

Three-Level Delta Modulation for Laplacian Source Coding

Bojan DENIC¹, Zoran PERIC¹, Vladimir DESPOTOVIC²

¹University of Nis, Faculty of Electronic Engineering, Aleksandra Medvedeva 14, 18000 Nis, Serbia

²University of Belgrade, Technical Faculty in Bor, Vojske Jugoslavije 12, 19210 Bor, Serbia
bojan.denic@elfak.rs

Abstract—This paper proposes a novel solution for coding of time varying signals with Laplacian distribution, which is based on delta modulation and three-level quantization. It upgrades the conventional scheme by introducing the quantizer with variable length code. Forward adaptive scheme is used, where the adaptation to the signal variance is performed on frame-by-frame basis. We employ configurations with simple fixed first-order predictor and switched first-order predictor utilizing correlation. Furthermore, we propose different methods for optimizing predictor coefficients. The configurations are tested on speech signal and compared to an adaptive two-level and four-level delta modulation, showing that proposed three-level delta modulation offers performance comparable to a four-level baseline with significant savings in bit rate.

Index Terms—delta modulation, Huffman coding, predictive coding, signal to noise ratio, speech coding.

I. INTRODUCTION

A signal is considered to be nonstationary if its frequency or spectral contents are changing with respect to time. Nonstationary signals are in general unpredictable and very hard to model or forecast. In order to receive consistent, reliable results, the nonstationary signals can be transformed into quasi-stationary ones, which are characterized by a nearly constant mean and variance over time. Speech is an example of nonstationary signal, but its frequency content varies slowly with time. Thus, the signal can be analyzed over a short-duration interval (frame or window) where it is considered to be quasi-stationary. It was shown that the distribution of speech signal is best modeled by Laplacian distribution for frame lengths shorter than 200 ms [1].

Delta modulation (ΔM) is an attractive technique that has been employed in various applications in signal processing. It is a simplified version of differential pulse code modulation (DPCM) and the class of predictive coding [2-6]. The standard ΔM architecture involves one-bit quantizer along with the fixed first-order predictor [5]. A good choice of parameters known as step size and sampling frequency can lead to an efficient implementation of ΔM system.

A vast amount of investigations are conducted over the years to enhance the performance of the standard (fixed size) ΔM . In the adaptive delta modulation (A ΔM) the step size is adapted to signal statistics (e.g. variance) for each signal frame [7]. If better quality is required, the same adaptation can be employed using modified Pulse Code Modulation (PCM) configuration with higher number of

quantization levels [8]. The analysis of nonadaptive and adaptive two-bit ΔM systems is presented in [9], [10], emphasizing the performance improvement over the standard one at the cost of higher complexity. Another special type of delta modulation called sigma delta modulation (S ΔM) has also been reported in the literature [11], where the short-term performance examination of S ΔM has been conducted, showing that its performance is highly correlated with the variation of input signal. The solution presented in [12] offers gain in performance (i.e. higher resolution and linearity, increased dynamic range) by applying high-level (three-bit) quantization.

Three-level quantization has been used in different areas of signal processing [5], [13-15]. Its application to coding for low-energy sensors in sensor network systems is discussed in [13]. An effective implementation of three-level quantization in ECG signal compression has been demonstrated in [14]. Furthermore, its incorporation in S ΔM structure, for the purpose of arithmetic processing has been shown in [15]. To the best knowledge of the authors, an implementation of three-level quantization in delta modulation system has not yet been reported for coding of Laplacian sources. We propose a three-level scalar quantizer designed for variable length code-words (VLC) employing Huffman coding, since it is highly efficient when operating with low number of quantization levels [3]. The representative levels are determined as centroids of their respective cells, as in the case of the Lloyd-Max quantization [2-4]. The following criteria should be fulfilled: the criterion of minimal distortion and the simultaneous criterion of minimal distortion and minimal bit rate. The appropriate performance discussion of different types of quantization, when Huffman coding is incorporated can be found in [16].

Furthermore, we proposed two configurations of delta modulation, based on two variants of three-level quantization, applied in the forward adaptive coding scheme [7], [17]. In contrast to the backward adaptation technique, forward adaptation provides gain in Signal to Quantization Noise Ratio (SQNR) in the wide dynamic range [18], as well as less sensitivity to transmission error [2]; however, it requires transfer of side information to the receiving part. The system is operating on frame-by-frame basis, where the quantizer is adapted to the short-term estimate of the frame variance.

Memoryless Laplacian source is often used for modeling of speech, image, video and bio-medical signals. In this paper, we use speech signal to test the performance of proposed configurations in a real environment.

This work was supported in part by the by the Ministry of Education and Science of the Republic of Serbia, grant no. TR32051 and TR32035, within the Technological Development Program.

In the first configuration, the fixed first-order predictor is considered. We propose the choice of optimal fixed-predictor coefficient to provide the highest performance measured by *SNR*. In the second configuration the switched first-order predictor based on correlation is used instead of fixed predictor. One effective way to use the correlation between samples within a frame is studied in [19], where the analysis is done for high speech signal quality by employing DPCM scheme with the switched first-order predictor. In our solution, the switched predictor has two fixed values of predictor coefficients on disposal, one for weakly correlated and one for highly correlated frames. The main idea presented here implies determination of the correlation coefficient for a particular frame of the input signal, which is used for selection of the appropriate predictor coefficient.

The performances of the proposed schemes are evaluated using *SNR* and Perceptual Evaluation of Speech Quality (*PESQ*) [20], as well as the bit rate. We use the adaptive two-level (conventional scheme) and four-level delta modulation schemes as baselines. Furthermore, the attained *PESQ* scores are compared to the wideband rate distortion bound given in [21], which is the lower bound of performance of several standardized ADPCM speech codecs for wideband speech, including G.722 standard [22] and its extensions.

