

Lloyd–Max’s Algorithm Implementation in Speech Coding Algorithm Based on Forward Adaptive Technique

Jelena NIKOLIC, Zoran PERIC

*Faculty of Electronic Engineering
Aleksandra Medvedeva 14, 18000 Nis, Serbia
e-mail: njelena@elfak.ni.ac.yu, peric@elfak.ni.ac.yu*

Received: July 2007

Abstract. In this paper a detail analysis of speech coding algorithm based on forward adaptive technique is carried out. We consider an algorithm that works on frame-by-frame basis, where a frame consists of a certain number of speech samples. Buffering frame-by-frame an estimation of the gain defined as squared root of the frame variance is enabled. The information about the gain (side information) and the code book of a nonadaptive quantizer, which is designed for the unit variance case of the input signal, are further used when designing an adaptive quantizer. In such a way better quantizer adaptation to the varying input statistics is provided. Observe that the goal of this paper is to investigate the preference that for the wide range of variance change could be achieved when implementing in the forward adaptive speech coding algorithm, the recently developed effective method for the Lloyd–Max’s algorithm initialization, which provides optimal Lloyd–Max’s quantizer performances for the unit variance case of the input signal. We destine to consider the speech coding algorithm based on forward adaptive technique since the backward adaptation provides $SQNR$ (signal to quantization noise ratio) within 1 dB of the forward adaptation. We provide theoretical and experimental results (performances of our algorithm) which are compared with the optimal results. Additionally, we discuss the performances of speech coding schemes designed according to G. 711 standard and we point out the benefits that can be achieved by using our algorithm. Finally, in order to find better solution for implementation of the proposed algorithm in practice we consider the performances of our algorithm when log-uniform as well as uniform scalar quantizer are used for gain quantizing.

Keywords: speech coding algorithm, forward adaptive technique, Lloyd–Max’s scalar quantizer.

1. Introduction

Speech coding is a procedure to represent a digitized speech signal using as few bits as possible, maintaining at the same time a reasonable level of speech quality (usually measured by $SQNR$) (Chu, 2003; Kondo, 2004; Gersho, 1994; Rabiner and Schafer, 1978; Pan, 1993; Pohlman, 1989; Jayant and Noll, 1984; Gersho and Gray, 1992; Minoli, 2002; Hersent *et al.*, 2005). A not so popular name having the same meaning is speech compression. Speech coding has matured to the point where it now constitutes an important application area of signal processing. Due to the increasing demand for

speech communication, speech coding technology has received augmenting levels of interest from the research, standardization, and business communities (Chu, 2003; Kondoz, 2004; Gersho, 1994; Rabiner and Schafer, 1978; Pan, 1993; Pohlman, 1989; Jayant and Noll, 1984; Gersho and Gray, 1992; Minoli, 2002; Hersent *et al.*, 2005; Hiwasaki *et al.*, 2006; Khasnabish, 2003). With the widespread availability of low-cost high performance digital signal processors many signal processing task done in the old days using analog circuitry are predominantly executed in the digital domain. Advantages of going digital are many: programmability, reliability, and the ability to handle very complex procedures, such as the operations involved in speech coders. Regarding to these advances and the vast availability of low-cost programmable processors and dedicated chips, the rapid transfer from research to the product development is enabled. This encourages the researchers to investigate the alternative schemes for speech coding with the objectives of overcoming deficiencies and limitations of the existing speech coding schemes. The results of such investigations are usually development of novel speech coding standards or improvement of existing standards. Until now, so much work has been done to reduce the bit rate while preserving the original quality of the digitized signal. For instance widely used G. 711 standard (Chu, 2003; Gersho and Gray, 1992; Minoli, 2002; Hersent *et al.*, 2005; Khasnabish, 2003; ITU-T, Recommendation G. 711, 1972) has been developed in order to processes a digital, linear, quantized signal on 12 bits. With this standard each 12 bits sample are converted to 8-bit code. The gain of using G. 711 is not in quality but in the resulting bit rate.

The importance and reality of the alternative speech coding scheme investigation, which satisfy G. 712 standard (ITU-T, Recommendation G. 712, 1992), i.e., which provides better quality of the speech signal when compared to the quality achieved by the G. 711 standard while attaining as higher as possible compression, occupied us and induce to think about possible solutions. Observe that selection of the particular speech coding scheme or development of novel scheme that fits in the available bandwidth and provides desired level of service (defined by G. 712 standard (ITU-T, Recommendation G. 712, 1992)) is not an obvious task. Namely, in order to make right choice it is necessary to have a deeper understanding on the internals of each candidate speech coders.

Let us define that a stationary process is a stochastic process whose probability distribution at a fixed time or position is the same for all times or positions. As a result, parameters such as the mean and variance also do not change over time or position. As an example, white noise is stationary. However, speech signals are not stationary because the statistic of speech signals constantly change with time (Chu, 2003; Kondoz, 2004; Jayant and Noll, 1984; Gersho and Gray, 1992; Minoli, 2002; Hersent *et al.*, 2005). By statistics, we mean the statistic mean, variance (or the dynamic range), and type of input pdf (probability density function). It is already shown that in order to process speech effectively it is necessary to use some kind of adaptation or system that work on frame-by-frame basis (Chu, 2003; Jayant and Noll, 1984; Gersho and Gray, 1992; Minoli, 2002; Hersent *et al.*, 2005), where a frame consists of a certain number of samples. The actual duration of the frame is known as length. Typically, length is selected between 10 and 30 ms or 80 and 240 samples (Chu, 2003; Jayant and Noll, 1984; Gersho and Gray, 1992; Minoli,

2002; Hersent *et al.*, 2005). Since in this short interval, properties of the speech signal remain roughly constant, it can be viewed as a local stationary signal.

