# Switched Scalar Quantization with Adaptation Performed on both the Power and the Distribution of Speech Signal

Goran M. Petković, Zoran H. Perić, and Leonid V. Stoimenov

*Abstract* — **This paper analyzes the models for switching scalar quantization of a source with the Laplacian and Gaussian distribution. We have analyzed the results of real telephone speech and proposed a model of switching scalar quantization, which, in addition to adaptation on the power of speech, includes the adaptation on the distribution of signals (Gaussian and Laplacian ), which resulted in a better quality of voice signal pronounced with Signal-to-Quantization-Noise Ratio.**

***Keywords*** **— Quantization algorithms, speech signal processing, switched scalar quantization**

## I. Introduction

QUANTIZERS have an important role in the theory and practice of modern signal processing [1]-[5]. The quantization procedure transforms the current value of the input signal, generally belonging to an uncountable set of values from the continual amplitude range into the closest possible value from the final discrete amplitude range. Quantization is not just a simple operation preceding the signal coding but rather represents an efficient technique for data compression [1]-[5].

It is widely known that real signals, such as a speech signal, are non-stationary processes expressing their characteristics through the changes of average power in time resulting in a wide dynamic range. Since the change of the average power of many real signals is slow, the process may be considered stationary in most cases with short time intervals. In order to achieve the maximum quality of the received speech signal when coding it, the use of adaptive coders is recommended for a given transfer speed, i.e. the use of quantizers that adapt to local source characteristics during quantization [6].

Apart from the commonly used adaptation to the speech signal power in the proposed model, the adaptation to the speech signal distribution is also applied, which improves the quantization quality. In this paper the Signal-to-Quantization-Noise Ratio (SQNR) is used as an objective measure of quality assessment which is common for high-quality digitized signals. The subjective measurements of the quality are most common for low quality signal coding.

Goran M. Petkovic is with High School of Applied Studies, Vranje, Serbia, (phone: 381-17-421589; e-mail: goranpetko@yahoo.com).

Zoran H. Peric is with Faculty of Electronic Engineering, University of Nis, Serbia, (phone: 381-63-691961, 381-18-529101, e-mail: zoran.peric@elfak.ni.ac.rs ).

Leonid V. Stoimenov is with Faculty of Electronic Engineering, University of Nis, Serbia, (phone: 381-18-642235, e-mail: leni@elfak.ni.ac.rs).

In section II, the theoretical basics of scalar quantization of a source with the Laplacian and Gaussian distribution will be presented. In section III, the new model of adaptive switching scalar speech-signal quantization will be presented. The experimental results will be presented in section IV.

## II. Companding Model

One way of realizing a non-uniform quantization was proposed by Bennet [1]-[3]. He introduced a companding technique by means of which the non-uniform quantization can be achieved by compressing the input signal $x$ by using a compressor with the nonlinear characteristic $c(\cdot)$, followed by the quantization of the compressed signal c(x) based on a uniform quantizer. Finally, the compressed quantized signal value is expanded by means of the nonlinear inverse compression characteristic $c^{-1}(\cdot)$ [1]-[3]. The structure of the described non-uniform quantizer and expander is known as a compander (Fig. 1).

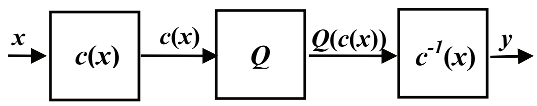


Fig. 1. Block diagram of the companding technique.

In speech coding, the exact value of the input variance is not known in advance. Moreover, it tends to change with time. In such a situation, the constant SNQR in a wide range of an input variance can be obtained by using the logarithmic compression law:

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

whereby the non-dimensional parameter $\mu$, known as compression factor, as well as $x_{max}$ refer to the maximum load amplitude of the quantizer defining the amplitude range of the scalar compander within the range $[-x_{max}, x_{max}]$. If the signal source characterized as a continual random variable with probability distribution $p(x)$, the distortion of non-uniform scalar quantizer is defined as the expected mean square error between the original and quantized signal. The final distortion consists of two components, granular distortion and the overload distortion. Symbolically,