The remaining of this paper is organized as follows: in Section II we present a three-level scalar quantizer design. In Section III the proposed configurations based on delta modulation are introduced. In Section IV the experimental results obtained using a real speech signal are presented and discussed. Finally, concluding remarks are given in Section V.

II. THREE-LEVEL QUANTIZER DESIGN

A scalar quantizer with $N=3$ levels is specified by the set of real numbers t_1, t_2 , called decision thresholds, satisfying $-\infty = t_0 < t_1 < t_2 < t_{N=3} = +\infty$ and set of numbers y_1, y_2, y_3 ,

called representation levels, satisfying $y_i \in \alpha_i = [t_{i-1}, t_i)$, for

$i = 1, 2, 3$. With α_i is denoted the quantization cell, where $\bigcup_{i=1}^{N=3} \alpha_i \in R$ and $\alpha_{i-1} \cap \alpha_i = 0$. The quantizer is defined as many-to-one mapping $Q : R \rightarrow \{y_1, y_2, y_3\}$ defined by $Q(x) = y_i$ where $x \in \alpha_i$. Additionally, for the assumed nonlinear input source, cell α_2 forms the granular region and is called granular cell, while α_1 and α_3 constitute an overload region and are called overload cells.

In this paper, we deal with the symmetrical three-level quantizer involving zero level y_2 . Due to the symmetry, threshold t_1 and level y_1 in the negative part of the quantizer characteristic are symmetrical to their counterparts in the positive part, i.e. the equations $-t_1=t_2$ and $-y_1=y_3$ will hold. Hence, only the positive threshold and level need to be determined.

Assuming the Laplacian source with zero mean and unit variance, the probability density function (pdf) becomes [1-4].

$$p(x) = \frac{1}{\sqrt{2}} \exp(-\sqrt{2}|x|) \quad (1)$$

Given pdf, the representative level y_3 can be determined as the centroid of the respective cell $\alpha_3=(t_2, \infty)$ [2-4].

$$y_3 = \frac{\int_{t_2}^{\infty} xp(x) dx}{\int_{t_2}^{\infty} p(x) dx} = t_2 + \frac{1}{\sqrt{2}} \quad (2)$$

The mean squared distortion D is usually used as a measure of irreversible error introduced during the quantization process. It can be expressed as the sum of distortions in the granular and overload region [2-4].

$$D = 2 \int_0^{t_2} x^2 p(x) dx + 2 \int_{t_2}^{\infty} (x - y_3)^2 p(x) dx = 1 - \exp(-\sqrt{2}t_2) \left(\frac{1}{2} + \sqrt{2}t_2 + t_2^2 \right) \quad (3)$$

We use *SQNR* as a measure of performance. Assuming the unit variance case it is determined as [2-4].

$$SQNR = 10 \log_{10} \left(\frac{1}{D} \right) \quad (4)$$

Let us denote by p_1, p_2 and p_3 the probabilities of occurrence of representative levels y_1, y_2 and y_3 , respectively.

$$p_2 = 2 \int_0^{t_2} p(x) dx = 1 - \exp(-\sqrt{2}t_2) \quad (5)$$

$$p_3 = p_1 = \int_{t_2}^{\infty} p(x) dx = \frac{1}{2} \exp(-\sqrt{2}t_2) \quad (6)$$

The bit rate R can be calculated as [3].

$$R = \sum_{i=1}^{N=3} p_i l_i \quad (7)$$

where l_i is the length of the Huffman codeword corresponding to the output y_i .

Equations (2)-(7) show that performance of the proposed quantizer is strongly dependent on decision threshold t_2 . Therefore, its corresponding value will be numerically determined in accordance with the criterion of minimal distortion (criterion 1) and the simultaneous criterion of minimal distortion and minimal bit rate (criterion 2).

Fig. 1 illustrates the bit rate R versus the distortion D for the proposed quantization model. Point A on the curve indicates the position of a three-level quantizer designed according to the criterion 1 ($t_2=0.707$), whereas the point B denotes the position when it is designed with the criterion 2 ($t_2=0.887$). We observe that the bit rate approaches to one as distortion tends to variance of the input signal (case $t_2 \gg 1$) [23].

Fig. 2 shows *SQNR* obtained for the Lloyd-Max quantization with $N=2$ and $N=4$ levels having bit rates $R=1$ bit/sample and $R=2$ bit/sample, respectively [2], [3]. The marked points above the curve indicate the achieved performance of three-level quantizer designed according to the aforementioned criteria. Note that the proposed model of quantizer satisfying the criterion 1 provides 1.1 dB higher

SQNR when compared to the expected *SQNR* value (specified by a point on a curve) for the same bit rate. In addition, when the criterion 2 is satisfied, the gain of 1.31 dB can be perceived.