The aforementioned adaptivity of speech coding system can be achieved by using an adaptive coder that attempts to make the encoder-decoder (quantizer) design adapt to the varying input statistics. Namely, there are two different types of adaptive coders: forward adaptive and backward adaptive coders (Chu, 2003; Kondo, 2004; Jayant and Noll, 1984; Gersho and Gray, 1992; Hersent *et al.*, 2005; Ortega and Vetterli, 1997). They conceptually differ regarding the features whether the adaptation is performed forward, i.e., from the input sequence or backward, i.e., from the coded output signal. Also, forward adaptive technique requires transmission of side information while it is not required by the backward adaptive technique since the adaptation performs according to what was transmitted the previous. It is obvious that the forward adaptation is less sensitive to transmission errors when compared to backward adaptation. Moreover, it is demonstrated that the backward adaptation provides $SQNR$ within 1 dB of the forward adaptation (Jayant and Noll, 1984). Hence, in this paper we focus our analysis on speech coding scheme based on forward adaptive technique. Such a coding scheme can be defined as an algorithm since speech coding is performed using numerous steps or operations. Furthermore, it is well known that the adaptivity of such a coding scheme can be achieved by adapting the input of the fixed or nonadaptive quantizer (encoder-decoder) according to the estimated and quantized gain of the frame. The same effect can be achieved by an adaptive quantizer with the code book obtained by multiplying with the estimated and quantized gain the code book of the aforementioned nonadaptive quantizer. Observe that nonadaptive quantizer is usually designed for the unit variance case of the input signal. Recall that the input speech signal can be modeled by the Laplacian distribution (Jayant and Noll, 1984; Gersho and Gray, 1992). Again recall that we already developed an effective method for initialization of Lloyd–Max’s algorithm of optimal scalar quantization for Laplacian source and unit variance case of the input signal (Peric and Nikolic, 2007). Regarding the previous results (Peric and Nikolic, 2007), we wonder here what speech coding performance would be achieved by implementing the Lloyd–Max’s quantizer, designed according to the method proposed in (Peric and Nikolic, 2007), in forward adaptive speech coding scheme. We destined to consider the Lloyd–Max’s model of quantizer since we already demonstrated in (Peric and Nikolic, 2007) that by using the proposed method for Lloyd–Max’s algorithm initialization the algorithm converges after only one iteration when the bit rates are 7 or greater and after two iterations when the bit rates are 4, 5, 6 while providing near optimal performances (measured by $SQNR$). Therefore, we hope that by implementing Lloyd–Max’s algorithm in forward adaptive speech coding algorithm near optimal performance would be achieved for the wide dynamic range of the input signal. The analysis that follows provides the answer on such a speculation.

Lloyd–Max’s Algorithm Implementation in Forward Adaptive Speech Coding Algorithm

As we already mentioned the procedure of adaptation can be performed by normalizing the input sequence, further coding with nonadaptive quantizer (encoder – decoder) and

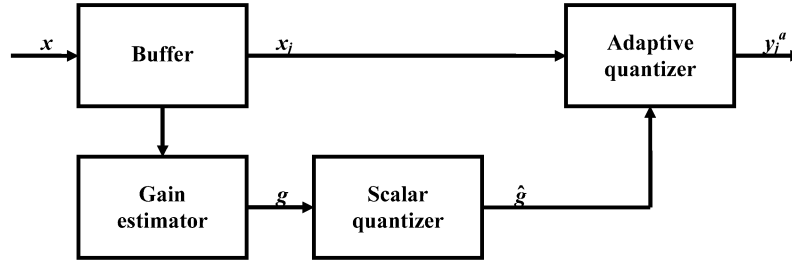


Fig. 1. General speech coding scheme based on forward adaptive technique.

finally denormalizing by the same quantized value of the calculated gain that is used for normalizing. Also, we point out that the same effect can be achieved by an adaptive quantizer (see Fig. 1) which code book is obtained by multiplying with the estimated and quantized gain the code book of the aforementioned nonadaptive quantizer (Gersho and Gray, 1992).

Since we already provide an effective method for optimal design of fixed Lloyd–Max’s quantizers (Peric and Nikolic, 2007), here we destined to implement such a designed quantizer in the general speech coding scheme that is depicted in Fig. 1. Namely, we appropriate to consider speech coding scheme consisting of a buffer, an adaptive N -level Lloyd–Max’s scalar quantizer, the gain estimator and the N_g -level scalar quantizer (for gain quantizing). Observe that every speech coding scheme can be defined as an algorithm since speech coding is performed using numerous steps or operations. Therefore, we provide here in detail a novel speech coding algorithm based on forward adaptive technique in which adaptive Lloyd–Max’s quantizer is implemented and designed according to the method proposed in (Peric and Nikolic, 2007).

The novel algorithm consists of following steps:

Step 1. Design of nonadaptive (fixed) Lloyd–Max’s scalar quantizer, by using initialization method that was proposed in (Peric and Nikolic, 2007), for the unit variance case ($\sigma_{ref}^2 = 1$) of the input signal x which is model by the Laplacian distribution (Gersho and Gray, 1992):

$$p(x, \sigma) = \frac{\sqrt{2}}{2\sigma} e^{-\frac{|x|\sqrt{2}}{\sigma}}. \quad (1)$$

Design of Lloyd–Max’s scalar quantizer is equal to finding its decision thresholds and the representation levels. The initial values for the decision thresholds $\{t_1^{(0)}, t_2^{(0)}, \dots, t_{N-1}^{(0)}\}$ and the representation levels $\{y_1^{(0)}, y_2^{(0)}, \dots, y_N^{(0)}\}$ of the Lloyd–Max’s quantizer are defined by the compandor model which in the first quadrant (the symmetry of decision thresholds and the representation levels are assumed) can be given by (Peric and Nikolic, 2007):

$$t_j^{(0)} = \frac{3}{\sqrt{2}} \ln \left(\frac{N}{2N - 2j + (2j - N) \exp \left(-\frac{\sqrt{2}}{3} x_{\max} \right)} \right) t_j^{(0)} > 0, \quad \frac{N}{2} < j < N, \quad (2)$$

$$y_j^{(0)} = \frac{3}{\sqrt{2}} \ln \left(\frac{N}{2N - 2j + 1 + (2j - 1 - N) \exp \left(-\frac{\sqrt{2}}{3} x_{\max} \right)} \right) y_j^{(0)} > 0, \quad (3)$$

$$\frac{N}{2} < j \leq N,$$

where x_{\max} is maximum input signal amplitude. Dependence of maximum input signal amplitude on the number of Lloyd–Max’s quantizer levels N is determined in (Peric and Nikolic, 2007) in the following form:

$$x_{\max} = \frac{3}{\sqrt{2}} \ln \left(\frac{N}{2} \right) + \sqrt{2}. \quad (4)$$

