x_{\max}}} - 1\right),\ i = 1,...,N_2/2. \tag{24}$$

Input signal sample x is quantized (encoded) in the following way. First, the interval to which x belongs (I_1 or I_2) is determined. Information about this is encoded with the codeword '**1**' if $x \in I_1$, otherwise it is encoded with codeword '**0**'. If $x \in I_1$, then x is quantized to the nearest representative level from I_1 and encoded with R_1 bits; otherwise x is encoded with R_2 bits. Note that codewords having lengths of R_1 or R_2 bits carry information about both the sign and magnitude of the sample. Since codewords of different length (R_1+1 or R_2+1 bits) can be generated at the output, the quantizer we propose is equivalent to the one utilizing variable length codewords.

In this paper, we consider the special case when $N_2 = 2 \cdot N_1$, i.e. $R_2 = R_1 + 1$, which implies $N = 3 \cdot N_1$. This also implies that $\Delta_1 = \Delta_2 = 2 \cdot x_{\max} / 3 \cdot N_1 = 2 \cdot x_{\max} / N = \Delta$ (see Section II-A). Under these circumstances the representational levels and decision thresholds are exactly the same as in the standard μ-law quantizer (same uniform quantizer with step size Δ is applied). If the critical design values, c and μ, are the same as in the standard μ-law quantizer, the proposed quantizer provides the same SQNR. On the other hand, the average bit rate can be defined as:

$$R_{av} = P_1(R_1 + 1) + (1 - P_1)(R_2 + 1), \tag{25}$$

where P_1 denotes the probability that the input sample belongs to I_1:

$$P_1 = P(x \in I_1) = \int_{-x_d}^{x_d} p(x,\sigma)\,dx = erf\left(\frac{x_d}{\sqrt{2}\sigma}\right), \tag{26}$$

and $R_2 = R_1 + 1$. Eventually, the average bit rate is obtained as:

$$\begin{aligned} R_{av}(\sigma) &= erf\left(\frac{x_d}{\sqrt{2}\sigma}\right)(R_1 + 1) + \left(1 - erf\left(\frac{x_d}{\sqrt{2}\sigma}\right)\right)(R_1 + 2) \\ &= R_1 + 2 - erf\left(\frac{x_d}{\sqrt{2}\sigma}\right). \end{aligned} \tag{27}$$

C. Performance analysis

The design is done for the reference variance $\sigma_0^2 = 1$, which is the standard approach in scalar quantization [1−4].

TABLE I. PERFORMANCE OF THE STANDARD AND THE PROPOSED QUASI-LOGARITHMIC *M*-LAW QUANTIZER FOR UNIT VARIANCE CASE

$N = 96$	c_{opt} ($x_{\max}^{\text{opt}}$)	μ_{opt}	SQNR [dB]	R [bit/sample]
Standard	3.74	3.54	34.75	7
Proposed	3.74	3.54	34.75	6.49

Table I shows the performance (SQNR and bit rate) for the standard (N = 96) and the proposed μ-law quantizer (N = 96, $N_1 = 2^5 = 32$, $N_2 = 2^6 = 64$). The key parameters c and μ are numerically specified such to satisfy (14). It can be seen that the proposed μ-law quantizer provides the same SQNR as the baseline, but approximately 0.5 bit/sample lower bit rate.

Fig. 1 illustrates the compression function used in the design of the MSE optimal μ-law quantizer with N = 96 levels, where the border (x_d = 0.69) separating the established intervals (I_1 and I_2) is shown as well.

Further, the MSE optimal μ-law quantizer with N = 96 levels is applied in quantization of Gaussian signals whose variance differs from the designed one ($\sigma^2 \neq \sigma_0^2 = 1$), i.e. the variance mismatched scenario [25]. SQNR and bit rate across wide range of the input signal variances ((-30dB, 30dB) in relation to σ_0^2) are plotted in Figs. 2 and 3, respectively.

[Downloaded from www.aece.ro on Thursday, July 03, 2025 at 19:28:02 (UTC) by 172.70.80.243. Redistribution subject to AECE license or copyright.]