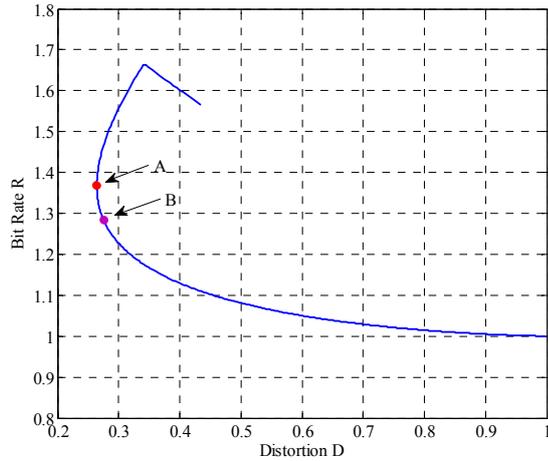


Figure 1. The dependence of bit rate *R* on distortion *D*

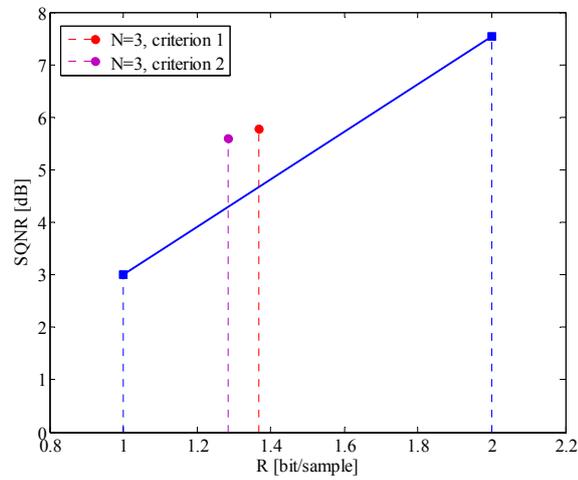


Figure 2. Performance of the proposed three-level quantizer compared to the Lloyd-Max quantizer with *N*=2 and *N*=4 levels

The detailed description of delta modulation configuration with fixed and switched prediction is given in the following section, where the proposed quantizer satisfying both criteria is implemented.

III. THE PROPOSED ADAPTIVE DELTA MODULATION SCHEMES

A. Delta Modulation with a Fixed Predictor and an Adaptive Three-Level Quantizer

A simple delta modulation scheme with a fixed first-order predictor is depicted in Fig. 3. In this scenario, the quantizer described in Section II is applied to the forward adaptive coding scheme that consists of a buffer, a variance estimator, a log-uniform quantizer having *L* levels and an adaptive three-level quantizer. The adaptation to the short-term estimate of the variance is performed for each frame of the signal difference. Consequently, it is necessary to modify the codebook frame-wise. The following procedure was conducted. The buffer was used for storage of *M* samples of the signal *e*[*n*], where *M* is a frame size and *e*[*n*] = *x*[*n*] - \hat{x} [*n*] is a prediction error. The predicted signal \hat{x} [*n*] = *a* · *x*[*n* - 1] is provided at the output of the fixed predictor. The first sample in each frame *x*[1] is predicted using the last sample from previous frame *x*[0] (i.e. the overlap between frames is one sample), except the first frame where *x*[0] = 0.

The average variance σ_e^2 for each frame is calculated in the variance estimator.

$$\sigma_e^2 = \frac{1}{M} \sum_{i=1}^M e_i^2 \quad (8)$$

and used for adaptation to the short-term signal statistics. Information about the variance has to be available at the decoder side for parameters scaling, hence it is quantized using the log-uniform quantizer (*Q_{LU}*). It was shown in [18] that such quantizer is preferred compared to the uniform one due to its improved performance in a wide dynamic range. It is designed to quantize logarithmic variance $10 \log_{10}(\sigma_e^2 / \sigma_{ref}^2)$ in the assumed range of (-25 dB, 25 dB), where σ_{ref}^2 is the reference variance.

In logarithmic domain, its thresholds are determined as.

$$l_i[\text{dB}] = -25 + \Delta_L i, \quad i = 0, \dots, L \quad (9)$$

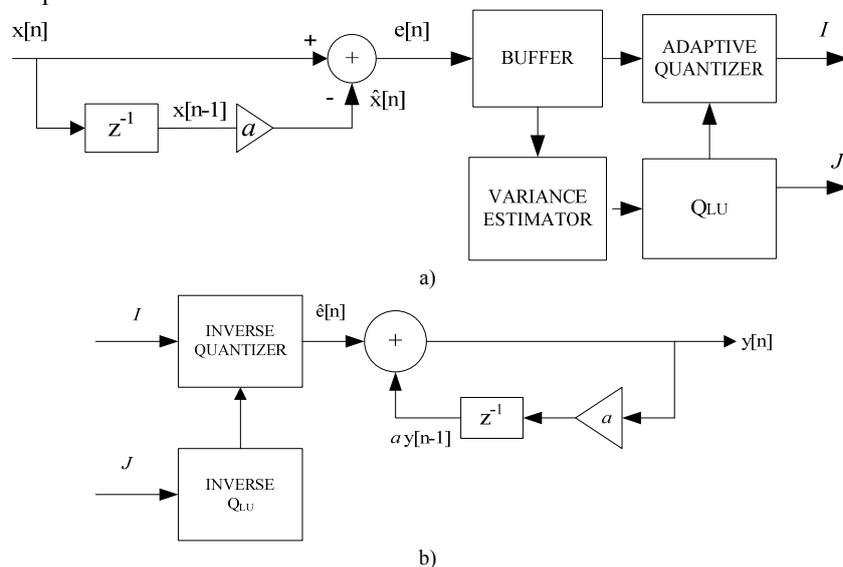


Figure 3. Delta modulation scheme with a fixed predictor and adaptive three-level quantizer: a) encoder, b) decoder