Using the proposed initialization, the new values for the decision thresholds and the representation levels of Lloyd–Max’s quantizer can be determined from the following relations (Lloyd–Max’s algorithm) (Jayant and Noll, 1984; Gersho and Gray, 1992; Peric and Nikolic, 2007):

$$t_j^{(i+1)} = \frac{(y_j^{(i)} + y_{j+1}^{(i)})}{2}, \quad 1 \leq j < N, \quad (5)$$

$$y_j^{(i+1)} = \frac{\int_{t_{j-1}^{(i+1)}}^{t_j^{(i+1)}} xp(x) dx}{\int_{t_{j-1}^{(i+1)}}^{t_j^{(i+1)}} p(x) dx}, \quad 1 \leq j \leq N. \quad (6)$$

Let us reword that stopping criteria that interrupts Lloyd–Max’s algorithm, which was settled in (Peric and Nikolic, 2007), can be satisfied after one iteration when the bit rates are 7 or greater and after two iterations when the bit rates are 4, 5, 6. Here we adopt such a criteria and perform design of nonadaptive Lloyd–Max’s scalar quantizers when the number of quantization levels N varies ($rate = \log_2 N$).

Step 2. Buffering of the input speech signal and gain estimation. A finite number M of input samples (frame) are buffered and used for gain estimation (squared root of the frame variance):

$$g = \sqrt{\frac{1}{M} \sum_{i=1}^M x_{j+i}^2}, \quad (7)$$

where $M \geq 1$ is a finite number, known as a frame length (Chu, 2003; Jayant and Noll, 1984; Gersho and Gray, 1992; Minoli, 2002; Hersent *et al.*, 2005).

Step 3. Quantization of the estimated gain. The estimated gain g is quantized by the scalar quantizer \hat{g} :

a) uniformly

$$\hat{g}_i = \sigma_{\min} + (2i - 1) \frac{\Delta}{2}, \quad i = 1, \dots, N_g; \quad \Delta = \frac{\sigma_{\max} - \sigma_{\min}}{N_g}, \quad (8)$$

where interval $(\sigma_{\min}, \sigma_{\max})$ defines dynamic range of the input signal;

b) log-uniformly

$$\begin{aligned} 20 \log(\hat{g}_i^{lu}) &= 20 \log(\sigma_{\min}) + (2i - 1) \frac{\Delta^{lu}}{2}, \quad i = 1, \dots, N_g; \\ \Delta^{lu} &= \frac{20 \log\left(\frac{\sigma_{\max}}{\sigma_{\min}}\right)}{N_g}, \end{aligned} \quad (9)$$

where interval $(20 \log \sigma_{\min}, 20 \log \sigma_{\max})$ defines dynamic range of the input signal in decibels.

Step 4. Design of adaptive Lloyd–Max’s scalar quantizer. The quantized gain \hat{g} is used for multiplying the code book of nonadaptive Lloyd–Max’s scalar quantizer that is designed in Step 1. The decision thresholds and the representation levels of the adaptive Lloyd–Max’s quantizer denoted with t_j^a and y_j^a respectively can be found from:

$$t_j^a = t_j^a(\hat{g}, \sigma_{ref}) = \hat{g} t_j(\sigma_{ref}), \quad j = 0, 1, \dots, N, \quad (10)$$

$$y_j^a = y_j^a(\hat{g}, \sigma_{ref}) = \hat{g} y_j(\sigma_{ref}), \quad j = 1, 2, \dots, N, \quad (11)$$

where $t_j = t_j(\sigma_{ref})$ and $y_j = y_j(\sigma_{ref})$ denote the decision thresholds and the representation levels of the nonadaptive Lloyd–Max’s quantizer.

Speech Coding Performance

The quality of a quantizer can be measured by the distortion of the resulting reproduction in comparison to the original. The most convenient and widely used measure of distortion between the input signal x and the quantized signal y_j^a is the average mean-squared error, i.e., quantization noise. Observe that time averaged distortion D_a^t , given by (Gersho and Gray, 1992):

$$D_a^t = \lim_{L \rightarrow \infty} \frac{1}{L} \sum_{j=1}^L (x_j - y_j^a)^2 \quad (12)$$

corresponds to the statistically averaged distortion D_a^s (Gersho and Gray, 1992) at one time instant:

$$D_a^s(\sigma) = \sum_{j=1}^N \int_{t_{j-1}^a}^{t_j^a} (x - y_j^a)^2 p(x) dx, \quad (13)$$

if the input signal is stationary ergodic process, where L denotes the number of considered samples x_j . In this paper we assume such equality $D_a = D_a^t = D_a^s$ despite the fact that we consider the local-stationary speech signal.