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

where the granular and the overload distortion have been defined as follows:

$$D_g = \sum_{i=2}^{N-1} \int_{t_{i-1}}^{t_i} (x - y_i)^2 p(x) dx \tag{3}$$

$$D_o = 2 \int_{t_{N-1}}^{\infty} (x - y_N)^2 p(x) dx \tag{4}$$

Equation (3) may be presented as follows:

$$D_g = \frac{\ln^2(1+\mu)}{3N^2}\sigma^2\left[\frac{1}{\mu^2}\frac{x_{\max}^2}{\sigma^2} + \frac{2}{\mu\sigma^2}x_{\max}\overline{|x|} + \frac{\overline{x^2}}{\sigma^2}\right] \tag{5}$$

where $N$ represents the number of quantization levels, and

$\overline{|x|}$ and $\overline{x^2}$ have been defined as:

$$\overline{|x|} = 2\int_0^{x_{\max}} x p(x) dx \tag{6}$$

$$\overline{x^2} = 2\int_0^{x_{\max}} x^2 p(x) dx \tag{7}$$

### *A. The distortion for the Laplacian source*

The speech signal is presented by the function of the Laplacian distribution, whose density for the independent variable $x$ with the unit variance may be presented as follows:

$$p(x,\sigma) = \frac{\sqrt{2}}{2\sigma} e^{-\frac{|x|\sqrt{2}}{\sigma}} \tag{8}$$

By changing (8), (7), (6) into (5) and after solving these two integrals and the specific approximation of the distortion, the following can be obtained [7]:

$$D_t = \frac{\ln^2(1+\mu)}{3N^2}\sigma^2\left[\frac{1}{\mu^2}\frac{x_{\max}^2}{\sigma^2} + \frac{x_{\max}}{\sigma}\frac{\sqrt{2}}{\mu} + 1\right] + \sigma^2 e^{-\frac{\sqrt{2}x_{\max}}{\sigma}} \tag{9}$$

The SQNR may be calculated for the quantization of the Laplacian source for the power in the wide dynamic range.

$$SQNR = 10\lg\frac{\sigma^2}{D_t} \tag{10}$$

In order to simplify the quantization performance analysis, the hypothesis of the mean value of the input signal being zero is commonly applied, so that the power of the input signal in the final equation is substituted by its variance. It can be observed that by applying this assumption the general aspect has not been lost.

### *B. The distortion for the Gaussian source*

Similarly to the previously conducted analysis, the speech signal may be analysed as based on the function of the Gaussian distribution:

$$p(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{x^2}{2\sigma^2}} \tag{11}$$

By substituting (11), (7), and (6) with (5), based on approximation and after solving the integral, the following equation for the distortion and the SNQR may be obtained:

$$D_t = \sigma^2\left(\frac{\ln^2(1+\mu)}{3N^2}\left[\frac{1}{\mu^2}c^2 + c\frac{2}{\mu}\sqrt{\frac{2}{\pi}} + 1\right] + \left[\left(1+c^2\right)\left(1 - erf\frac{c}{\sqrt{2}}\right) - ce^{-\frac{c^2}{2}}\sqrt{\frac{2}{\pi}}\right]\right) \tag{12}$$

$$SQNR = 10\lg\frac{\sigma^2}{D_t} \tag{13}$$

## III. The Model of Switched Quantization Based on Average Power Adaptation and Signal Distribution

One of the ways of achieving a constant SNQR is the switched adaptive scalar quantization with a codebook. Based on this, the input frame determines whether the block of the samples belonging to a particular statistic class in range $k$ in comparison to all possible classes. An index identifying the class or the codebook is transferred as additional information.

The models of the class of switching scalar quantization differ among themselves in the way of adaptation. The proposed model shows the adaptation based on both the power and the distribution of speech signal. Errors in the transmission channel do not significantly affect the SQNR and in this paper they have not been considered [1],[4].

### *A. Adaptation to average power*

The quantization technique in the proposed model is based on a frame-by-frame principle of input sample processing. The processing procedure for the signal begins with buffering. After the buffering of the $j$th frame of $M$ length, containing samples marked as $x_{j+i}$, $i$=0,1,...,$M$-1, a variance approximation of the buffered frame $\sigma_j^2$ is completed followed by its non-uniform quantization $\hat{\sigma}_j^2$ [3],[4]. Based on the switching technique, the selection of one $s$th quantizer from among the available k quantizers projected for the variances $\hat{\sigma}_p^2$, $p$=1…$k$ distributed in a log-uniform way in the dynamic range of the variances $B = 20\log(\sigma_{\max}/\sigma_{\min})$:

$$20\log\hat{\sigma}_p = 20\log\sigma_{\min} + (2p-1)B/(2k),\ p = 1,\dots k$$

is enabled.