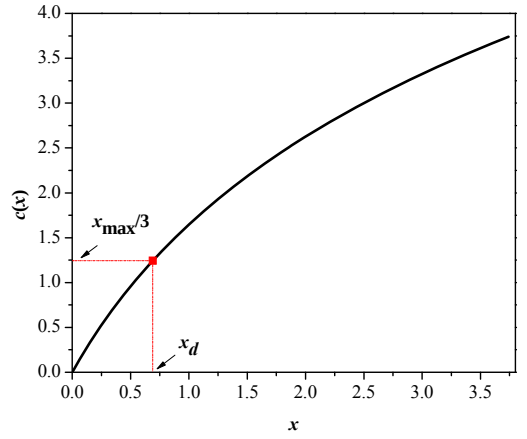

Figure 1. Compression function defined by (1) used in design of the optimal μ-law quantizer with $N = 96$ levels ($x_{max} = 3.74$ and $\mu = 3.54$)

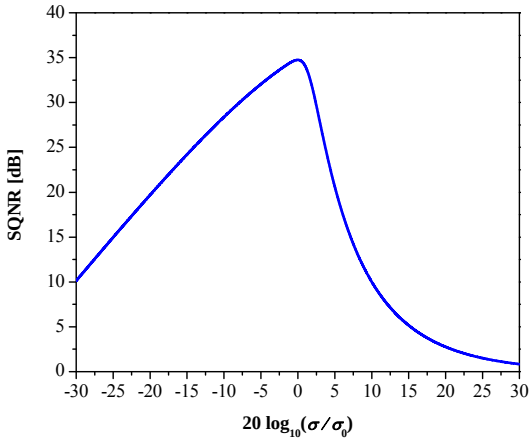

Figure 2. SQNR as a function of the input signal variance of the μ-law quantizer with $N = 96$ levels ($N_1 = 32$, $N_2 = 64$, $c = 3.74$, $\mu = 3.54$, $R = 6.49$ bit/sample)

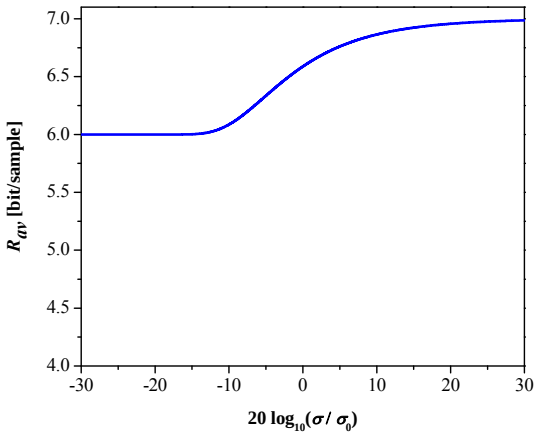

Figure 3. The average bit rate as a function of the input signal variance of the μ-law quantizer with $N = 96$ levels ($N_1 = 32$, $N_2 = 64$, $c = 3.74$, $\mu = 3.54$)

Considering Fig. 2, one can conclude that such quantizer is not efficient enough, since the desired SQNR (see Table 1) is achieved in a very small range of the input variances (near the designed variance, 0 dB point in log scale), while it significantly decreases in the rest of the range. In addition, average bit rate is also not constant across the range (Fig. 3), which is out of interest. In order to improve the performance over the entire variance range (i.e. to achieve approximately constant SQNR and bit rate) the switched quantization technique is employed.

III. The Proposed Algorithm Based on Switched Quantization Technique

A. Algorithm Description

Switched quantization technique is an effective way for achieving the robustness (expressed by approximately constant SQNR in a wide dynamic range), especially required in processing of non-stationary signals [1], [20−22]. According to this technique, the range of the input signal variances is divided into a certain number of sub-ranges, and for each sub-range the quantizer is designed such that it achieves acceptable performance. Hence, in switched quantization there is a set of quantizers at disposal, and switching among them is done with respect to the estimated variance of the input signal. The quantizer developed in Section II-B is implemented in the switched quantization model, and the appropriate analysis is given. The algorithm processes the input signal frame-by-frame according to the following steps:

Step 1. Frame buffering. Input signal frame with the finite length, $x_j(n)$, $n = 1,\ldots,M$, $j = 1,\ldots, F$, where M denotes the frame size, j stands for the frame index and F represents a total number of frames, is stored within the buffer.

Step 2. Variance calculation. For the j-th signal frame, the variance σ_j^2 is estimated as [1], [26]:

$$\sigma_j^2 = \frac{1}{M}\sum_{n=1}^{M} x_j^2(n),\ j = 1,\ldots,F\,. \tag{28}$$

Step 3. Classification of the calculated variance. The dynamic range of the signal frame variances ($V_{\min}[\text{dB}] = 20\cdot\log_{10}\sigma_{j,\min}$, $V_{\max}[\text{dB}] = 20\cdot\log_{10}\sigma_{j,\max}$) is uniformly divided into k sub-ranges, whose borders are specified as:

$$x_{\sigma_i} = V_{\min} + i\Delta_k,\ i = 0,1,\ldots,k\,, \tag{29}$$

where $\Delta_k[\text{dB}] = (V_{\max} - V_{\min}) / k$ is the width of the sub-range. For each sub-range $\left[x_{\sigma_i}, x_{\sigma_{i+1}}\right]$, one variance value is chosen as the representative one and determined as the midpoint:

$$y_{\sigma_i} = V_{\min} + \left(i - \frac{1}{2}\right)\Delta_k,\ i = 1,\ldots,k\,. \tag{30}$$

Finally, after the estimated variance σ_j^2 is classified into one of k available sub-ranges, it is mapped to the appropriate representative point for that interval. The described process is equivalent to the log-uniform quantization [26].