Observe now that similar to Eq. 13, distortion denoted with $D(\sigma_i)$, for the particular variance of the speech signal σ_i , can be defined with (Gersho and Gray, 1992):

$$D(\sigma_i) = \sum_{j=1}^N \int_{t_{j-1}^a}^{t_j^a} (x - y_j^a)^2 p(x, \sigma_i) dx. \quad (14)$$

Such a defined distortion determines the appropriate signal to quantization noise ratio $SQNR(\sigma_i)$ (Gersho and Gray, 1992):

$$SQNR(\sigma_i) = 10 \log \left(\frac{\sigma_i^2}{D(\sigma_i)} \right). \quad (15)$$

Defining the average signal to quantization noise ratio $SQNR_a$ as follows:

$$SQNR_a = \frac{1}{k} \sum_{i=1}^k 10 \log \left(\frac{\sigma_i^2}{D(\sigma_i)} \right), \quad (16)$$

where k defines the number of the particular variances that are considered, theoretical results can be obtained and hence numerical comparison with the theoretical optimal values of signal to quantization noise ratio $SQNR^{opt}$ (Jayant and Noll, 1984) is enabled.

Since in this paper we perform both theoretical and experimental analysis we provide here the expressions for average signal to quantization noise ratio for both cases. In order to provide the experimental values of the average signal to quantization noise ratio within the each of F frames we define the following relation:

$$SQNR_p^{ex} = 10 \log \frac{(\sigma_p^{ex})^2}{D_p^{ex}}, \quad p = 1, \dots, F, \quad (17)$$

where $(\sigma_p^{ex})^2$ denotes the variance of the input speech samples within the p th frame, $p = 1, \dots, F$:

$$(\sigma_p^{ex})^2 = \frac{1}{M} \sum_{q=1}^M x_{pq}^2, \quad p = 1, \dots, F, \quad (18)$$

and D_p^{ex} denotes the average distortion for the p th frame, $p = 1, \dots, F$,

$$D_p^{ex} = \frac{1}{M} \sum_{q=1}^M (x_{pq} - y_{pq}^a)^2, \quad p = 1, \dots, F. \quad (19)$$

Finally, x_{pq} and y_{pq}^a denote the input speech samples and the outputs of the adaptive Lloyd–Max’s quantizer respectively. Recall that M is the number of speech samples within the frame (frame length) and F is the number of frames that are considered in experiment. By averaging the signal to quantization noise ratios within the each of F

frames (Eq. 17), we can obtain experimental results (average values of $SQNR$, denoted with $SQNR_a^{ex}$):

$$SQNR_a^{ex} = \frac{1}{F} \sum_{p=1}^F SQNR_p^{ex}. \quad (20)$$

Relying on the proposed speech coding algorithm and using the afore defined expressions for the signal to quantization noise ratio determining we provide theoretical and experimental results. Such a results are further compared with theoretically optimal results (Jayant and Noll, 1984) in order to find the particular speech coding scheme that for a fixed number of quantization levels N provides as high as possible average $SQNR$ ($SQNR_a$ or $SQNR_a^{ex}$) while satisfying the criteria:

$$\Delta SQNR_a \geq 0.03 \text{ dB}, \quad (21)$$

The introduced criteria states that there is no point to further increase the number of bits for side information if the difference $\Delta SQNR_a$:

$$\begin{aligned} \Delta SQNR_a = & SQNR_a(N = 2^{q_N}, N_g = 2^{q_{N_g}}) \\ & - SQNR_a(N = 2^{q_N}, N_g/2 = 2^{q_{N_g}-1}), \end{aligned} \quad (22)$$

considered for $q_N = 5, 6, 7$, $q_{N_g} = 1, 2, 3, 4, 5, 6$ is less than 0.03. Finally, observe that we obtain the value of 0.03 considering the well known fact that the increase of one bit per frame consisting of $M = 200$ samples causes expected increase of $SQNR$ of about $6 \text{ dB}/200 = 0.03 \text{ dB}$ (Jayant and Noll, 1984; Gersho and Gray, 1992).

Results

Dynamic range is a term used frequently in numerous fields to describe the ratio between the smallest and largest possible values of a changeable quantity. Changeable quantity that we consider is the signal variance. It is well known that the speech signal modeled by the Laplacian distribution have the wide dynamic range (Jayant and Noll, 1984). Since, for the wide dynamic range it is usually assumed range of 40 dB, in order to obtain numerical results we assume such a dynamic range (Jayant and Noll, 1984). We dispose of 10200 speech samples (8 kHz speech) and perform the buffering with the frame length $M = 200$. It is important to point out that the selection of frame length or the learning period M (Jayant and Noll, 1984) can be one of critical issues. If the learning period is short, the adaptation to the local statistics will be effective, but the side information needs to be sent frequently (more bits needs to be used for sending the side information). If the learning period is long, the number of bits used for side information decrease, the adaptation becomes less sensitive to changing statistics and both processing delay and storage required increase. In practice, a proper compromise between quantity of side

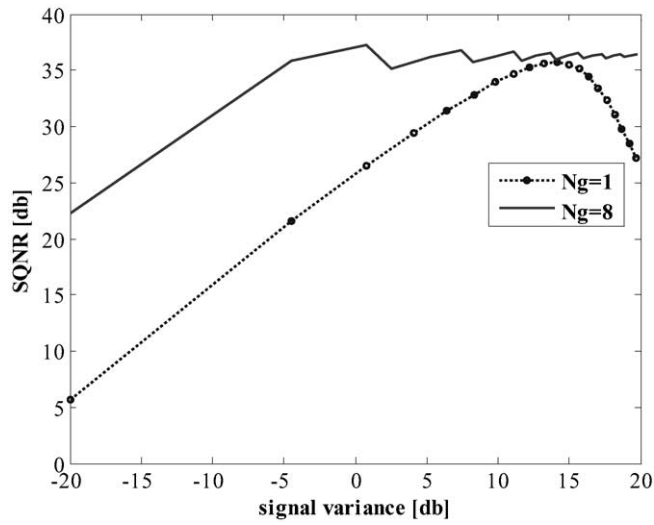


Fig. 2. Theoretical results: $SQNR$ dependence on the signal variance for $N = 128$ levels when $N_g = 1$ and $N_g = 8$ levels uniform scalar quantizer is used.