When projecting each of the available $k$ quantizers in each of the $k$ ranges that the entire dynamic range of the variances $B$ are distributed upon, the optimal value of the maximum load amplitude is determined in such a way that the total distortion for $\sigma = \hat{\sigma}_p$ is minimal. The optimization of the total distortion has been performed for the fixed standard value of the parameter $\mu$=255 by optimizing the parameter $c$:

$$\frac{\partial D_t}{\partial c} = 0 \Rightarrow c = c_{opt} \tag{14}$$

representing the $x_{\max}/\hat{\sigma}_p$ ratio and called the factor of relative quantizer range.

The size of the codebook $N$, which equals the number of quantization levels, depends on the number of bits used for the coding $n$. The relationship between $N$ and $n$ is $N=2^n$, where $n$ represents the bit number per sample.

If there are no limits regarding either memory size or the processing power, there is a possibility of choice of the optimal number of quantizers in a system model. By that, a high quality of the quantized signal measured by SQNR is provided in the wide input range of signal (Fig. 2 and Fig. 3).

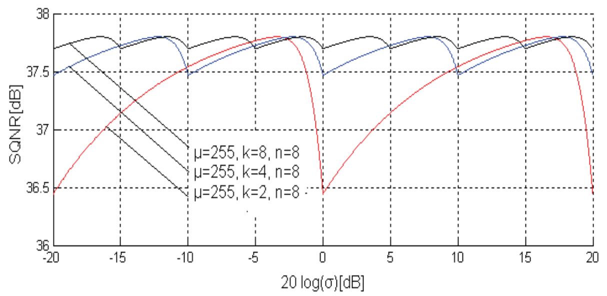


Fig. 2. SQNR for a model with two, four or eight quantizers for the Laplacian source.

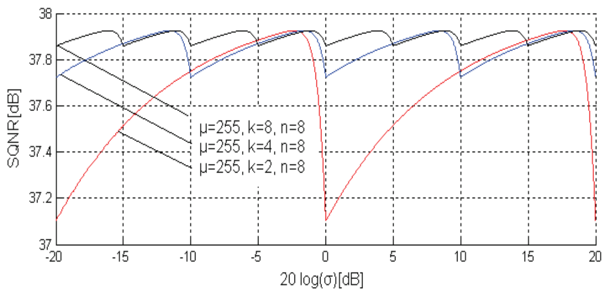


Fig. 3. SQNR for a model with two, four or eight quantizers for the Gaussian source.

### B. Adaptation to signal distribution

The Laplacian distribution represents a solid approximation of the real probability density function for current values of a high-quality recorded speech signal. For a speech signal recorded with a low-quality microphone, the deviations from this distribution are considerably more prominent [1].

When observing the time domain of the speech signal, it can be noticed that the voiced phonemes (vocals and semi-vocals) are of a high current value. A relatively low current value may be observed with voiceless phonemes (fricatives, affricates, etc.) Since the wave forms of voiceless phonemes are stochastic (turbulence, air friction in the vocal apparatus of the speakers), similar to noise in the case of a recording and speech transfer system, the Gaussian distribution may turn out to be the dominant one. For these reasons sometimes the Gaussian and sometimes Laplacian distribution may describe the speech signal in a better way [1].

In the proposed model, after choosing the quantizer by a criterion of the average frame power, quantization is being performed. Quantization is being performed both by the quantizer projected for the Gaussian distribution, and by the quantizer for the Laplacian distribution, and the distortion for such quantized signals is being calculated. Afterwards, the quantizer, which provides better quality is being chosen.

Additional information is sent to the receiver in the form of an index identifying the class, i.e. the codebook based on the average frame power and the one-bit information determining whether the quantizer used is projected for the Laplacian or the Gaussian distribution. These additional pieces of information are usually either at the beginning or at the end of the block.