Step 4. Design of the quantizer for a sub-range. The parameter determined in the Step 3 is further used to design the employed μ-law quantizer (Section II-A) for the i-th sub-range denoted by Q_i, $i = 1,\ldots, k$, as follows:

$$x_d^i = g\cdot x_d(\sigma_0)\,, \tag{31}$$

$$x_{\max}^i = g\cdot x_{\max}(\sigma_0)\,, \tag{32}$$

where

$$g = 10^{\frac{y_{\sigma_i}}{20}}\,, \tag{33}$$

$x_d = x_d(\sigma_0)$ is given by (16), and $x_{\max} = x_{\max}(\sigma_0)$ is given in Table I. Observe that that the design parameter μ is independent of σ.

[Downloaded from www.aece.ro on Thursday, July 03, 2025 at 19:28:02 (UTC) by 172.70.80.243. Redistribution subject to AECE license or copyright.]

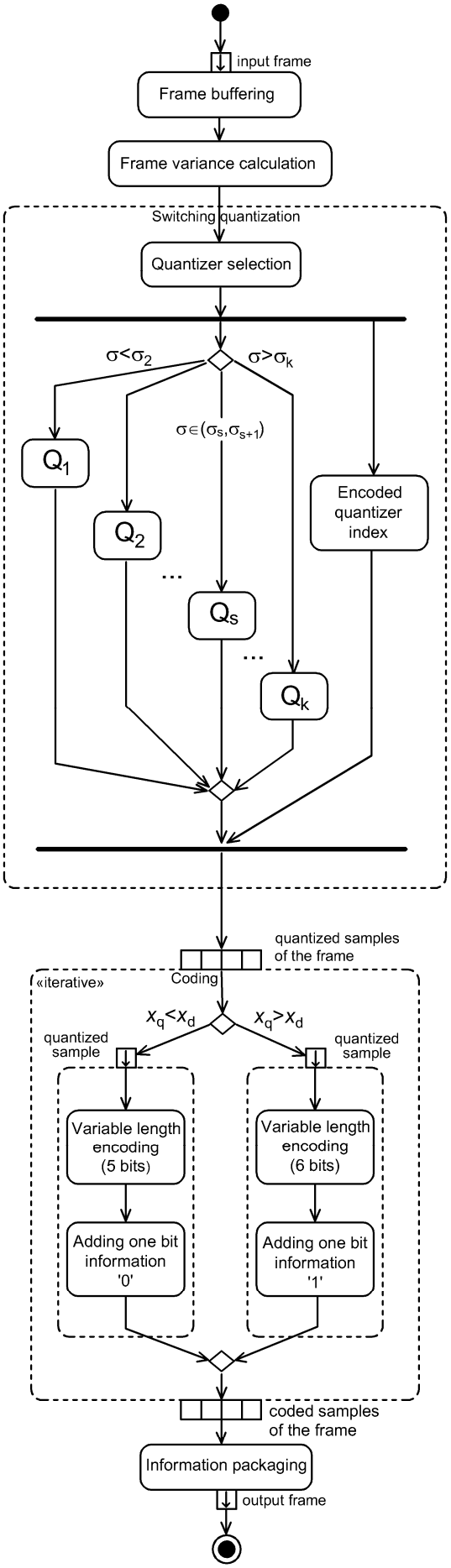

Figure 4. Flowchart of the proposed algorithm based on switched quantization technique

As there are k quantizers in total, the index of employed quantizer has to be encoded with $\log_2 k$ bits and sent to the decoder once per each frame.

Step 5. Coding (quantization) process. When the quantizer Q_i for the j-th frame is selected, the proposed algorithm chooses one of two available codeword lengths to encode an individual frame sample. In particular, the sample value $x_j(n)$ is compared to the threshold x_d^i, and the algorithm makes the decision which codeword should be generated. If $x_j(n) < x_d^i$, the codeword consisted of $R_1 + 1$ bits is generated; otherwise the generated codeword consists of R_1+2 bits.

Step 6. Repeat the previously defined steps to process all frames.

Bit rate and SQNR are used to evaluate the performance of the algorithm. Recall that the information about the quantizer index is required at the decoder part in addition to the encoded samples, leading to the total bit rate:

$$R_{tot} = R_{av} + \frac{\log_2 k}{M} \text{ [bit/sample]}, \tag{34}$$

where R_{av} is obtained by substituting (31) in (27) and the second term is the side information.