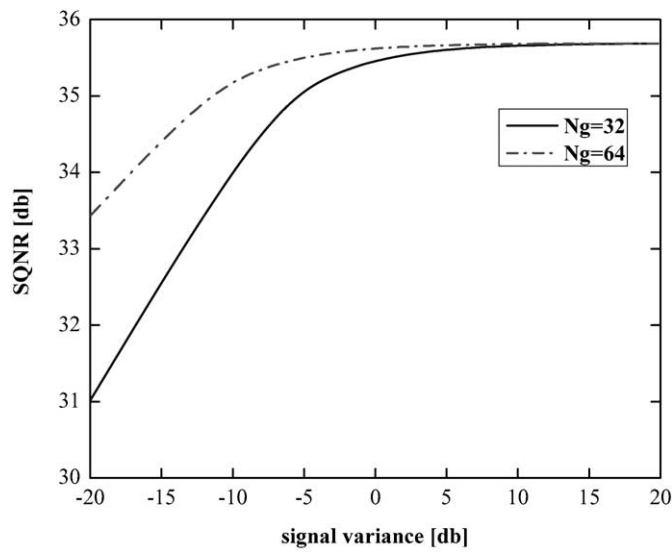


Fig. 3. Theoretical results: $SQNR$ dependence on the signal variance for $N = 128$ levels when $N_g = 32$ and $N_g = 64$ levels uniform scalar quantizer is used.

information and effectiveness of adaptation produces a good selection of the learning period.

As we already mentioned we perform our analysis when the frame length is $M = 200$ and dynamic range of the speech signal is 40 dB. For such a case using Eq. 15 we provide the $SQNR$ dependence on signal variance when $N = 128$ levels Lloyd–Max’s quantizer

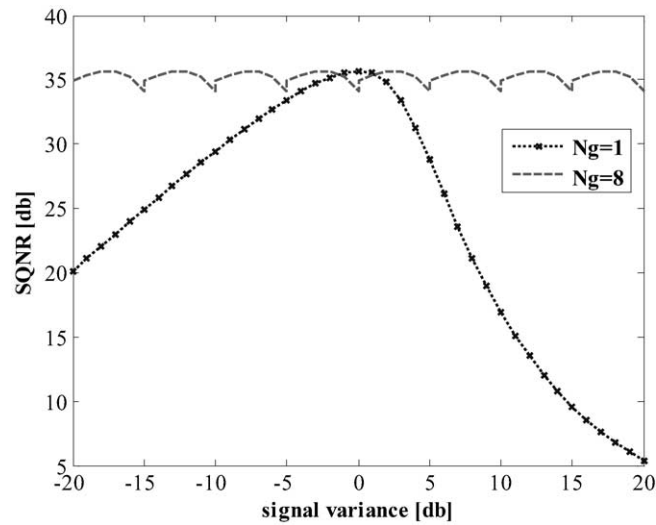


Fig. 4. Theoretical results: $SQNR$ dependence on the signal variance for $N = 128$ levels when $N_g = 1$ and $N_g = 8$ levels log-uniform scalar quantizer is used.

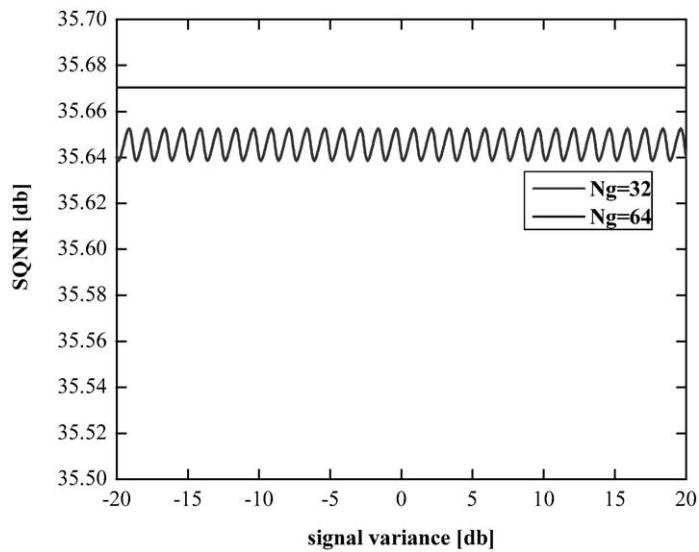


Fig. 5. Theoretical results: $SQNR$ dependence on the signal variance for $N = 128$ levels when $N_g = 32$ and $N_g = 64$ levels log-uniform scalar quantizer is used.

is considered and $N_g = 1$ and $N_g = 8$ levels uniform scalar quantizer are used (see Fig. 2). Using the same equation we provide the $SQNR$ dependence on signal variance when again $N = 128$ levels Lloyd–Max’s quantizer is considered and $N_g = 1$ and $N_g = 8$ levels log-uniform scalar quantizer are used (see Fig. 4). Moreover, we provide the $SQNR$ dependence on signal variance for the same $N = 128$ levels Lloyd–Max’s