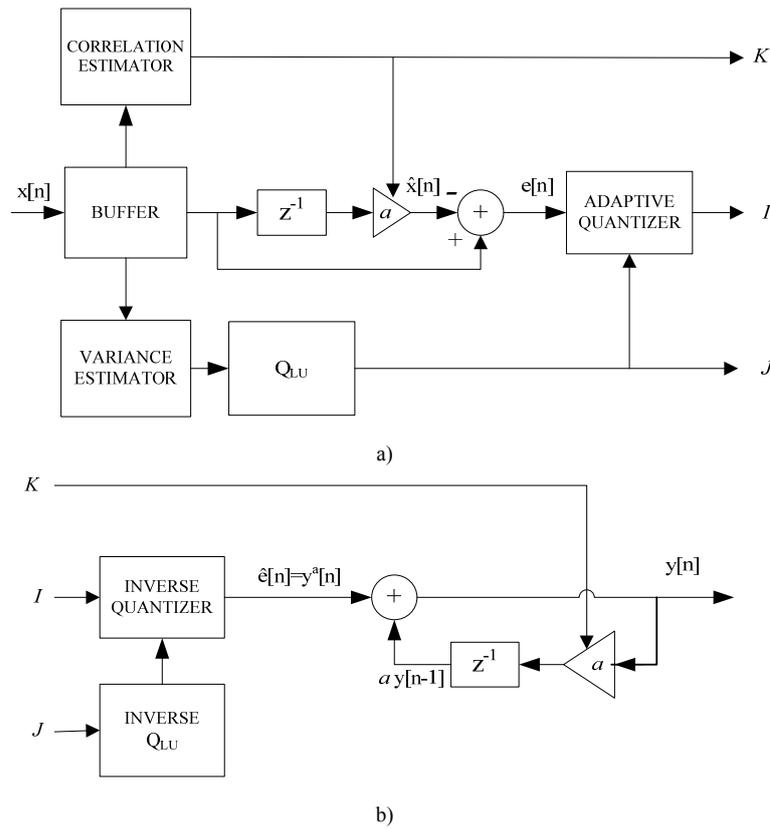


Figure 4. Delta modulation scheme with a switched predictor and adaptive three-level quantizer: a) encoder, b) decoder

and levels are determined as.

$$\hat{l}_i[\text{dB}] = -25 + \Delta_L \left(i - \frac{1}{2} \right), \quad i = 1, \dots, L \quad (10)$$

where $\Delta_L = 50/L$ is the step size.

In the linear domain, thresholds and levels are given by (11) and (12), respectively.

$$r_i = \sigma_{ref}^2 10^{\frac{l_i}{10}}, \quad i = 0, \dots, L \quad (11)$$

$$\hat{r}_i = \sigma_{ref}^2 10^{\frac{\hat{l}_i}{10}}, \quad i = 1, \dots, L \quad (12)$$

The quantization rule is given as $Q_{LU}(\sigma_e^2) = \hat{r}_i$ when $\sigma_e^2 \in (r_{i-1}, r_i)$.

The representative level y_3^a and the decision threshold t_2^a , for $\sigma_e^2 \in (r_{i-1}, r_i)$ are determined as: $y_3^a = g_i \cdot y_3^f$ and $t_2^a = g_i \cdot t_2^f$, where $g_i = \sqrt{\hat{r}_i}$ and y_3^f and t_2^f denote the level and the threshold of fixed three-level quantizer (designed for unit variance), respectively.

On the encoder output we have two digital signals I and J that are transferred to the decoder for each frame, as illustrated in Fig. 3b. Index I represents code-words for M samples in the current frame. Index J consists of $\log_2 L$ bits per frame and it contains information about the level of Q_{LU} that has been used for frame variance quantization. Note that index J carries additional or side information.

Then, the bit rate is given by.

$$R^I = R^f + \frac{\log_2 L}{M} \quad (13)$$

where R^f is a bit rate of fixed three-level quantizer.

In addition, Fig. 3b shows that the reconstructed signal

$y[n]$ can be obtained as.

$$y[n] = a \cdot y[n-1] + y^a[n] \quad (14)$$

where $y^a[n]$ is the output of the adaptive quantizer.

B. Delta Modulation with a Switched Predictor and an Adaptive Three-Level Quantizer

In order to further improve the performance of the solution presented in Section 3.A, we have introduced a model that involves switched first-order predictor and adaptive three-level quantization, as shown in Fig. 4. The switched predictor employs one of two available coefficients a_1 and a_2 based on the correlation coefficient ρ estimated for each frame of the input signal.

$$\rho = \frac{\sum_{i=1}^{M-1} x_i x_{i+1}}{\sum_{i=1}^M x_i^2} \quad (15)$$

where M is the frame size.

The predictor makes a selection between two possible values: if $\rho < 0.8$ the input frame is classified as weakly correlated and the switched predictor takes the coefficient a_1 , otherwise the frame is considered as highly correlated and the coefficient a_2 is selected. The values a_1 and a_2 should be optimal for weakly correlated and highly correlated frames, respectively. By using the predictor coefficients that are optimal for the available signal one can get performance improvements.

After selecting the appropriate value of the predictor coefficient, the same procedure is conducted as in the solution given in Section 3.A. Note that adaptation to the signal variance $\sigma_e^2 = \sigma_e^2 (1 - \rho^2)$ is performed for each frame,

where σ_x^2 is the variance of input signal and σ_e^2 is the variance of prediction error.

The bit rate for the proposed scheme is.

$$R'' = R^f + \frac{1 + \log_2 L}{M} \quad (16)$$

The number of bits per frame (side information) is increased by one in regard to the equation (13), due to the fact that information about the selected predictor coefficient (index K in Fig. 4) needs to be transferred to the receiving part as well.

C. Application in Speech Coding and Evaluation Metrics

It has been found that the Laplacian distribution accurately models speech signal [1], hence we test the proposed configurations with speech. As an objective measure of performance, we use Signal to Noise Ratio (SNR) which is composed of two components: the prediction gain G and $SQNR$. Let us assume a speech signal involving F frames, where the particular frame contains M samples. $SQNR$ for the particular frame can be calculated as.

$$SQNR_e^j = 10 \log_{10} \left(\frac{(\sigma_e^j)^2}{D_e^j} \right), \quad j = 1, \dots, F \quad (17)$$

where $(\sigma_e^j)^2$ is the average variance of the j -th frame, as given in (8), and D_e^j is the average distortion of the j -th frame.

$$D_e^j = \frac{1}{M} \sum_{i=1}^M (e_{ji} - y_{ji}^a)^2, \quad j = 1, \dots, F \quad (18)$$

where y_{ji}^a are the outputs of adaptive quantizer.