On the other hand, SQNR can be evaluated using eqs. (11)−(13) by substituting $x_{\max}$ with $x_{\max}^i$ defined in (32).

B. Theoretical Analysis of the Proposed Algorithm

In this subsection, performance of the proposed algorithm is analyzed, for the parameters of the quantizer adopted in Section II-C (N = 96, N_1 = 32 (R_1 = 5 bit/sample), N_2 = 64 (R_2 = 6 bit/sample), c = 3.74, μ = 3.54 and x_d = 0.69). The frame size is set to M = 160 samples.

We investigate the effect of a number of μ−law quantizers k (k = 2, 4, 8, 16, 32, 64 and 128, i.e. k is a power of 2) on the switched algorithm performance measured by both SQNR and bit rate.

Fig. 5 demonstrates the SQNR obtained using the proposed switched algorithm over the broad range of input variances when k varies, where for the purpose of better visibility only the cases k = 2, k = 8, k = 32 and k = 128 are shown. We observe that SQNR is periodic with total number of periods equals to k. In order to provide better insight into SQNR performance, let us further introduce the following parameters:

$$\text{SQNR}^{\text{dynamic}} = \text{SQNR}_{\max} - \text{SQNR}_{\min}, \tag{35}$$

$$\text{SQNR}^{\text{average}} = \frac{1}{m}\sum_{i=1}^{m} \text{SQNR}(\sigma_i), \tag{36}$$

Note that $\text{SQNR}^{\text{dynamic}}$ represents the difference between the theoretical maximal ($\text{SQNR}_{\max}$) and minimal ($\text{SQNR}_{\min}$) value of SQNR achieved in the considered variance range of 60 dB width. $\text{SQNR}^{\text{average}}$ defines the average SQNR achieved in the considered range, where m = 2000 denotes the number of particular variances σ_i^2 in that range.

Fig. 6 plots the $\text{SQNR}^{\text{dynamic}}$ and $\text{SQNR}^{\text{average}}$ versus $\log_2 k$. Observe that highest $\text{SQNR}^{\text{dynamic}}$ and the lowest $\text{SQNR}^{\text{average}}$ values are achieved for $\log_2 k$ = 1, i.e. k = 2; in that scenario the lower performance bound is achieved by the proposed algorithm. As k increases, $\text{SQNR}^{\text{dynamic}}$ decreases whereas $\text{SQNR}^{\text{average}}$ increases, improving the algorithm performance.

[Downloaded from www.aece.ro on Thursday, July 03, 2025 at 19:28:02 (UTC) by 172.70.80.243. Redistribution subject to AECE license or copyright.]

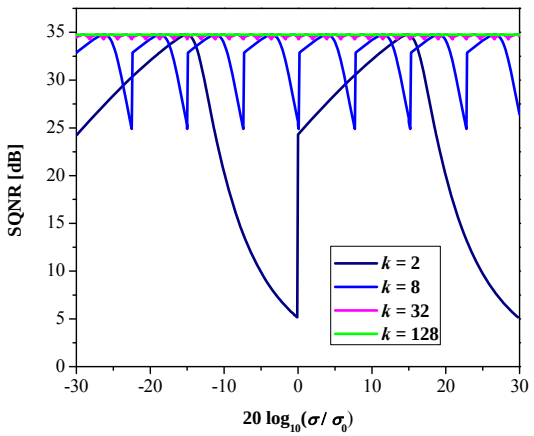

Figure 5. Theoretical results: SQNR vs. input signal variances of the proposed switched algorithm (N = 96, N_1 = 32, N_2 = 64, c = 3.74, μ = 3.54 and x_d = 0.69) when the number of quantizers k varies

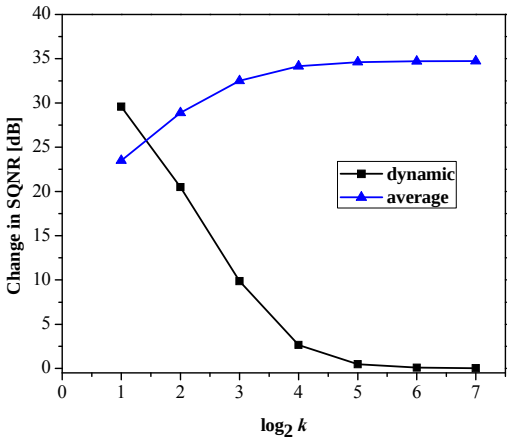

Figure 6. Theoretical results: $\text{SQNR}^{\text{dynamic}}$ and $\text{SQNR}^{\text{average}}$ vs. the number of quantizers k

When k is sufficiently high, $k \geq 32$ ($\log_2 k \geq 5$), $\text{SQNR}^{\text{dynamic}}$ tends to zero (the SQNR curve becomes flat over the entire range, as Fig. 5 shows) while $\text{SQNR}^{\text{average}}$ tends to maximal SQNR given in Table I, which is an upper performance bound of the proposed algorithm. Furthermore, Fig. 8 suggests that for $k > 32$ $\text{SQNR}^{\text{average}}$ curve saturates, which means that there is no need to further increase the number of quantizers since the contribution to the $\text{SQNR}^{\text{average}}$ is negligible.