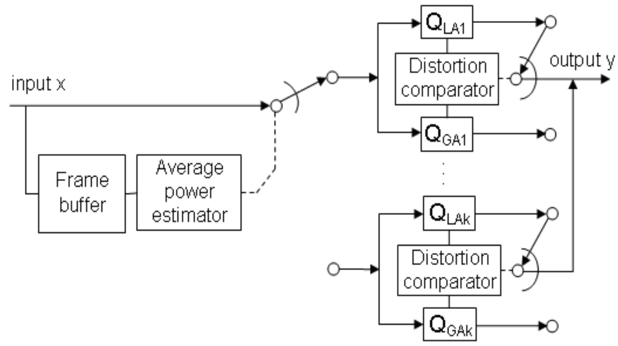


Fig. 4. Switched adaptive scalar quantization with adaptation to average power and speech signal distribution

If each one of $k$ codebooks is of size $N$ and $M$ is the length of the frame, the number of bits per sample is:

$$R = \log_2 N + \frac{\log_2 k}{M} + \frac{1}{M} \tag{15}$$

The information about the codebook increases the bit rate for addend $\log_2(k)/M$, and the transfer of the one-bit information about the selected distribution (Gaussian or Laplacian) for addend $1/M$. The transfer of these pieces of information has a significant influence on the bit rate for very short frames (Fig. 5).

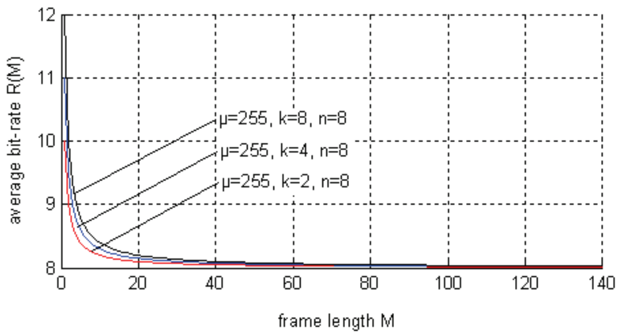


Fig. 5. Average bit-rate for the quantizer with adaptation to average power and speech signal distribution

## IV. EXPERIMENTAL RESULTS

Three available voice signals of three different speakers in the dialogue, marked as SP01, SP02 and SP03, have been analysed. Their frequency was 8 kHz and the samples were coded with 16 bits. The signals were filtered by a 300Hz-3400Hz-range filter. The quantizers were projected for a particular dynamic signal range (-20dB 20dB). The quantization of 8-bit samples was analysed. The analysis was performed for different frame lengths.

Tables 1-3 display the values for the SNQR for a different number of codebooks for different frame lengths. La and Ga are used to mark the results of the switched quantization for projected quantizers based on the Gaussian and Laplacian distribution and adaptation only to an average frame power, whereas LaGa is used to mark the results obtained by means of the proposed model with adaptation to both average power and distribution.

TABLE 1: SQNR FOR THE MODEL WITH TWO QUANTIZERS.

| *k=2* | | ***Frame length*** | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | 5 | 10 | 20 | 40 | 80 | 160 | 320 |
| SP 01 | La | 37.15 | 37.19 | 37.25 | 37.32 | 37.31 | 37.32 | 37.33 |
| | Ga | 37.57 | 37.60 | 37.64 | 37.67 | 37.67 | 37.66 | 37.67 |
| | LaGa | 38.79 | 38.53 | 38.34 | 38.20 | 38.01 | 37.89 | 37.82 |
| SP 02 | La | 37.12 | 37.11 | 37.10 | 37.08 | 37.08 | 37.08 | 37.09 |
| | Ga | 37.57 | 37.56 | 37.56 | 37.55 | 37.55 | 37.54 | 37.40 |
| | LaGa | 38.71 | 38.33 | 38.07 | 37.88 | 37.74 | 37.65 | 37.61 |
| SP 03 | La | 37.09 | 37.12 | 37.13 | 37.14 | 37.15 | 37.17 | 37.19 |
| | Ga | 37.49 | 37.51 | 37.52 | 37.53 | 37.53 | 37.54 | 37.53 |
| | LaGa | 38.71 | 38.36 | 38.11 | 37.93 | 37.78 | 37.69 | 37.64 |

TABLE 2: SQNR FOR THE MODEL WITH FOUR QUANTIZERS.

| *k=4* | | ***Frame length*** | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | 5 | 10 | 20 | 40 | 80 | 160 | 320 |
| SP 01 | La | 37.56 | 37.56 | 37.57 | 37.60 | 37.60 | 37.60 | 37.61 |
| | Ga | 37.77 | 37.78 | 37.79 | 37.80 | 37.80 | 37.71 | 37.71 |
| | LaGa | 39.26 | 38.93 | 38.65 | 38.47 | 38.25 | 38.09 | 38.00 |
| SP 02 | La | 37.57 | 37.60 | 37.60 | 37.62 | 37.61 | 37.62 | 37.63 |
| | Ga | 37.80 | 37.82 | 37.81 | 37.82 | 37.82 | 37.82 | 37.69 |
| | LaGa | 39.32 | 38.97 | 38.64 | 38.42 | 38.21 | 38.08 | 37.99 |
| SP 03 | La | 37.55 | 37.57 | 37.58 | 37.59 | 37.59 | 37.61 | 37.61 |
| | Ga | 37.78 | 37.79 | 37.79 | 37.79 | 37.79 | 37.80 | 37.77 |
| | LaGa | 39.25 | 38.88 | 38.57 | 38.35 | 38.16 | 38.04 | 37.95 |