Prediction gain G for the particular frame can be determined as.

$$G^j = 10 \log_{10} \left(\frac{\frac{1}{M} \sum_{i=1}^M x_{ji}^2}{\frac{1}{M} \sum_{i=1}^M e_{ji}^2} \right), \quad j = 1, \dots, F \quad (19)$$

where x_{ji} are the samples of input speech signal.

By averaging $SQNR$ and G given in (17) and (19) over all frames we obtain.

$$SNR_{DM} = \frac{1}{F} \sum_{j=1}^F SQNR_e^j + \frac{1}{F} \sum_{j=1}^F G^j \quad (20)$$

Let us assume that P out of F frames are classified as weakly correlated. Then, using the equation (20) we obtain.

$$SNR_{DM}^{(1)} = \frac{1}{P} \sum_{j=1}^P SQNR_e^j + \frac{1}{P} \sum_{j=1}^P G^j \quad (21)$$

$$SNR_{DM}^{(2)} = \frac{1}{F-P} \sum_{j=1}^{F-P} SQNR_e^j + \frac{1}{F-P} \sum_{j=1}^{F-P} G^j \quad (22)$$

Finally, SNR_{DM} of the solution presented in Section 3.B has the form.

$$SNR_{DM} = w SNR_{DM}^{(1)} + (1-w) SNR_{DM}^{(2)} \quad (23)$$

where $w = P/F$ is the probability of occurrence of weakly correlated frame.

In addition to SNR we use $PESQ$, which is an objective perceptual speech quality measure that assesses how speech

quality is perceived by people. $PESQ$ uses a perceptual model to convert the original and the degraded speech into an internal representation. The degraded speech is time-aligned with the original signal to compensate for the delay that may be caused by degradation. The difference in the representations of two signals is used by the cognitive model to estimate the Mean Opinion Score (MOS). $PESQ$ was standardized by the ITU-T P.862 standard [20] with extensions to wideband speech [24].

D. Selection of Optimal Predictor Coefficients

An analysis of the influence of predictor coefficients on performances of the proposed delta modulation configurations with an adaptive three-level quantization is carried out in order to determine the optimal coefficients which maximize SNR_{DM} . We analyze both the configurations with fixed predictor (see Section 3.A) and switched predictor (see Section 3.B).

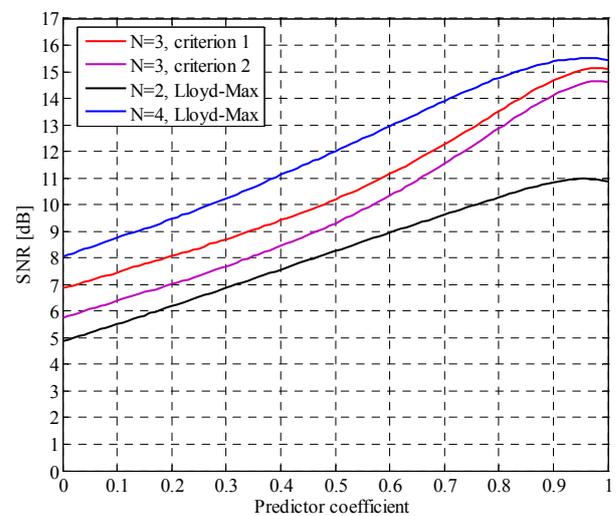


Figure 5. Selection of optimal predictor coefficient for various quantizers implemented in a delta modulation scheme with fixed predictor

The impact of the predictor coefficient choice on SNR_{DM} for frame size $M=80$ samples is illustrated in Fig. 5 for ΔM with a fixed predictor and an adaptive three-level quantization implemented according to the criterion 1 and criterion 2. The results are obtained using (20) when the prediction error signal $e[n]$ is formed using the fixed predictor coefficient a ranging from 0 to 1. Note that the maximum value of SNR_{DM} in this example was found for $a=0.98$. We observe that the proposed three-level ΔM attains significantly higher SNR_{DM} value compared to the one achieved by two-level ΔM (Lloyd-Max quantization with $N=2$ levels), and very close to the one achieved by four-level ΔM (Lloyd-Max quantization with $N=4$ levels).

Regarding the configuration depicted in Fig. 4, we have established two methods for choosing optimal coefficients a_1 and a_2 of the switched predictor. The first method assumes that weakly correlated frames ($\rho < 0.8$) and highly correlated frames ($\rho \geq 0.8$) are processed separately using scheme in Fig. 3, and optimal coefficients are determined according to the criterion of maximal SNR_{DM} (using (20)) in the specific region (see also Fig. 6). Note that coefficients $a_1=0.12$ and $a_2=0.99$ satisfy the required criterion for frame size $M=80$ samples.

The second method for determination of switched

predictor coefficients is based on the correlation coefficient of the input speech. Thus, the value a_1 is obtained as the average correlation coefficient calculated over all weakly correlated frames in the available speech signal. Vice-versa, the value a_2 corresponds to the average correlation coefficient calculated over all highly correlated frames. Optimal predictor coefficients using this criterion, for $M=80$ samples, are $a_1=0.23$ and $a_2=0.95$.