The results for the total bit rate (eq. (34)) in the dynamic range established above are presented in Fig. 7. The parameters R_{tot}^{dynamic} and R_{tot}^{average} are given by eqs. (37) and (38), respectively:

$$R_{tot}^{\text{dynamic}} = R_{tot,\max} - R_{tot,\min}, \tag{37}$$

$$R_{av}^{\text{average}} = \frac{1}{m}\sum_{i=1}^{m} R_{av}(\sigma_i), \tag{38}$$

They are plotted in Fig. 8 versus $\log_2 k$. $R_{tot,\min}$ and $R_{tot,\max}$ in (37) denote the minimal and maximal values of the total bit rate in the observed range. Fig. 8 reveals a small impact of k to total bit rate, since R_{tot}^{average} slightly increases with k (it has the values near to 6.49 bit/sample presented in Table I). Based on the detailed analysis for both SQNR and bit rate, we can accept k = 32 as the optimal number of quantizers for implementation of the proposed switched algorithm.

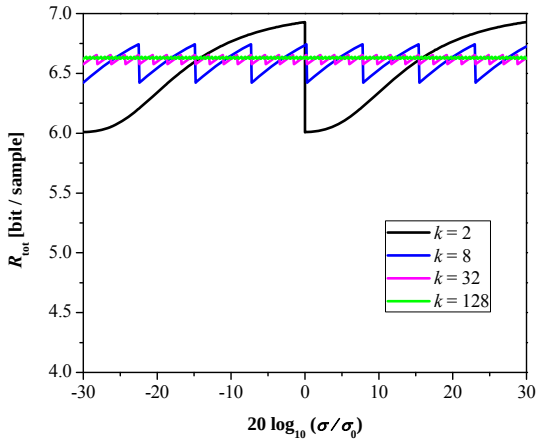

Figure 7. Theoretical results: Total bit rate vs. input signal variances of the proposed switched algorithm (N = 96, N_1 = 32, N_2 = 64, c = 3.74, μ = 3.54 and x_d = 0.69) when the number of quantizers k varies

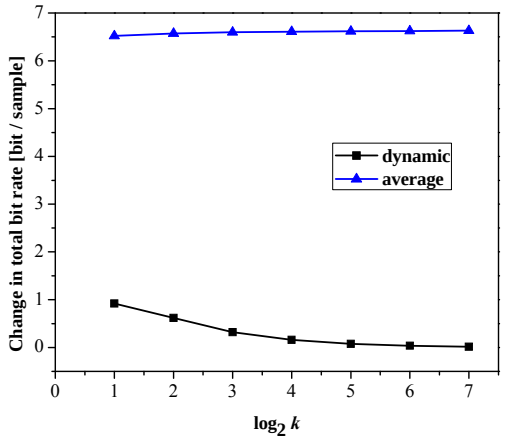

Figure 8. Theoretical results: R_{tot}^{dynamic} and R_{tot}^{average} vs. the number of quantizers k

Table II shows the bit allocation per frame with the length of 160 samples for the proposed algorithm and k = 32.

TABLE II. BIT ALLOCATION PER SIGNAL FRAME OF 160 SAMPLES

Parameter	Bits/frame
Quantizer index	5
Interval selection	1
Selected interval	**1** 6 bits × 160 samples **0** 7 bits × 160 samples

IV. SIMULATION RESULTS AND DISCUSSION

To verify the previously obtained theoretical results the simulation process is conducted. The simulations are performed in MATLAB and the block diagram is depicted in Fig. 9. The input signal was the Gaussian random variable with zero-mean and variance σ_i [dB] $\in$ [$20\cdot\log_{10}\sigma_{\min}$, $20\cdot\log_{10}\sigma_{\max}$] = [−30 dB, 30 dB], i = 1, …, 2000.

