Reduced Complexity LSF Vector Quantization with Switched-Adaptive Prediction

Davorka Petrinovic, Davor Petrinovic

Faculty of Electrical Engineering and Computing
University of Zagreb, Croatia
davorka, davor.petrinovic@fer.hr

Abstract

A modification of a classical Predictive Vector Quantization (PVQ) technique with switched-adaptive prediction for line spectrum frequencies (LSF) quantization is proposed in this paper, enabling significant reduction in complexity. Lower complexity is achieved through use of higher number of switched prediction matrices but with reduced number of their non-zero elements. The structures of such matrices and optimal matrix elements are obtained to maximize the quantizer closed-loop prediction gain. A comparison of the proposed quantizer to the ones with full prediction matrices as well as to the quantizer incorporating diagonal matrices is given. The effectiveness of the proposed approach is shown and the trade-off between complexity and quality of the quantizer is analyzed.

1. Introduction

It has been shown by many authors that applying memory-based techniques for speech spectrum quantization results with lower bit-rates compared to the memoryless VQ. This has also been substantiated in a recent survey [1] showing that quantization schemes based on predictive vector quantization (PVQ) approach theoretical limit that can be achieved with interframe coding techniques. The line spectrum frequencies (LSF) representation of the LPC spectrum has been established as the best parameter set for such quantization schemes since LSF vectors of successive speech frames exhibit strong correlation. Adaptation of the predictor to the input LSF process is usually performed by switching between several pre-designed prediction matrices as in [2]. Bit allocation between switched predictor (SP) and VQ determines the trade-off between quantizer complexity and quality. The SP-VQ scheme with only two switched matrices (the one with the highest VQ complexity) has been shown [3] to result with the lowest average spectral distortion and has therefore been applied by a number of authors.

In this work the aspect of computational complexity of the SP-VQ has primarily been investigated. The research is based on the fact that by increasing the number of switched predictor matrices VQ complexity can be considerably reduced. This, however, can sometimes result with significant increase of complexity of prediction that has to be dealt with. Prediction of any input vector is usually performed by multiplication of the vector with all switched matrices (typically of dimension 10×10) and selecting the one that results with the closest prediction. One possible way of simplifying this task is utilizing binary-tree search of prediction matrices instead of the optimal exhaustive search [4]. Another approach is setting to zero most of the elements of switched matrices that has to be supported by efficient computational procedures then performed only on the remaining nonzero elements. For example, full prediction matrices can be approximated with diagonal ones, as mentioned by some authors in [1], [2] and later investigated in [5].

Although a certain loss of quantization quality is inevitable if prediction matrix elements are zeroed, a fixed structure (pattern of zero and nonzero elements) of diagonal prediction matrices may not result with the best quality of reconstructed speech for that number of nonzero elements. A general technique that does not impose any restrictions on the matrix structures was introduced in [6]. Moreover, in this approach matrix element values are calculated to be optimal for the obtained structures. Application of this technique in LSF quantization is presented in this paper with detailed analysis of the achievable results. It will be shown that the obtained sparse prediction matrix quantizers (called 'sparse quantizers' for short) with 32 or 64 switched matrices can have several times lower overall complexity compared to the baseline SP-VQ systems with only 2 matrices. Furthermore, sparse quantizers outperform

quantizers with diagonal prediction matrices, for cases with equal total number of predictor nonzero elements.

2. Sparse switched predictor design

In the classical SP-VQ quantization scheme, design of the switched predictor is an iterative procedure [3]. In each iteration input vectors of the training sequence are encoded and classified into N classes based on prediction matrices used for their prediction. For each of the classes covariance matrices are calculated as in (1) and (2):

$$\mathbf{C}_m = \sum_{n \in \text{class } m} \mathbf{x}(n-1)\,\mathbf{x}^{\mathsf{T}}(n-1) \quad \in \quad \mathbf{R}^{k \times k}, \tag{1}$$

$$\mathbf{\Gamma}_m = \sum_{n \in \text{class } m} \mathbf{x}(n-1)\,\mathbf{x}^{\mathsf{T}}(n) \quad \in \quad \mathbf{R}^{k \times k}, \tag{2}$$

and then used for determination of a new set of prediction matrices $\mathbf{A}_m \in \mathbf{R}^{k \times k}, m = 1, \ldots, N$ for next iteration as in:

$$\mathbf{C}_m \cdot \mathbf{A}_m^{\mathsf{T}} = \mathbf{\Gamma}_m \ . \tag{3}$$

Design of sparse switched predictors is also an iterative procedure resulting with a new set of optimal matrices in each iteration that, in addition, have less and less nonzero elements as iterating progresses. Computational complexity of prediction is proportional to the total number of nonzero elements in all N switched matrices. Therefore each iteration p is characterized by a number called the element reduction factor, η, defined as the ratio between total number of elements in full matrices of dimension k and total number of nonzero elements S in all sparse matrices:

$$\eta(p) = N \cdot k^2 / S(p) \,. \tag{4}$$

One design iteration consists of the two basic steps: 1. iterative predictor refinement for the current sparse structure and 2. determination of the new predictor structure. Suppose that p-1 design iterations have already been performed. The p^{th} iteration starts from N switched matrices determined in the second step of the previous iteration. Matrix elements are optimal for their current structures but were calculated from covariance matrices (vector classification) corresponding to the previous iteration. Therefore, keeping the structures fixed, a set of prediction matrices is iteratively refined in the first step of the p^{th} iteration. This is performed analogously to the classical design procedure by repetitive encoding (reclassification) and predictor calculation until the change in prediction gain is small enough. Due to the sparse structure, prediction matrix elements are calculated differently than in (3), by way of modified algorithm [6].

At this point the resulting matrices are saved since they represent optimal solutions for given structures and factor $\eta(p)$. Second design step is then performed in order to obtain switched matrices of higher element reduction. New structures are realized by setting to zero an arbitrary number of nonzero elements selected among all elements of all switched matrices by applying the reduction criterion that ensures the highest prediction gain. This criterion together with the modified algorithm for calculation of the optimal element values of sparse matrices is briefly described in the following section. After element reduction and recalculation, the current iteration is finished.

The above design procedure is repeated until all sets of sparse matrices of desired element reduction factors η are obtained. Since there is no restriction on the choice of the removed elements, the method results with sparse matrices having different number of nonzero elements located on different positions in each of the matrices. This way the matrix structures and element values are inherently adjusted to model only the strongest correlations of the input LSF vector process.

An example of a sparse predictor with N = 8 matrices having in total 1/10-th of elements of the full predictor is shown in Figure 1. The black fields represent nonzero elements while the white fields represent zeros.

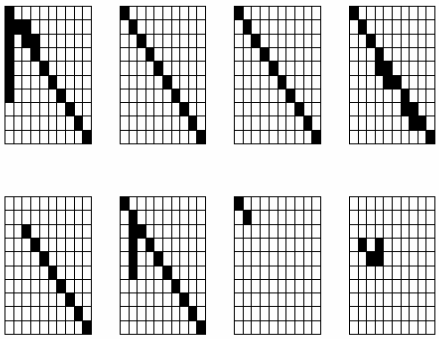

Figure 1. Sparse prediction matrices with k=10, N=8 and η = 10

2.1. Sparse predictor computation

An algorithm has been developed [6] for calculating optimal sparse predictors that also incorporates the optimal criterion for element reduction. Sparse matrices are always calculated row by row since the structure of each of them can be different. Each row of any sparse prediction matrix can be treated as a scalar predictor of the order equal to the number of nonzero elements in that particular row. If an additional nonzero element of the i^{th}

row is set to zero, predictor order is reduced by one, resulting with monotonic increase of the prediction residual energy. This increase can be reduced if the remaining elements of that row are recomputed to be optimal for the new predictor order. Therefore the elements are set to zero in a way that the increase of the residual energy after row re-computation is the least.

Let $\boldsymbol{\alpha}_i^{(r(i))}$ denote the optimal i^{th} row of any sparse switched matrix (matrix index m is omitted for simplicity). Number $r(i)$ denotes the number of elements already set to zero in that particular row. According to the algorithm, $\boldsymbol{\alpha}_i^{(r(i))}$ is calculated as:

$$(\boldsymbol{\alpha}_i^{(r(i))})^{\mathsf{T}} = \boldsymbol{\gamma}_i^{\mathsf{T}}\, \mathbf{B}_i^{(r(i))}, \qquad (5)$$

where vector $\boldsymbol{\gamma}_i$ is the i^{th} column of the covariance matrix $\boldsymbol{\Gamma}_m$. Matrix $\mathbf{B}_i^{(r(i))}$ is an auxiliary matrix associated with $\boldsymbol{\alpha}_i^{(r(i))}$, having $r(i)$ rows and $r(i)$ columns equal to zero vectors. The indexes of zero rows and columns correspond to the positions of zeroes in $\boldsymbol{\alpha}_i^{(r(i))}$. Initially, for the row with all nonzero elements ($r(i) = 0$), matrix $\mathbf{B}_i^{(0)}$ is equal to $\mathbf{C}_m^{-1}$. Criterion for reduction of one additional element of the i^{th} row is given by:

$$D_{i,j}^{(r(i)+1)} = \frac{1}{b_{(i)j,j}^{(r(i))}} (a_{i,j}^{(r(i))})^2 \,, j \in \{(1,2,\ldots,k) \,\big|\, a_{i,j}^{(r(i))} \neq 0\}, \qquad (6)$$

i.e. as the ratio of the square of every nonzero element of that row $a_{i,j}^{(r(i))}$ and corresponding diagonal element $b_{(i)j,j}^{(r(i))}$ of the matrix $\mathbf{B}_i^{(r(i))}$. If this criterion is applied to all rows $i = 1,\ldots,k$ independently, the element $a_{I,J}^{(r(I))}$ that results with the smallest ratio across the whole matrix is found and then set to zero. Only the matrix $\mathbf{B}_I^{(r(I))}$ corresponding to the selected row has to be recalculated. The recursive expression for calculation is given by:

$$\mathbf{B}_i^{(r(i)+1)} = \mathbf{B}_i^{(r(i))} - \frac{1}{b_{(i)j,j}^{(r(i))}} \boldsymbol{\beta}_{(i)j}^{(r(i))} \cdot (\boldsymbol{\beta}_{(i)j}^{(r(i))})^{\mathsf{T}}, \qquad (7)$$

where $\boldsymbol{\beta}_{(i)j}^{(r(i))}$ represents the j^{th} column of $\mathbf{B}_i^{(r(i))}$.

The above operation performed by substituting $i = I$ and $j = J$ will zero the J^{th} row and column of $\mathbf{B}_I^{(r(I)+1)}$. New optimal row $\boldsymbol{\alpha}_I^{(r(I)+1)}$ is then obtained as in (5) thus increasing the prediction residual energy by $D_{I,J}^{(r(I)+1)}$. In the switched predictor scheme, the element search is performed over all N matrices and the element resulting with the least increase is selected as the one for reduction.

3. Simulation results

Extensive simulations were carried out with a number of different LSF quantizers. Speech database used for quantizer design contained 20 minutes of speech (119960 LSF vectors), spoken by 2 male and 2 female speakers. Quantizers were evaluated on a different speech material obtained from 7 male and 3 female speakers (59980 vectors). Both databases were created from speech data sampled at 8kHz and analyzed using 10^{th} (k=10) order LPC analysis performed on 25ms speech segments at two different LPC frame rates: FR = 100 and 50 frames/s.

Designed quantizers combined switched prediction and full-searched split vector quantization (split 4-6) of the prediction residual. The total number of bits per frame, B, was varied from 19 to 23 bits for FR = 50 fr./s and from 15 to 19 for FR = 100 fr./s. The number of bits specifying the switched prediction matrix was $b = 1, 4, 5$ and 6 (number of matrices $N = 2^b$). For the given B, the bit allocations (for predictor b, and the first and the second sub-vector codebook, $b1$-$b2$) producing the best results are given in Table 1.

Table 1. Bit allocations for quantizers

FR=50 fr./s					
$b \setminus B$	19	20	21	22	23
4	6- 9	7- 9	7-10	8-10	8-11
5	6- 8	6- 9	7- 9	7-10	8-10
6	5- 8	6- 8	6- 9	7- 9	7-10
FR=100 fr./s					
$b \setminus B$	15	16	17	18	19
4	5- 6	5- 7	5- 8	6- 8	6- 9
5	4- 6	5- 6	5- 7	5- 8	6- 8
6	4- 5	4- 6	5- 6	5- 7	6- 7

Initial predictors and vector quantizers obtained in the open-loop design procedure were improved in the closed-loop as in [3]. Besides the proposed sparse quantizers with different element reduction factors η, a group of quantizers with normal (full) prediction matrices was also designed for comparison (called 'full quantizers').

Quality of the proposed quantization scheme was evaluated by the average log spectral distortion, $\overline{SD}$, determined on the evaluation database as the RMS error between the original and the quantized LPC log power spectra. Percentage of the outlier frames with $\overline{SD}$ greater then 2 dB was also calculated. Another important performance measure was the overall quantizer complexity, denoted with K. It was calculated as the total number of arithmetic operations (multiplications, additions and comparisons) for both prediction and VQ that are required for quantization of a single input vector. Since this work was aimed towards reducing complexity

while maximizing the quantizer quality, most of the results are presented as complexity (K) vs. distortion ($\overline{SD}$) curves.