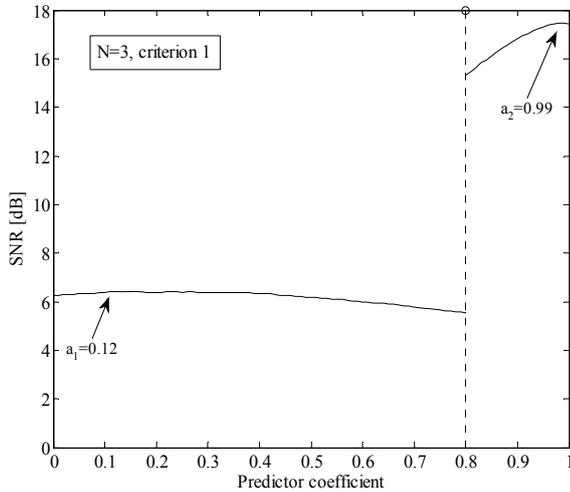


Figure 6. Selection of optimal values of switched predictor coefficients when three-level quantizer is implemented in ΔM with switched predictor

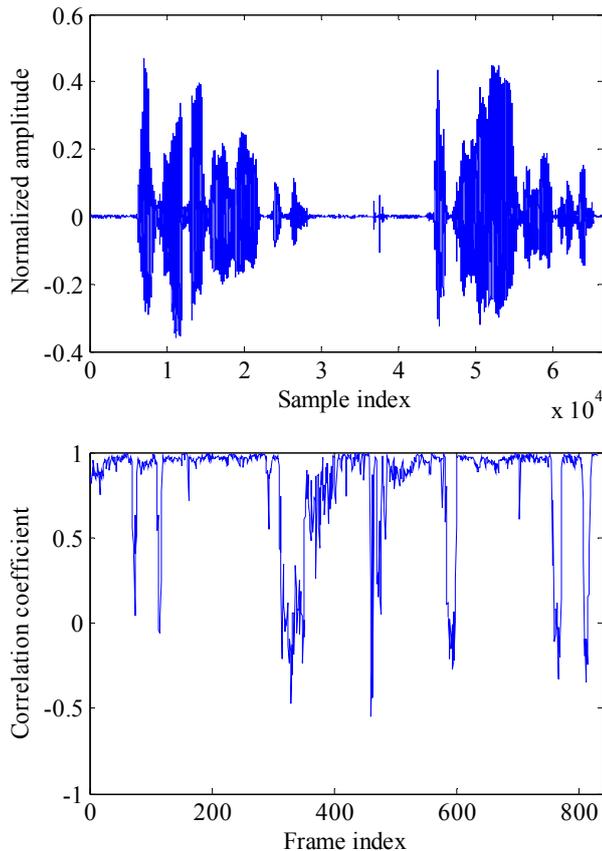


Figure 7. The correlation coefficient for different signal frames

IV. EXPERIMENTAL RESULTS AND DISCUSSION

All experiments were performed using the speech signal that consists of 66500 samples, sampled at 16 kHz. Training sequence of approximately 3 minutes of speech was used to obtain predictor coefficients for configuration with the

switched predictor. The coder was then applied to speech that was not included in the training sequence. Log-uniform quantization with $L=32$ levels was used for variance quantization, whereas 50 dB range of the variances is assumed and the reference variance is fixed at $\sigma_{ref}^2=4 \times 10^{-4}$.

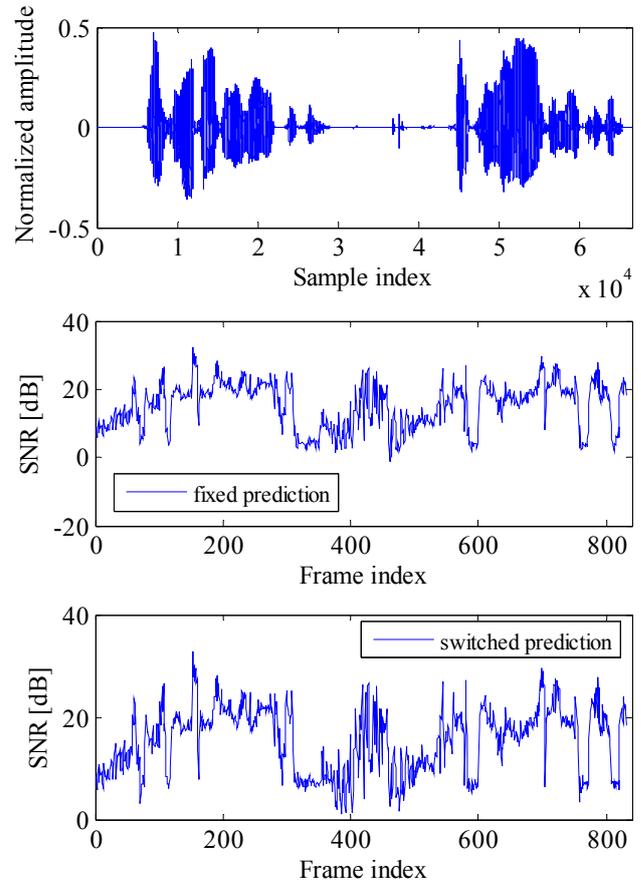


Figure 8. SQNR for different signal frames ($M=80$) for $N=3$ levels quantizer satisfying criterion 1 implemented in the scheme with the fixed predictor and the scheme with the switched predictor (the coefficients are chosen according to the method 1)

Fig. 7 illustrates the estimated correlation coefficient using (15) for each frame of the considered speech. Note that in the area of active speech values of ρ are close to 1, indicating the high predictability of the signal.

In Fig. 8, we present the signal to noise ratio over all signal frames ($M=80$) for configurations with the three-level quantizer designed according to the criterion of minimal distortion (criterion 1) with the fixed and the switched predictor. It can be seen in both that for the active speech higher values of the SNR are obtained (up to 30 dB), while in inactive speech frames SNR decreases up to 0 dB.

In Table I we present the average values of signal to noise ratio along with the bit rate, for various frame lengths (i.e. $M=80, 160, 200, 240$ and 320 samples). SNR_{DM} denotes the average signal to noise ratio obtained using a coding scheme with fixed predictor (see Section 3.A) obtained using (20). SNR_{DM}^1 and SNR_{DM}^2 denote the average values for the configuration with switched predictor (see Section 3.B) obtained according to (23), when the predictor coefficients are optimally selected according to the first and second method for each frame length, as given in Section 3.D. The last column SNR_{DM}^a serves as an upper bound and presents results obtained in an ideal case, when an adaptive first-order predictor was used.