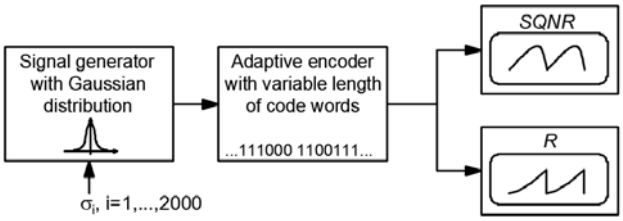

Figure 9. Block diagram of the simulation process

[Downloaded from www.aece.ro on Thursday, July 03, 2025 at 19:28:02 (UTC) by 172.70.80.243. Redistribution subject to AECE license or copyright.]

For each σ_i, 400000 values are generated and the frame length is set to $M = 160$ (total number of frames is $F = 2500$). Specifically, we consider the switched algorithm with the proposed μ-law quantizer with the same settings as in Section II-C and $k = 32$.

To provide simulation value of SQNR for the particular variance σ_i we use the relation:

$$SQNR_S(\sigma_i) = \frac{1}{F}\sum_{j=1}^{F} 10\log_{10}\left(\frac{\frac{1}{M}\sum_{n=1}^{M} x_n^2}{\frac{1}{M}\sum_{n=1}^{M}\left(x_n - \hat{x}_n\right)^2}\right), \quad (39)$$

where x_i are the input samples for a Gaussian input with the variance σ_i^2, and $\hat{x}_i$ are the quantized samples. In addition, the average bit rate is determined according to:

$$R_S(\sigma_i) = \frac{1}{F}\sum_{j=1}^{F}\left(\frac{1}{M}\sum_{n=1}^{M} R(x_n)\right), \quad (40)$$

where $R(x_n)$ denotes the bit rate used for the sample x_n.

The simulation values of SQNR in case of single μ-law quantizer ($N_1 = 32$, $N_2 = 64$, $c = 3.74$, $\mu = 3.54$) are illustrated in Fig. 10. Figs. 11 and 12 shows the values of SQNR and bit rate, respectively, in case of the proposed switched algorithm ($N_1 = 32$, $N_2 = 64$, $c = 3.74$, $\mu = 3.54$, $k = 32$).

Let us further compare the theoretical and simulation results. To this end we use Tables III and IV where the values of two assessment measures, SQNR and bit rate, are summarized.

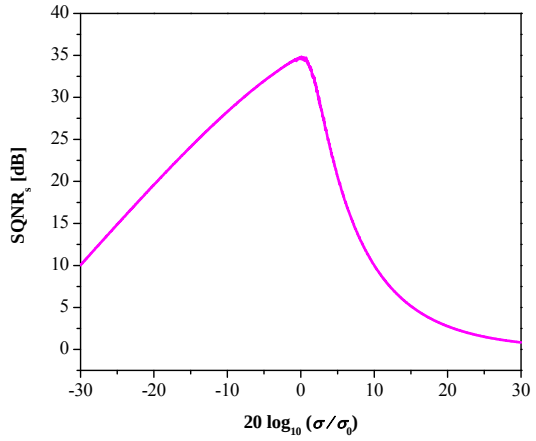

Figure 10. Simulation results: SQNR dependence on the input signal variance for the proposed quantizer model ($N_1 = 32$, $N_2 = 64$, $c = 3.74$, $\mu = 3.54$)

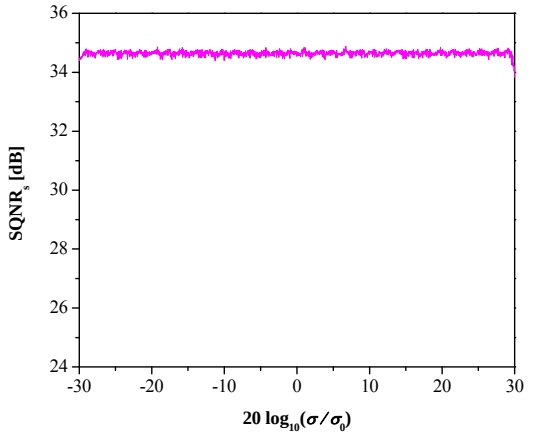

Figure 11. Simulation results: SQNR over the wide dynamic range of the input signal variances for the switched algorithm with the proposed quantizer ($N_1 = 32$, $N_2 = 64$, $c = 3.74$, $\mu = 3.54$, $k = 32$)

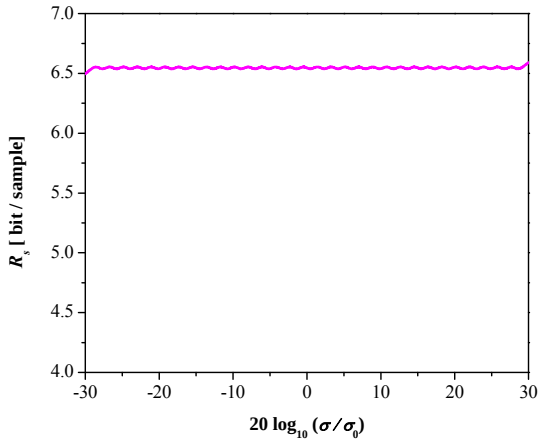

Figure 12. Simulation results: Bit rate in dependence on the input signal variance for the switched algorithm with the proposed quantizer ($N_1 = 32$, $N_2 = 64$, $c = 3.74$, $\mu = 3.54$, $k = 32$)

In particular, we consider minimum, maximum and average values of the respective measures attained in the established variance range.

