Switched Scalar Optimal µ-law Quantization with Adaptation Performed to both the Variance and the Distribution of Speech Signal

Goran M. Petkovic[1], Zoran H. Peric[2], Leonid V. Stoimenov[2]
[1]College of Applied Studies,
Filipa Filipovica 20, 17500 Vranje, Serbia
[2]Faculty of Electronic Engineering, University of Nis,
Aleksandra Medvedeva 14, 18000 Nis, Serbia
goran.petkovic@visokaskola.edu.rs

Abstract—**This paper proposes a new model of switching scalar quantization that besides the usual adaptation to the variance uses a switching technique for adaptation to the distribution of the speech signal source. By applying the new model a smaller quantization errors achieved in comparison to the conventional models with a usual adaptation.**

Index Terms—**Signal processing algorithms, adaptive quantization, speech processing, signal to noise ratio.**

I. INTRODUCTION

As the procedure of signal discretization per amplitude quantization is an integral part of almost all the systems for digital signal processing. Quantization process transforms continuous values of signal amplitude into discrete values of a predetermined set of amplitude values. Although during the decades of the development of digital signal processing many quantization techniques were developed, out of which many are described in the standards, there is a continuing need for new solutions in this area [1]–[5].

The speech signal is by nature a non-stationary process, which expresses its non-stationarity by changing parameters during the time. Since the parameters changes in time are slow, if the speech signal is observed in a short speech intervals, the process can be considered as stationary. Due to parameters changes of the speech signal in the time, it is recommended the use of adaptive quantizers that are during quantization adapted to local statistical characteristics of the source [1]–[4].

The most used adaptation in the process of quantization is adaptation to the variance of the speech signal. As it is presented, in a new model of quantization besides the usual adaptation, the adaptation is performed to the distribution of the amplitude of the speech signal sources. The result of the proposed model application is an improved quality of the reconstructed signal expressed by the ratio Signal-to-Quantization-Noise Ratio (SQNR).

After analysis of logarithmic quantization in the case of a source of the signal with the Laplacian and Gaussian distribution is exposed to the theoretical model of the new quantization techniques. After that, numeric results of

quantization of usual technique and new suggested model are compared on the example of real speech signals.

II. LOGARITHMIC QUANTIZATION LAW

The general form of compression characteristics can be defined by a number of different mathematical functions. It is necessary to provide the best compression characteristic of the reconstructed signal. Good selection of mathematical functions is logarithmic law by which small signal intensities are quantized with greater resolution or smaller steps of quantization in relation to a large signal intensity.

In order that the law of compression should satisfy practical application, the law of compression is modified so that the function passes through the point of origin. One of modification of the logarithmic law compression is a quasi-logarithmic ~ law described by function

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

where μ is a non-dimensional parameter that is called the factor of compression, and $x_{\max}$ the amplitude of the maximum load of quantizer which defines the scope of the scalar amplitude within the limits $[-x_{\max}\ x_{\max}]$.
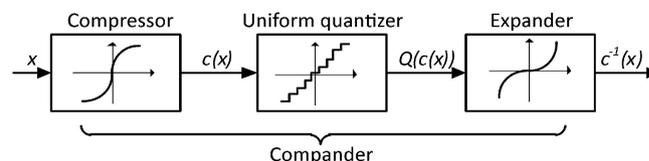

Fig. 1. The technique of non-uniform quantization.

Projection of non-uniform quantizer *i.e.* determining the codebook, is often based on companding technique by which non-uniform quantization is achieved by compression of the input signal x using a compressor with a non-linear characteristic $c(\bullet)$, then by the quantization of the compressed signal $c(x)$ applying uniform quantizer and finally by expanding the quantized values of the compressed signal using a non-linear inverse compression characteristics $c^{-1}(\bullet)$. The structure of the described non-uniform quantizer

that is made of regularly associated compressor, expander and a uniform quantizer is called compander [1] (Fig. 1).