For illustration of the influence of element reduction factor η, sparse quantizers with 10 different values of η (from 1 to 10 in steps of 1) were designed for $B = 21$ ($FR = 50$) and for $B = 17$ bits/frame ($FR = 100$). The results are shown in Figures 2 and 3 where each point on the curves represents one quantizer obtained with different values of η and b. As expected, higher η (less nonzero elements) results with lower complexity at the cost of somewhat increased $\overline{SD}$. Obviously the highest reduction of complexity occurs for quantizers with the highest number of matrices ($b = 6$) but for these $\overline{SD}$ is also increased the most. It is interesting to observe that K falls abruptly as η is increased from 1 to 6, but only minor complexity reductions can be gained afterwards.

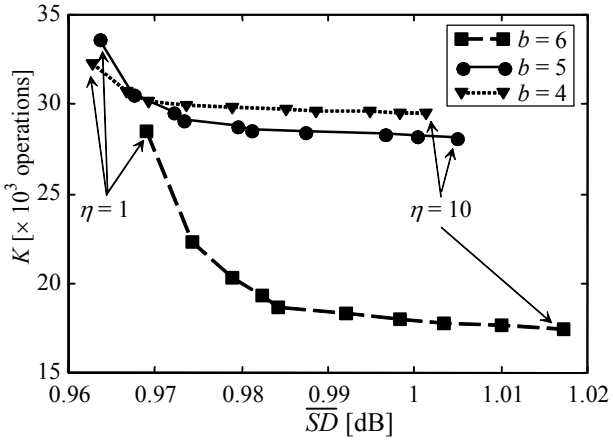

Figure 2. Number of operations vs. $\overline{SD}$ for sparse quantizers with B = 21 and different η (FR = 50 fr./s)

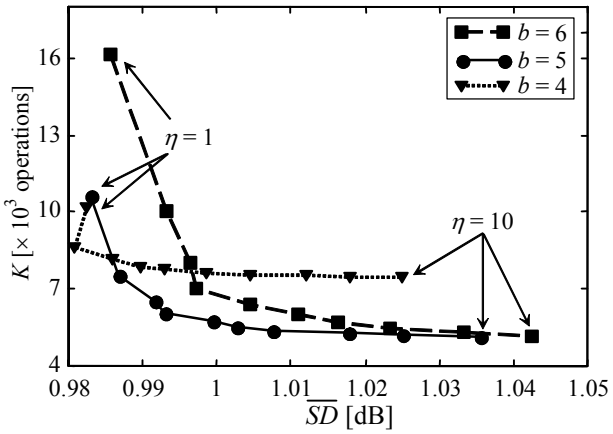

Figure 3. Number of operations vs. $\overline{SD}$ for sparse quantizers with B = 17 and different η (FR = 100 fr/s)

Quantizers for all other values of B were designed with three different element reduction factors η equal to 3, 6 and 10. Their performance obtained with the highest tested number of switched matrices ($b = 6$) is shown in Figure 4 for $FR = 50$ and Figure 5 for $FR = 100$. Besides these sparse quantizers, full quantizers with $b = 1$ and $b = 6$ are also shown for reference since the former result with the highest quality while the latter group demonstrates the influence of the increased number of switched prediction matrices. In the figures, quantizers of different types are represented with different symbols while the lines connect quantizers having the same total number of bits per frame.

First, the difference between full quantizers designed for two different frame rates is discussed. Namely, increasing the number of matrices from 2 ($b = 1$) to 64 ($b = 6$) causes great reduction of number of operations at $FR = 50$ fr./s. On the other hand, for higher frame rate VQ codebooks are smaller so the number of operations for prediction (with b=5 or 6) prevails over the one for VQ. As a consequence, full quantizers with b=6 and B<18 require even more operations compared to those with b=1 (Fig. 5).

As can be seen on Figures 4 and 5 for quantizers with (η>1), the proposed technique with sparse matrices solves the problem at higher frame-rate and also enables additional reduction of complexity for quantizers at $FR = 50$ fr./s. Complexity reduction factor κ is introduced as ratio of total number of operations for quantization (K) between any full quantizer with 2 switched prediction matrices and the sparse quantizer designed for the same B. The highest factors κ are obtained for sparse quantizers with element reduction factor η=10 and predictors with $b = 5$ or 6, depending on the value of B. Factors are in the range 3.0-6.1 for $FR = 50$, and 1.5-4.0 for $FR = 100$.