TABLE III. COMPARISON OF THE THEORETICAL AND SIMULATION VALUES OF SQNR OF THE SWITCHED ALGORITHM ($N = 96$, $N_1 = 32$, $N_2 = 64$), $K = 32$, $M = 160$

Parameter	Theoretical results [dB]	Simulation [dB]
$SQNR_{min}$	34.27	33.82
$SQNR_{max}$	34.75	34.89
$SQNR^{average}$	34.65	34.65

TABLE IV. COMPARISON OF THE THEORETICAL AND SIMULATION VALUES OF THE TOTAL BIT RATE OF THE SWITCHED ALGORITHM ($N = 96$, $N_1 = 32$, $N_2 = 64$), $K = 32$, $M = 160$

Parameter	Theoretical results [bit/sample]	Simulation [bit/sample]
R_{min}	6.58	6.50
R_{max}	6.65	6.55
$R^{average}$	6.62	6.59

It can be seen that the simulation results are well matched with the theoretical ones, proving the correctness of the developed theoretical model.

V. CONCLUSION

In this paper, an efficient algorithm with μ-law quantizer and variable length codewords for high-rate encoding of time-varying signals described by Gaussian distribution has been proposed. The main advantage of this algorithm lies in the application of improved μ-law quantizer that offers better performance over the classical μ-law quantizer in terms of bit rate (savings in the bit rate up to 0.5 bit/sample has been provided), preserving the same signal quality and the quantizer complexity. The performance analysis of the switched algorithm when it includes a various number of quantizers has been conducted in a wide dynamic range of input signal variances, where the implementation with $k = 32$ quantizers has been adopted as the optimal one. In addition, simulations have been performed to confirm the validity of theoretical results. Based on the achieved promising results, the proposed algorithm can be considered as a good candidate for quantization of signals that follow the Gaussian pdf, e.g. weights in neural network.

[Downloaded from www.aece.ro on Thursday, July 03, 2025 at 19:28:02 (UTC) by 172.70.80.243. Redistribution subject to AECE license or copyright.]

ACKNOWLEDGMENT

This work has been supported by the Ministry of Education, Science and Technological Development of the Republic of Serbia and by the Science Found of the Republic of Serbia (Grant No. 6527104, AI- Com-in-AI).

REFERENCES

[1] N. S. Jayant, P. Noll, Digital Coding of Waveforms: Principles and Applications to Speech and Video. New Jersey, Prentice Hall, Chapter 4, pp. 115−188, 1984.

[2] K. Sayood, Introduction to Data Compression. San Francisco, Elsevier Science, Chapter 9, pp. 227−270, 2005. doi: 10.1016/B978-0-12-620862-7.X5000-7

[3] D. Salomon, "Variable-length codes for data compression,". London, Springer, Chapter 1, pp. 9−69, 2007. doi:10.1007/978-1-84628-959-0

[4] W. C. Chu, "Speech Coding Algorithms: Foundation and evolution of standardized coders," New Jersey, John Wiley & Sons, Chapter 5, pp. 143−158, 2003. doi:10.1002/0471668850

[5] Z. Peric, B. Denic, V. Despotovic, "Three-level delta modulation with second-order prediction for Gaussian source coding," Advances in Electrical and Computer Engineering, vol. 18, no. 3, pp. 73–78, 2018. doi:10.4316/AECE.2018.03017

[6] J. Nikolic, Z. Peric, A. Jovanovic, "Two forward adaptive dual-mode companding scalar quantizers for Gaussian source," Signal Processing, vol. 120, pp. 129–140, 2016. doi:10.1016/j.sigpro.2015.08.016

[7] J. Nikolic, Z. Peric, D. Aleksic, D. Antic, "Linearization of optimal compressor function and design of piecewise linear compandor for Gaussian source," Advances in Electrical and Computer Engineering, vol. 13, no. 4, pp. 73–78, 2013. doi:10.4316/AECE.2013.04013

[8] A. D. Lyon, "The μ-law CODEC," Journal of Object Technology, vol. 7, no. 8, pp. 17–31, 2008. doi:10.5381/jot.2008.7.8.c2