III. DISTORTION FOR LOGARITHMIC QUANTIZATION LAW

Observe the source of the speech signal, which is described by the distribution of probability $p(x)$. N^{th} degrees of uniform scalar quantizer has got the distortion defined as the expected mean square error between the original and the quantized signal.

The total distortion D_t can be expressed as the sum of granular distortion D_g and overload distortion D_o

$$D_t = D_g + D_o. \tag{2}$$

Define the mean absolute value of the signal with

$$\overline{|x|} = 2 \int_0^{x_{max}} x p(x) dx, \tag{3}$$

and the mean square value of the signal with

$$\overline{x^2} = 2 \int_0^{x_{max}} x^2 p(x) dx. \tag{4}$$

If the source of the speech signal is represented by Laplacian distribution function of signal samples x, probability density function of the random variable x with variance is represented by the expression

$$p(x, \dagger) = \frac{\sqrt{2}}{2\dagger} \exp\left(-\frac{|x|\sqrt{2}}{\dagger}\right), \tag{5}$$

and total distortion by the expression [3]

$$D_t = \frac{\ln^2(1+\sim)}{3N^2} \dagger^2 \left[\frac{1}{\sim^2} \frac{x_{max}^2}{\dagger^2} + \frac{x_{max}}{\dagger} \frac{\sqrt{2}}{\sim} + 1\right] +$$
$$+ \dagger^2 \exp\left(-\frac{\sqrt{2}x_{max}}{\dagger}\right). \tag{6}$$

In order to simplify the analysis of quantizer it is usually introduced the assumption that the average value of the speech signal is zero, in which no generality is lost in the analysis. Following this assumption in the final expressions the average power of the input signal is replaced by its variance.

If the source of the speech signal is described by Gaussian probability distribution function

$$p(x) = \frac{1}{\sqrt{2f}\dagger} \exp\left(-\frac{x^2}{2\dagger^2}\right), \tag{7}$$

total distortion can be represented by the expression [4]

$$D_t = \dagger^2 \left(\frac{\ln^2(1+\sim)}{3N^2}\left[\frac{1}{\sim^2}c^2 + c\frac{2}{\sim}\sqrt{\frac{2}{f}} + 1\right] +$$

$$+ \left[(1+c^2)\left(1 - \text{erf}\left(\frac{c}{\sqrt{2}}\right)\right) - c\exp\left(-\frac{c^2}{2}\right)\sqrt{\frac{2}{f}}\right]\right). \tag{8}$$

Following the distortion it can be calculated Signal-to-Quantization-Noise Ratio (SQNR) source

$$SQNR = 10\lg\frac{\dagger^2}{D_t}. \tag{9}$$

IV. NEW MODEL OF SWITCHING SCALAR QUANTIZATION

One of the way to achieve constant SQNR is adaptive switched scalar quantization with codebook, with which it is determined to what statistical class from the set of k possible classes belong the samples of the frame of the speech signal. Frame samples are quantized with quantizer whose parameters are set per a certain statistical class, and to the receiver as additional information is transmitted index that identifies a class or codebook, which allows the reconstruction of the original signal. If n is the number of bits per samples used to encode, the size of the codebook is $N = 2^n$. Many models were developed from this class, which mutually differ in the way of adaptation. The most common way of adaptation is the adaptation to the variance of the signal in which the frame sample variance calculated and one from a set of quantizers is selected. The dynamic range of the assumed variance is divided into sub-ranges, and for each of them is designed a separate quantizer whose parameters are set so that the quantization error is minimal [2]–[4].

The probability density function of current values of the qualitatively recorded speech signal can be approximated by Laplacian distribution. On the other hand, by the analysis of the speech signal, it is concluded that the sound voices have large current value, while the voiceless sounds are represented by relatively small current values. Since the wave forms of unvoiced sounds are of stochastic nature (turbulence of airflow in the vocal tract speakers) as it is the noise of the system for recording and transmission of speech, it has got the impact on the fact that the dominant distribution is Gaussian. It can be concluded that sometimes Laplacian and sometimes Gaussian distribution is used for better description of speech signal, and due to the nature of speech and technical factors in processing and transmission, it can be expressed either to the entire signal or by segments of the speech signal.