TABLE I. PERFORMANCE OF THE PROPOSED DELTA MODULATION SCHEMES WITH THREE-LEVEL SCALAR QUANTIZER (CRITERION 1), FOR VARIOUS FRAME LENGTH

M	SNR_{DM} [dB]	R^I [bit/sample]	SNR_{DM}^1 [dB]	SNR_{DM}^2 [dB]	R^{II} [bit/sample]	SNR_{DM}^a [dB]
80	15.135	1.429	15.382	15.274	1.441	15.516
160	14.888	1.398	15.152	15.091	1.404	15.290
200	14.881	1.391	15.168	15.114	1.396	15.297
240	14.843	1.387	15.110	15.075	1.391	15.268
320	14.739	1.382	15.015	14.958	1.385	15.142

TABLE II. PERFORMANCE OF THE PROPOSED DELTA MODULATION SCHEMES WITH THREE-LEVEL SCALAR QUANTIZER (CRITERION 2), FOR VARIOUS FRAME LENGTH

M	SNR_{DM} [dB]	R^I [bit/sample]	SNR_{DM}^1 [dB]	SNR_{DM}^2 [dB]	R^{II} [bit/sample]	SNR_{DM}^a [dB]
80	14.652	1.342	14.891	14.747	1.355	15.017
160	14.481	1.311	14.733	14.616	1.317	14.872
200	14.487	1.305	14.756	14.679	1.310	14.886
240	14.468	1.301	14.745	14.667	1.305	14.880
320	14.378	1.296	14.638	14.551	1.299	14.763

The results in a given table indicate that the maximal value in terms of SNR_{DM} in all considered scenarios is obtained for the frame length $M=80$, which is expected as the quantizer parameters are adjusted more often.

Comparing results to a delta modulation system with two-level and four-level Lloyd-Max quantization we see that the proposed scheme with fixed predictor attains almost the same SNR_{DM} value as four-level delta modulation, with reduction in the bit rate of 0.63 bit/sample, whereas outperforming two-level delta modulation for nearly 4.2 dB (see Table III). The reason of such benefit lies in applying the variable length coding.

TABLE III. PERFORMANCE OF A DELTA MODULATION SYSTEM WITH FIXED PREDICTOR USING TWO-LEVEL AND FOUR-LEVEL LLOYD-MAX QUANTIZER

M	$SNR_{N=2}$ [dB]	$R_{N=2}$ [bit/sample]	$SNR_{N=4}$ [dB]	$R_{N=4}$ [bit/sample]
80	10.978	1.062	15.529	2.062

Furthermore, schemes with the switched predictor (see SNR_{DM}^1 and SNR_{DM}^2) offer better performance than the one with fixed predictor (SNR_{DM}), at the cost of slightly increased bit rate. Slightly higher performance is obtained when the switched predictor coefficients are chosen according to the first optimization method (SNR_{DM}^1). All proposed schemes have a performance that is very close to upper bound results, as given by SNR_{DM}^a .

The results for a three-level delta modulation schemes when a three-level quantizer designed according to the simultaneous criterion of minimal distortion and minimal bit rate (criterion 2) is used are summarized in Table II. The table is organized in the same way as Table I, and similar conclusions can be drawn. We obtain slightly lower SNR_{DM} values compared to the results in Table I, but with savings in bit rate. This model is preferred when higher compression is desired.

Moreover, we compare performance of the proposed schemes with the wideband rate distortion bound, which represents the lower bound of performance of several standardized ADPCM speech codecs for wideband speech, including the G.722 standard [22] and its extensions, as shown in [21]. Since distortion is not the best evaluation measure of speech quality, we generate a mapping of distortion-to- $PESQ$ using the mapping function [21].

$$PESQ(D) = ae^{-b \cdot D} + 4.5 - a \quad (24)$$

where a and b are estimated by the least squares fit of the distortion D and $PESQ$ pairs of ADPCM waveform codecs.

The results for the proposed configurations (with three-level quantization satisfying criterion 1), as well as baselines, for $M=80$ are given in Fig. 9, along with the rate distortion bound (solid line). We can observe that delta modulation with the switched predictor provides $PESQ$ scores slightly better than the one of four-level delta modulation baseline, and it outperforms configurations with fixed predictor and two-level Lloyd-Max's quantization, at the same time satisfying the rate distortion lower bound.

The complexity of the proposed algorithm remains unchanged compared to the baselines, it is equal to $O(N^2)$.

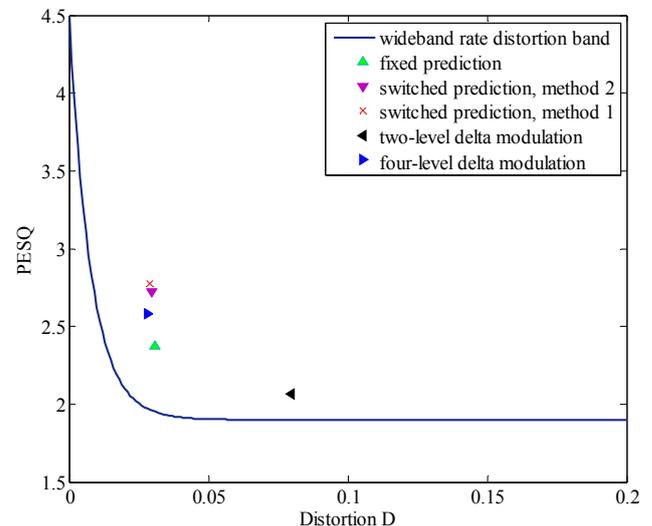


Figure 9. The wideband rate distortion bound, as defined in [21]

V. CONCLUSION

Two configurations based on delta modulation with adaptive three-level scalar quantization are proposed in this paper, one with a simple fixed first-order predictor and one with a switched predictor utilizing correlation. Switched predictor divides frames into weakly and highly correlated and chooses an appropriate, optimal coefficient for selected frame. $PESQ$, SNR and bit rate are used as a measure of performance. Predictor coefficients are optimized frame-

wise, with maximal SNR being used as criterion for optimization. It has been demonstrated in this paper that both proposed solutions with fixed and switched predictor provide over 4 dB higher SNR compared to the traditional adaptive two-level delta modulation, while comparable performance in terms of SNR , with substantial savings in bit rate, was achieved in comparison to four-level delta modulation baseline. Moreover, all proposed configurations have a performance that is very close to upper bound results, obtained using delta modulation with an adaptive three-level quantizer and an adaptive fixed first-order predictor. The obtained performances in terms of $PESQ$ of considered solutions are further compared to rate distortion bound, showing that it provides satisfying perceived quality of speech.