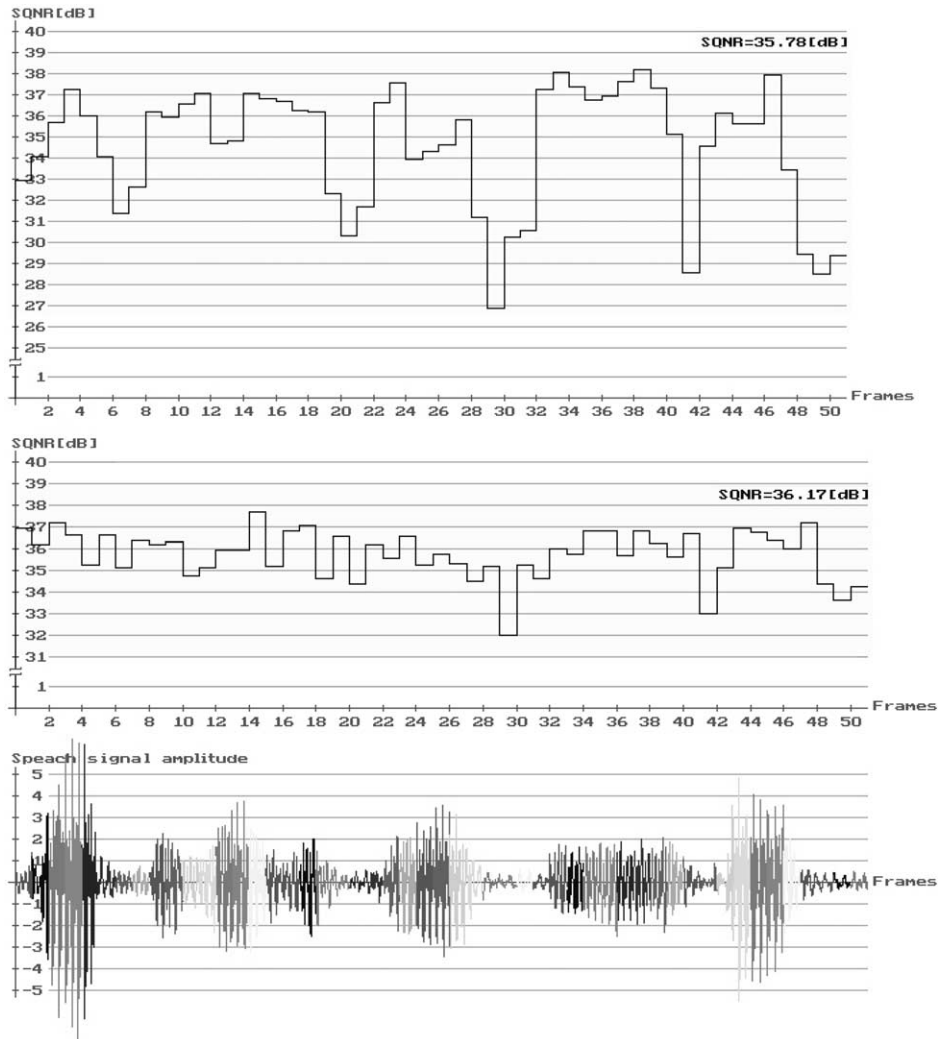


Fig. 6. Experimental results: $SQNR$ dependence on the frames for $N = 128$ levels when:
 a) $N_g = 32$ levels log-uniform scalar quantizer is used for gain quantizing and
 b) $N_g = 64$ levels log-uniform scalar quantizer is used for gain quantizing.

quantizer when the uniform scalar quantizer as well as log-uniform scalar quantizer have $N_g = 32$ and $N_g = 64$ levels (see Fig. 3 and Fig. 5). Since the more constantly $SQNR$ in a wide dynamic range of the input signal is achieved by using log-uniform quantizer we destine to consider such a quantizer. Therefore, using Eqs. 16 and 20 we provide theoretical and experimental values of average signal to quantization noise ratios when the log-uniform scalar quantizer, used for gain quantizing, have different number of quantization levels $N_g = 1$, $N_g = 2$, $N_g = 4$, $N_g = 8$, $N_g = 16$, $N_g = 32$, $N_g = 64$ (see Tables 1–3). Additionally, we tabulate the number of bit/sample that corresponds to the

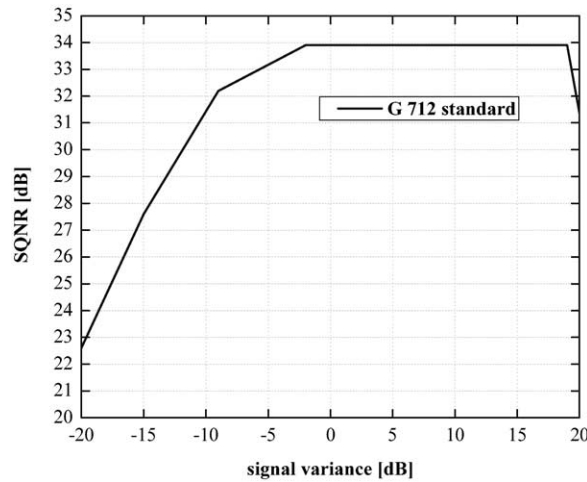


Fig. 7. $SQNR$ dependence on the signal variance defined by ITU-T G. 712 standard.

particular speech coding scheme. Finally, in order to have better insight in the performances of the considered speech coding scheme we provide (using Eq. 17) the values of the signal to quantization noise ratio within the each of F frames (see Fig. 6).

It is important to point out that experimental results overreach the optimal $SQNR$ value that corresponds to optimal Lloyd–Max quantizer (Jayant and Noll, 1984). Observe that the reason of such overreaching for $N_g = 32$ and $N_g = 64$ when log-uniform scalar quantizer is used for gain quantizing is already explained in (Jayant and Noll, 1984) for the case of uniform adaptive coders. Recall that the main reason of such overreaching follows from the fact that optimal results are obtained in case of stationary input signal while we perform our experiment assuming the local-stationary input signal. Observe that the-

Table 1

Experimental and theoretical results for $N = 128$ levels when N_g -level log-uniform scalar quantizer is used for gain quantizing

$N = 128$	Average $SQNR$ (optimal value $SQNR^{opt} = 35.69$)		Number of bits/sample
	N_g -level log-uniform quantizer	Experimental results	
1 (0bit/frame)	14.25	23.75	$7 + 0/200 = 7$
2 (1bits/frame)	19.76	30.39	$7 + 1/200 = 7.005$
4 (2bits/frame)	25.55	33.99	$7 + 2/200 = 7.010$
8 (3 bits/frame)	30.84	35.17	$7 + 3/200 = 7.015$
16 (4 bits/frame)	35.00	35.58	$7 + 4/200 = 7.020$
32 (5 bits/frame)	35.78	35.65	$7 + 5/200 = 7.025$
64 (6 bits/frame)	36.17	35.67	$7 + 6/200 = 7.030$