Belonging of the speech signal, as per parts, to different distributions of the functions probability density, and the possibility of optimal setting of parameters μ characteristic to features distribution is in the fundament of the new proposed quantizer model. Besides the usual adaptation to the variance in the proposed model adaptation, it is performed the adaptation to the distribution of the speech signal [3].

The assumed dynamic range of variance B is divided into k range, and for each range are designed quantizer for variance $\dagger_p^2$, $p=1,...,k$ which are log-uniformly distributed in the dynamic range of variance $B = 20\log(_{max}/_{min})$ [6]

$$20\log \dagger_p = 20\log \dagger_{min} + \frac{(2p-1)B}{2k}, \qquad (10)$$

where $p = 1,...k$. For each of the k range two quantizers are designed: 1) quantizer which provides the smallest quantization error or minimum distortion if the assumed distribution of the amplitude of the signal source is described by Laplacian function and 2) quantizer which provides the smallest quantization error when assumed distribution of sources is described by Gaussian function.

While quantizers designing, the factor of relative range of the quantizer c value, which represents the ratio $x_{max} / \dagger_p$, and the non-dimensional factor ~ are determined so that the overall distortion is minimal. Optimal values for c and ~ are determined by numerical Nelder-Mead method for two variables function being optimized. In the Table I calculated optimal numerical values of the parameters c and ~ and SQNR for the number of quantization levels $N = 256$, for the case of Laplacian and Gaussian signal source are presented. In Fig. 2 is presented SQNR for switching scalar quantizer with adaptation to variance in case of source of Laplacian distribution function, and in Fig. 3 with Gaussian distribution function.

TABLE I. OPTIMAL PARAMETER VALUES c AND ~.

Source	c	μ	SQNR
Laplacian	8.69	15.72	40.51
Gaussian	4.20	4.54	42.83

The quantization technique in the proposed model processes incoming samples per frames. The procedure of signal processing starts with buffering. After buffering j^{th} of frame length M which contains samples labelled with $x_{(j-1)M+i}$, $i=1,2,...,M$, calculation of the variance of buffered frame $\dagger_j^2$ is performed [6]

$$\dagger_j^2 = \frac{1}{M}\sum_{i=1}^{M} x_{(j-1)M+i}^2. \qquad (11)$$

In the next step, the quantization by quantizers designed for s^{th} range of these variances and quantizer designed for Gaussian distribution and quantization for the Laplacian distribution is performed. Then the distortions are calculated for both quantized signals, quantizer which provides less distortion is chosen, and at the receiver side are sent samples quantized by the better one.
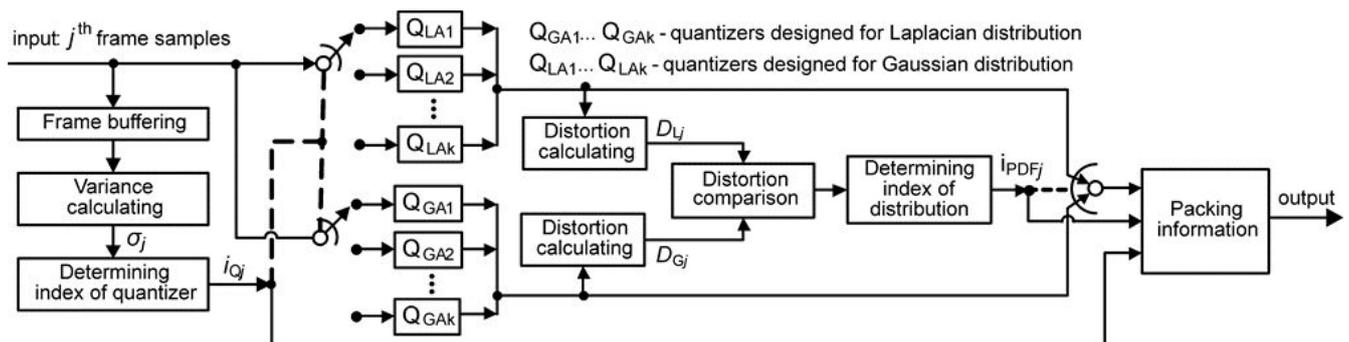
Following the calculated variance out of k ranges signal variance, using the switched technique, the choice of sth range is performed.
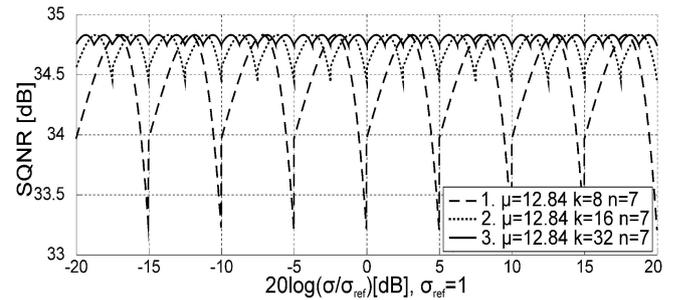