ACKNOWLEDGMENT

This work was supported in part by the by the Ministry of Education and Science of the Republic of Serbia, grant no. TR32051 and TR32035, within the Technological Development Program.

REFERENCES

- [1] J. Jensen, I. Batina, R. C. Hendriks, R. Heusdens, "A study of the distribution of time-domain speech samples and discrete fourier coefficients," in Proc. IEEE First BENELUX/DSP Valley Signal Processing Symposium, 2005, pp. 155-158.
- [2] N. S. Jayant, P. Noll, Digital Coding of Waveforms. New Jersey, Prentice Hall, Chapter 4, pp. 115-188, Chapter 8, pp. 372-417, 1984.
- [3] K. Sayood, Introduction to Data Compression. San Francisco, Elsevier Science, Chapter 9, pp. 227-270, 2005.
- [4] W. C. Chu, Speech Coding Algorithms: Foundation and Evolution of Standardized Coders. John Wiley & Sons, New Jersey, Chapter 5, pp. 143-158, 2003.
- [5] D. G. Zrilic, Circuits and Systems Based on Delta Modulation. Springer, Chapter 1, pp. 1-27, 2005.
- [6] J. D. Gibson, "Speech compression," Information, vol. 7, no. 32, pp. 1-22, 2016. doi:10.3390/info7020032
- [7] Z. Peric, B. Denic, V. Despotovic, "Delta modulation system with a limited error propagation" in Proc. XIII International Conference SAUM, Nis, Serbia, 2016.
- [8] 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
- [9] H. Zheng, Z. Lu, "Research and design of a 2-bit delta modulator encoder/decoder," in Proc. 24th Chinese Control and Decision Conf. (CCDC), Taiyuan, 2012. doi:10.1109/CCDC.2012.6244574
- [10] E. A. Prosalentis, G. S. Tombras, "2-bit adaptive delta modulation system with improved performance," EURASIP Journal on Advances in Signal Processing, Article ID 16286, 2007. doi:10.1155/2007/16286
- [11] M. Lewandowski, "A short-term analysis of a digital sigma-delta modulator with a nonstationary audio signals," in Proc. International AES Convention, Warsaw, 2015.
- [12] T. Ziquan, Y. Shaojun, J. Yueming, D. Naiying, "The design of a multi-bit quantization sigma-delta modulator," International Journal of Signal Processing, Image Processing and Pattern Recognition, vol. 6, no. 5, pp. 265-274, 2013. doi:10.142 57/ ijsip.2013.6.5.24
- [13] C. Canudas De Wit, J. Jaglin, C. Siclet, "Energy-aware 3-level coding and control co-design for sensor network systems," in Proc. IEEE International Conference on Control Applications, Singapore, 2007, pp. 1012 - 1017. doi: 10.1109/CCA.2007.4389366
- [14] M. Azarbad, A. Ebrahimzadeh, "ECG compression using the three-level quantization and wavelet transform," International Journal of Computer Applications, vol. 59, no. 1, pp. 28-38, 2012. doi:10.5120/9515-3916
- [15] Z. A. Sadik, J. P. O'Shea, "Realization of ternary sigma-delta modulated arithmetic processing modules," EURASIP Journal on Advances in Signal Processing, vol. 6, no. 5, pp. 665-676, 2009. doi:10.1155/2009/574627
- [16] M. Dincic, Z. Peric, "Design of quantizers with Huffman coding for Laplacian source," Elecrtonika IR Electrotehnika, vol. 106, no. 10, pp. 129-132, 2010.
- [17] 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.
- [18] A. Ortega, M. Vetterly, "Adaptive scalar quantization without side information," IEEE Trans. on Image Processing, vol. 6, no. 5, pp. 665-676, 1997. doi:10.1109/83.568924
- [19] V. Despotovic, Z. Peric, L. Velimirovic, V. Delic, "DPCM with forward gain-adaptive quantizer and simple switched predictor for high quality speech signals," Advances in Electrical and Computer Engineering, vol. 10, no. 4, pp. 95-98, 2010. doi:10.4316/AECE.2010.04015
- [20] ITU-T, Recommendation P. 862: "Perceptual evaluation of speech quality (PESQ), an objective method for end-to-end speech quality assessment of narrowband telephone networks and speech codecs", 2001.
- [21] J. D. Gibson, Y. Y. Li, "Rate distortion performance for wideband speech," in Proc. Information Theory and Applications Workshop (ITA), San Diego, 2012, pp. 186-191. doi:10.1109/ITA.2012.6181803
- [22] ITU-T, Recommendation G.722.1: "Low-complexity coding at 24 and 32kbit/s for hands-free operation in systems with low frame loss", 2005.
- [23] D. Marco, D. Neuhoff, "Low resolution scalar quantization for Gaussian and Laplacian sources with absolute and squared distortion measures," Technical Report, 2006.
- [24] ITU-T Recommendation P.862.2: "Wideband extension to recommendation P.862 for the assessment of wideband telephone networks and speech codecs", 2007.