Table 2

Experimental and theoretical results for $N = 64$ levels when N_g -level log-uniform scalar quantizer is used for gain quantizing

$N = 64$	Average $SQNR$ (optimal value $SQNR^{opt} = 29.74$)		Number of bits/sample
	N_g -level log-uniform quantizer	Experimental results	
1 (0bit/frame)	12.10	19.31	$6 + 0/200 = 6$
2 (1bits/frame)	13.72	25.07	$6 + 1/200 = 6.005$
4 (2bits/frame)	19.52	28.21	$6 + 2/200 = 6.010$
8 (3 bits/frame)	24.92	29.32	$6 + 3/200 = 6.015$
16 (4 bits/frame)	28.82	29.62	$6 + 4/200 = 6.020$
32 (5 bits/frame)	29.94	29.71	$6 + 5/200 = 6.025$
64 (6 bits/frame)	30.31	29.73	$6 + 6/200 = 6.030$

Table 3

Experimental and theoretical results for $N = 32$ levels when N_g -level log-uniform scalar quantizer is used for gain quantizing

$N = 32$	Average $SQNR$ (optimal value $SQNR^{opt} = 23.87$)		Number of bits/sample
	N_g -level log-uniform quantizer	Experimental results	
1 (0bit/frame)	6.55	14.77	$5 + 0/200 = 5$
2 (1bits/frame)	7.20	19.82	$5 + 1/200 = 5.005$
4 (2bits/frame)	13.59	22.50	$5 + 2/200 = 5.010$
8 (3 bits/frame)	19.03	23.48	$5 + 3/200 = 5.015$
16 (4 bits/frame)	23.05	23.75	$5 + 4/200 = 5.020$
32 (5 bits/frame)	24.00	23.84	$5 + 5/200 = 5.025$
64 (6 bits/frame)	24.15	23.86	$5 + 6/200 = 5.030$

oretical results do not overreach the appropriate optimal $SQNR$ value since the theoretical results are carried out assuming equality of Eqs. 12 and 13. Moreover, by increasing the number of bits required for sending the side information (information about the gain) the performances of the considered speech coding scheme approaches to optimal. It is very important to notice that the performances ($SQNR$) obtained by using the proposed speech coding algorithm are closer to the optimal and the less bits/sample are required (approximately about 1 bits/sample compression) when compared to adaptive coders designed according to G. 711 standard (ITU-T, Recommendation G. 711, 1972).

Finally, the selection of the particular speech coding scheme can be performed according to the introduced criteria Eq. 21. Observe that when $N_g = 64$ level log-uniform scalar quantizer is used for gain quantizing experimental results demonstrate that the in-

roduced criteria is satisfied, while theoretical results satisfied criteria when $N_g = 32$ level log-uniform scalar quantizer is used. Since the theoretical results for $N_g = 64$ are closer to optimal we destine to adopt as better solution the speech coding algorithm consisting of $N_g = 64$ level scalar quantizer. Finally, considering the Figs. 2, 3, 4 and 5 it is obvious that the best performances of the considered speech coding schemes can be achieved by using $N_g = 64$ level log-uniform scalar quantizer ($SQNR$ constant in a wide dynamic range of input signal). Therefore we propose implementation in practice of the speech coding scheme consisting of $N_g = 64$ level log-uniform scalar quantizer. Finally, it is important to point out that the proposed speech coding scheme consisting of $N_g = 64$ level log-uniform scalar quantizer satisfies G. 712 standard (see Fig. 7) for a wide dynamic range of the input signal.

Conclusion

In this paper we proposed a novel speech coding algorithm based on forward adaptive technique in which adaptive Lloyd–Max’s quantizer is implemented and designed according to the method proposed in (Peric and Nikolic, 2007). In order to have accurate insight in the performances of the adaptive coders designed according to the proposed algorithm we have performed the comparison of such performances with the theoretically optimal performances (Jayant and Noll, 1984). Moreover, we discussed the performances of speech coding schemes designed according to G. 711 standard and argue that with the proposed algorithm the nearest to the optimal performances can be achieved while obtaining approximately about 1bits/sample compression. We provided theoretical and experimental results (performances of our algorithm). We demonstrated that experimental results ($SQNR_a^{ex}$) overreach the optimal results ($SQNR^{opt}$) and explained that the reason of such overreaching follows from the fact that we performed our experiment assuming the local-stationary input signal while optimal results are obtained in case of stationary input signal. Further, we demonstrated that theoretical results approach to optimal results by increasing the side information. In order to limit the increase of side information as well as to find the particular speech coding scheme that for a fixed number of quantization levels N provides as high as possible average $SQNR$ ($SQNR_a$ or $SQNR_a^{ex}$) we introduced the criteria which do not allow further increase of the number of bits for side information if the difference $\Delta SQNR_a$ is less than 0.03. Finally, in order to find better solution for implementation of the proposed algorithm in practice we considered the performances of our algorithm when log-uniform as well as uniform scalar quantizer are used for gain quantizing. We demonstrated that in a wide dynamic range of input signal constant $SQNR$ can be achieved when $N_g = 64$ level log-uniform scalar quantizer is used for gain quantizing. Additionally, we demonstrated that afore considered speech coding scheme satisfies G. 712 standard for a wide dynamic range of the input signal. Therefore we hope that our algorithm provides a good solution for the practical realization of adaptive speech coders. Particularly, we believe that adaptive speech coders designed according to our algorithm can be widely used for voice transmission over internet (Minoli, 2002; Hersent *et al.*, 2005), as well as in public switched telephone networks (PSTN) where the high quality voice signal represented with less than 8 bits/sample are required.