TABLE 3: SQNR FOR THE MODEL WITH EIGHT QUANTIZERS.

| *k=8* | | ***Frame length*** | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | 5 | 10 | 20 | 40 | 80 | 160 | 320 |
| SP 01 | La | 37.66 | 37.70 | 37.70 | 37.70 | 37.70 | 37.70 | 37.71 |
| | Ga | 37.84 | 37.85 | 37.86 | 37.85 | 37.86 | 37.75 | 37.43 |
| | LaGa | 38.66 | 38.58 | 38.47 | 38.37 | 38.21 | 38.11 | 38.04 |
| SP 02 | La | 37.66 | 37.69 | 37.70 | 37.71 | 37.71 | 37.72 | 37.72 |
| | Ga | 37.85 | 37.86 | 37.86 | 37.87 | 37.87 | 37.86 | 37.59 |
| | LaGa | 38.71 | 38.57 | 38.41 | 38.30 | 38.17 | 38.07 | 38.02 |
| SP 03 | La | 37.66 | 37.67 | 37.69 | 37.69 | 37.69 | 37.70 | 37.70 |
| | Ga | 37.82 | 37.83 | 37.83 | 37.84 | 37.84 | 37.83 | 37.81 |
| | LaGa | 38.64 | 38.50 | 38.35 | 38.22 | 38.10 | 38.03 | 37.97 |

It may be concluded that the values obtained in the experiment performed on real signals as related to the quantization error, i.e. the SNQR, lie within the range predicted by means of theoretical consideration (Fig. 2 and Fig. 3).

The comparison of the model with adaptation only to the average frame power leads to the conclusion that better results are obtained by the model projected for the Gaussian distribution in comparison to the model projected for the Laplacian distribution whereas the differences are more expressed with shorter frames. This may be explained by the fact that the recorded speech is better described by the Gaussian distribution, since no high-quality technical components were used during the recording process. More prominent differences in quality may be observed between models with two quantizers. The proposed model with adaptation to average power and speech signal distribution provides better results for all frame lengths whereby the improvement is specifically evident with shorter frames when the difference in comparison to previous models is larger than 1 dB. According to the analysis of the numerical results and the implementation complexity, it can be concluded that the best solution is to choose a model with four quantizers.

With the proposed model, the demands related to the processor power have been raised. This is because the signal needs to be quantized, the distortion determined for both the case with the Gaussian and the Laplacian distribution, and one of two quantizers chosen. The development of contemporary hardware components has resulted in a less prominent factor of limitation in the process of realization.

## V. CONCLUSION

The proposed model of switched quantization results in a better quality of the transfer of a speech signal in comparison to the model of switched scalar quantization where adaptation is obtained only to the signal power. This improvement may be obtained by slightly increasing the transfer information and by the increased demands related to the processor power and memory requirements.

## REFERENCES

[1] G. Lukatela, D. Drajic, G. Petrovic, R. Petrovic, *Digital Telecommunications*, Gradjevinska knjiga, Beograd 1981.

[2] W. C. Chu, *Speech Coding Algorithms, Foundation and Evolution of Standardized Coders*, New Jersey: John Wiley & Sons, 2003.

[3] A. Gersho and R. M. Gray, *Vector Quantization and Signal Compression*, Boston, Dordrecht, London: Kluwer Academic Publishers, 1992.

[4] N. S. Jayant and P. Noll, *Digital Coding Of Waveforms, Principles and Applications to Speech and Video*, New Jersey: Prentice Hall, 1984.

[5] L. R. Rabiner and R. W. Schafer, "Introduction to Digital Speech Processing", *Foundations and Trends in Signal Processing*, vol. 1, pp. 1-194, Jan. 2007.

[6] Z. Peric, G. Petkovic, J. Nikolic, " Switched Scalar Quantization of Speech Signal in a Wide Dynamic Range of Variances ", *ETRAN 2010*, Jun 2010.

[7] J.Nikolic and 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.