[9] S. Tomic, Z. Peric, J. Nikolic, "Modified BTC algorithm for audio signal coding," Advances in Electrical and Computer Engineering, vol. 16, no.4, pp. 31–38, 2016. doi:10.4316/AECE.2016.04005

[10] Z. Peric, M. Dincic, D. Denic, A. Jocic, "Forward adaptive logarithmic quantizer with new lossless coding method for Laplacian source," Wireless Personal Communications, vol. 59, pp. 625–641, 2010. doi:10.1007/s11277-010-9929-3

[11] G. K. Venayagamoorthy, W. Zha, "Comparison of nonuniform optimal quantizer designs for speech coding with adaptive critics and particle swarm," IEEE Transactions on Industry Applications, vol. 43, no. 1, pp. 238−244, 2007. doi:10.1109/TIA.2006.885897

[12] M. Rahali, H. Loukil, M. S. Bouhlel, "New image compression method using logarithmic quantization," in Proc. Int. Conf. on Information and Digital Technologies (IDT), Hammamet, Tunisia, 2016. doi:10.1109/SETIT.2016.7939909

[13] M. Mounir, M. B. El_Mashade, "On the selection of the best companding technique for PAPR reduction in OFDM systems," Journal of Information and Telecommunication, vol. 3, no. 3, pp. 400−411, 2019. doi:10.1080/24751839.2019.1606878

[14] Z. Ting, L. Junmin, "Robust iterative learning control of multi-agent systems with logarithmic quantizer," in Proc. 34th Chinese Control Conference (CCC), Hangzhou, China , 2015. doi:10.1109/chicc.2015.7260751

[15] M. Dincic, Z. Peric, D. Denic, Z. Stamenkovic, "Design of robust quantizers for low-bit analog-to-digital converters for Gaussian source," Journal of Circuits, Systems and Computers, vol. 28, no. supp01, 1940002, 2019. doi:10.1142/S0218126619400024

[16] T. Ueki, K. Iwai, T. Matsubara, T. Kurokawa, "Learning accelerator of deep neural networks with logarithmic quantization," in Proc. 7th Int. Congress on Advanced Applied Informatics (IIAI-AAI), Yonago, Japan, 2018. doi:10.1109/iiai-aai.2018.00133

[17] S. Gazor, W. Zhang, "Speech probability distribution," IEEE Signal Process. Letters, vol. 10, no. 7, pp. 204–207, 2003. doi:10.1109/LSP.2003.813679

[18] Y. Hou, G. Liu, Q. Wang, W. Xiang, "Performance optimization of digital spectrum analyzer with Gaussian input signal," IEEE Signal Processing Letters, vol. 20, no. 1, pp. 31–34, 2013. doi:10.1109/LSP.2012.2227255

[19] R. Banner, Y. Nahshan, E. Hoffer, D. Soudry, "Analytical clipping for integer quantization of neural networks," arXiv2018, arXiv: 1810.05723.

[20] G. Petkovic, Z. Peric, L. Stoimenov, "Switched scalar optimal μ-law quantization with adaptation performed to both the variance and the distribution of speech signal," Elektronika ir Elektrotechnika, vol. 22, no. 1, pp. 64−67, 2016. doi:10.5755/j01.eee.22.1.14111

[21] N. Vucic, Z. Peric, G. Petkovic, "Design of switched quantizers and speech coding based on quasi-logarithmic compandor," Elektronika Ir Elektrotechnika, vol. 24, no. 6, pp. 82−86, 2018. doi:10.5755/j01.eie.24.6.22295

[22] A. Mosic, Z. Peric, M. Savic, S. Panic, "Switched semilogarithmic quantization of Gaussian source with low delay," Elektronika ir Elektrotechnika, vol. 108, no. 2, pp. 71–74, 2011. doi:0.5755/j01.eee.108.2.148

[23] S. Na, D. Neuhoff, "On the support of MSE-optimal fixed–rate scalar quantizers," IEEE Transactions on Information Theory, vol. 47, no. 7, pp. 2972–2982, 2001. doi:10.1109/18.959274

[24] S. Na, "On the support of fixed-rate minimum mean-squared error scalar quantizers for a Laplacian source," IEEE Transactions on Information Theory, vol. 50, no. 5, pp. 937–944, 2004. doi:10.1109/TIT.2004.826686

[25] S. Na, "Asymptotic formulas for variance-mismatched fixed-rate scalar quantization of a Gaussian source," IEEE Transactions on Signal Processing, vol. 59, no. 5, pp. 2437–2441, 2011. doi:10.1109/TSP.2011.2112354

[26] J. Nikolic, Z. Peric, "Lloyd-Max's algorithm implementation in speech coding algorithm based on forward adaptive technique," Informatica, vol. 19, no. 2, pp. 255–270, 2008.