Fig. 2. Quality of the quantization expressed by SQNR, for the model with 8, 16 and 32 quantizers for Laplacian distribution of the source of signal.
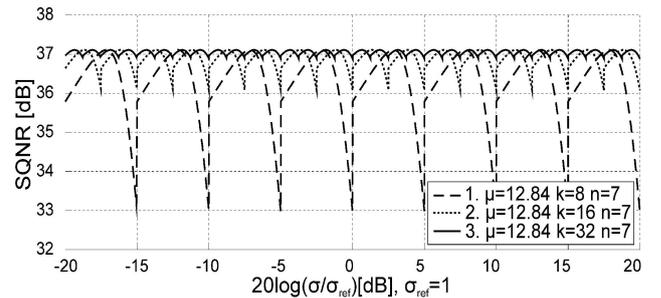

Fig. 3. Quality of the quantization expressed by SQNR, for the model with 8, 16 and 32 quantizer for Gaussian distribution of the source of signal.

To the receiver as additional information are transmitted: 1) index that identifies a class determined by frame variance 2) index of distribution by which it is identified which quantizer is used, one designed for Laplacian or the other designed for Gaussian distribution. Information of index of distribution is length of one bit. (Fig. 4).

If any of k codebook has got a size N, the bit rate expressed in number of bits per sample is

$$R = \log_2 N + \frac{\log_2 k}{M} + \frac{1}{M}. \qquad (12)$$

The transfer of information of one bit, which describes the distribution, increases the bit rate for addition $1/M$ (12) compared to the models when the adaptation is performed only to the variance of the signal. Numerical values of the bit rate depending on the number of codebooks, for different frame lengths, are presented in Table II, for $N = 128$. It may be noted that the transfer of information of the distribution and variance more significantly affects the bit rate when very short frames are used.

Fig. 4. Scheme of the new-proposed model of switched scalar quantization with adaptation to both the variance and the distribution of signal.

TABLE II. BIT RATE FOR 128 QUANTIZATION LEVELS ($N = 128$).

Number of quantizers	Frame length						
	5	10	20	40	80	160	320
2	7.4	7.2	7.1	7.05	7.03	7.01	7.01
4	7.6	7.3	7.15	7.08	7.04	7.02	7.01
8	7.8	7.4	7.2	7.1	7.05	7.03	7.01
16	8.0	7.5	7.25	7.13	7.06	7.03	7.02
32	8.2	7.6	7.3	7.15	7.08	7.04	7.02

Compared to models with adaptation only to the variance, with the suggested model it is necessary to carry out one more quantization, calculation of the distortion for two frames and comparison of distortion, which requires additional processing power. Technological development and reduction of hardware costs have contributed that the need for the increase of processing power is a less expressed factor of constraint in the implementation.

V. Numerical Results of Quantization on the Example of Real Signal

The new suggested model is compared with the usual models of quantization on the example of an available signal of a speaker in the dialogue duration of 120 s. The recording of the signal has been performed by the audio system of PC computer with sampling frequency of 8 kHz, encoding samples with 16 bits. The signal was filtered by a filter range from 300 Hz to 3400 Hz.