References

- Chu, W.C. (2003). *Speech Coding Algorithms*. John Wiley & Sons, New Jersey, Chapters 5–6, pp. 143–183.
- Gersho, A. (1994). Advances in speech and audio compression. In *Proceedings of the IEEE*, Vol. 82(6), pp. 900–918.
- Gersho, A., and R.M. Gray (1992). *Vector Quantization and Signal Compression*. Kluwer, Academ. Pub., Chapters 5–6, pp. 133–202.
- Hersent, O., J. Petit and D. Gurle (2005). *Beyond VoIP Protocols – Understanding Voice Technology and Networking Techniques for IP Telephony*. John Wiley & Sons, New Jersey, Chapters 1–2, pp. 1–88.
- Hiwasaki, Y., H. Ohmuro, T. Mori, S. Kurihara and A. Kataoka (2006). A G. 711 embedded wideband speech coding for VoIP conferences. *IEICE Transactions on Info and Systems*, **E89-D(9)**, 2542–2552.
- ITU-T, Recommendation G. 711 (1972). *Pulse Code Modulation (PCM) of Voice Frequencies*. International Telecommunication Union.
- ITU-T, Recommendation G. 712 (1992). *Transmission Performance Characteristics of Pulse Code Modulation (PCM)*. International Telecommunication Union.
- Jayant, N.S., and P. Noll (1984). *Digital Coding of Waveforms*. Prentice-Hall, New Jersey, Chapter 4, pp. 129–139.
- Khasnabish, B. (2003). *Implementing Voice over IP*. John Wiley & Sons, New Jersey.
- Kondo, A. (2004). *Digital Speech, Coding for Low Bit Rate Communication Systems*. John Wiley & Sons, New Jersey.
- Minoli, D. (2002). *Voice over MPLS – Planning and Designing Networks*. McGraw-Hill, Chapters 1–2, pp. 1–134.
- Ortega, A., and M. Vetterli (1997). Adaptive scalar quantization without side information. *IEEE Transactions on Image Processing*, **6(5)**, 665–676.
- Pan, D. (1993). Digital audio compression. *Digital Technical Journal*, **5(2)**.
- Peric, Z., and J. Nikolic (2007). An effective method for initialization of Lloyd–Max’s algorithm of optimal scalar quantization for Laplacian source. *Informatika*, **18(2)**, 279–288.
- Pohlman, K. (1989). *Principles of Digital Audio*. Howard W. Sams and Co., Indianapolis.
- Rabiner, L., and R. Schafer (1978). *Digital Processing of Speech Signals*. Prentice-Hall.

J.R. Nikolic was born in Prokuplje, Yugoslavia, in 1978. She received the BSc degree in telecommunications from the Faculty of Electronic Engineering, Nis, in 2003, and MSc degree in telecommunications from the University of Nis, in 2006. She is currently teaching assistant on Faculty of Electronic Engineering at the Department of Telecommunications.

Her current research interests include the information theory, source coding and signal processing. She is particularly working on scalar quantization techniques in speech. She was author and coauthor in over 30 papers in digital communications.

Z.H. Peric was born in Nis, Yugoslavia, in 1964. He received the BSc degree in electronics and telecommunications from the Faculty of Electronic Engineering, Nis, Serbia, Yugoslavia, in 1989, and MSc degree in telecommunications from the University of Nis, in 1994. He received the PhD degree from the University of Nis, also, in 1999.

He is currently professor at the Department of Telecommunications and vice-dean of the Faculty of Electronic Engineering, University of Nis, Serbia. His current research interests include the information theory, source and channel coding and signal processing. He is particularly working on scalar and vector quantization techniques in speech and image coding. He was author and coauthor in over 150 papers in digital communications. Dr Peric has been a reviewer for IEEE Transactions on Information Theory. He is member editorial board of journal "Electronics and Electrical Engineering".

Lloyd–Max’s algoritmo įgyvendinimas šnekos kodavimo algoritme, grindžiamame ėjimo į priekį adaptyviu metodu

Jelena NIKOLIC, Zoran PERIC

Straipsnyje detaliai analizuojamas šnekos kodavimo algoritmas, grindžiamas ėjimo į priekį adaptyvaus metodo panaudojimu. Nagrinėjamas algoritmas, operuojantis kadrais, kuriuose pateikiamas tam tikras šnekos pavyzdžių imčių skaičius. Kaupiant kadrus stiprinimo įvertis apibrėžiamas kaip kvadratinė šaknis iš kadrų imčių dispersijos. Informacija apie stiprinimą ir neadaptyvaus kvantuoklio kodinę knygą, kuri sukurta vienietinės ėjimo signalo dispersijos atveju, yra toliau naudojama kuriant adaptyvų kvantuoklį. Tokiu būdu atliekama geresnė kvantuoklio adaptacija besikeičiančiai ėjimo statistikai. Pastebėsime, kad šio straipsnio tikslas yra tirti, kad plačiame dispersijos kitimo intervale pirmenybė galėtų būti pasiekta panaudojant ėjimo į priekį adaptyvų kalbos kodavimo algoritmą, neseniai sukurta efektyvų metodą Lloyd–Max’s algoritmo inicializacijai, kuri duoda optimalų Lloyd–Max’s kvantuoklio veikimą ėjimo signalo vienietinės dispersijos atveju. Mes numatome nagrinėti kalbos kodavimo algoritmą, paremtą ėjimo į priekį adaptyviu metodu, kadangi grįžimo atgal adaptacija duoda SQNR (signalų ir kvantavimo triukšmo santykį) tiesioginės adaptacijos 1 dB ribose. Mes pateikiame teorinius ir eksperimentinius rezultatus (mūsų algoritmo darbingumą), kuris yra palyginamas su optimaliais rezultatais. Priedo, mes nagrinėjame kalbos kodavimo schemas, sukurtas pagal G 711 standartą, darbingumą ir pabrėžiame naudą, kuri gali būti gauta naudojant mūsų algoritmą. Galiausiai, kad rasti geresnius sprendimus pasiūlyto algoritmo įgyvendinimui praktikoje, mes nagrinėjame mūsų algoritmo darbingumą, kai logaritmiškai tolygus, o taip pat tolygus skaliarinis kvantuoklis yra naudojami stiprinimo kvantavimui.