Quantizers are designed for the dynamic range of the signal (–20 dB 20 dB). Quantization of the samples with 7 bits was performed for a different number of codebooks and different frame length.

In the Tables III–Tables V the values of the SQNR of available voice signal are presented. With AdLa and AdGa are denoted the results of switching quantization for designed quantizers for Gaussian and Laplacian distribution and adaptation respectively only with the adaptation to the variance of the frame, and with AdLaGa, the results obtained by the proposed model with adaptation to the variance and to the distribution of samples of the signal.

It can be concluded that the values obtained by experiment on the real signal, referring to the quantization error and SQNR, are in accordance with the theoretical consideration (Fig. 2 and Fig. 3).

In the given example, the proposed model, which, besides the adaptation to the variance, uses switching technique for adaptation to the distribution of speech signal, provides higher SQNR for all frame lengths, considering different number of codebooks, in relation to the models designed for the speech source with Laplacian and Gaussian distribution

function when the adaptation is performed only to the variance of the signal (Table III–Table V).

TABLE III. SQNR [DB] FOR THE MODELS WITH 8 QUANTIZERS.

Quantization model	Frame length			
	40	80	160	320
AdLa	34.59	34.59	34.60	34.67
AdGa	36.83	36.47	36.21	35.11
AdLaGa	36.88	36.86	36.86	36.86

TABLE IV. SQNR [DB] FOR THE MODELS WITH 16 QUANTIZERS.

Quantization model	Frame length			
	40	80	160	320
AdLa	34.70	34.72	34.72	34.77
AdGa	37.09	36.99	36.72	35.87
AdLaGa	37.11	37.11	37.11	37.11

TABLE V. SQNR [DB] FOR THE MODELS WITH 32 QUANTIZERS.

Quantization model	Frame length			
	40	80	160	320
AdLa	34.73	34.73	34.75	34.79
AdGa	37.16	37.07	36.82	35.88
AdLaGa	37.17	37.15	37.15	37.14

VI. Conclusions

The new proposed model provides better quality of quantization than conventional models. In the example mentioned above, depending on the length of the frame and the number of quantizers, SQNR is higher up to 2 dB. The application of the proposed model requires additional processing power and additional transfer of information from one bit per frame in comparison to usual models with adaptation only to the variance of signal.

VII. References

[1] N. S. Jayant, P. Noll, *Digital Coding Of Waveforms, Principles and Applications to Speech and Video*. New Jersey, Prentice Hall, 1984.
[2] L. R. Rabiner, R. W. Schafer, "Introduction to digital speech processing", *Foundations and Trends in Signal Processing*, vol. 1, pp. 1–194, 2007. [Online]. Available: http://dx.doi.org/10.1561/2000000001
[3] P. M. Goran, P. H. Zoran, S. V. Leonid, "Switched scalar quantization with adaptation performed on both the power and the distribution of speech signal", *Telfor Journal*, vol. 3, no. 2, pp. 96–99, 2011. [Online]. Available: http://2010.telfor.rs/files/radovi/TELFOR2010_05_25.pdf
[4] A. Mosic, Z. Peric, M. Savic, S. Panic, "Switched semilogarithmic quantization of Gaussian source with LowDelay", *Elektronika ir Elektrotechnika*, no. 2, pp. 71–74, 2011. [Online]. Available: http://dx.doi.org/10.5755/j01.eee.108.2.148
[5] A. Anskaitis, A. Kajackas, "The tool for quality estimation of short voice segments", *Elektronika ir Elektrotechnika*, no. 8, pp. 103–108, 2010.
[6] Z. Peric, J. Nikolic, "An adaptive waveform coding algorithm and its application in speech coding", *Digital Signal Processing*, vol. 22, no. 1, pp. 199–209, 2012. [Online]. Available: http://dx.doi.org/10.1016/j.dsp.2011.09.001