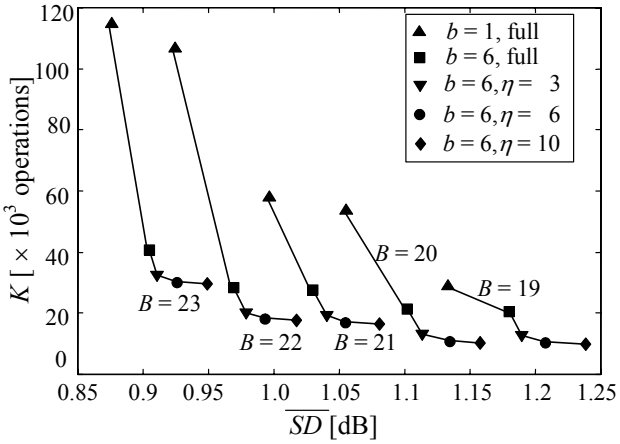

Figure 4. Number of operations vs. $\overline{SD}$ for different sparse and full quantizers (FR = 50 fr./s)

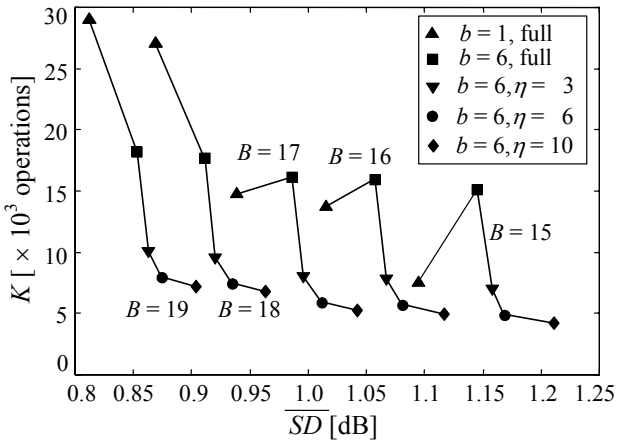

Figure 5. Number of operations vs. $\overline{SD}$ for different sparse and full quantizers (*FR* = 100 fr./s)

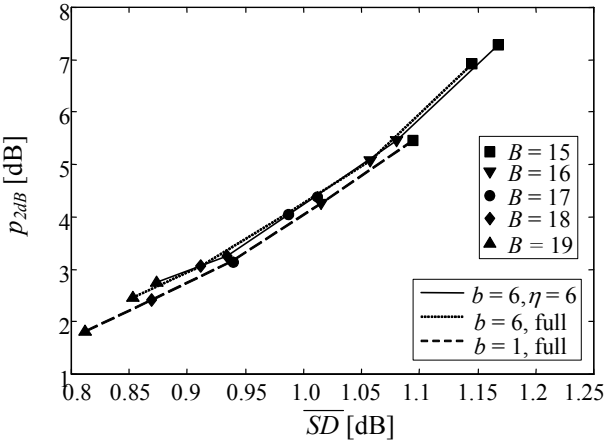

Figure 6. Percentage of outliers vs. $\overline{SD}$ for different quantizers (*FR* = 100 fr./s)

Sparse predictor structure as well as the increased number of switched matrices result with certain loss in quality of sparse quantizers compared to the full quantizers with $b = 1$ (both having the same number of bits per frame, B). This loss can be expressed as the required increase of bit-rate, ΔB, of sparse quantizers for achieving the quality ($\overline{SD}$) of baseline quantizers. These values are almost independent of B so they were averaged across all used bit-rates. The resulting values of $\overline{\Delta B}$ for different sparse quantizers are given in Table 2. It is obvious that the numbers are very similar at both frame-rates and increase with both b and η. Comparing the values in Table 2 with results on Figures 4 i 5, it is obvious that the best complexity-performance trade-off can be achieved with sparse quantizers with $\eta = 6$.

Table 2. The average increase of bit-rate $\overline{\Delta B}$ for different sparse quantizers compared to the full quantizers with $b = 1$

η	b	$\overline{\Delta B}$ [bit]	
		FR = 50	*FR* = 100
3	5	0.72	0.76
	6	0.74	0.76
6	5	0.89	0.91
	6	0.97	0.93
10	5	1.23	1.33
	6	1.34	1.38

As can be seen in the example on Figure 6, the percentage of outliers, $p_{2\text{dB}}$, is not increased by the sparse matrix technique itself. It is somewhat higher for quantizers with higher b compared to quantizers with $b = 1$ since the same VQ codebooks are used for all residual vectors (regardless of the class).

To emphasize the advantage of the proposed technique, sparse quantizers with $\eta = 10$ are compared to quantizers designed with diagonal prediction matrices (called 'diagonal quantizers' for short). These diagonal quantizers are chosen for comparison since they represent the simplest and most straightforward approach for reducing the complexity of prediction. Both the sparse and the diagonal type of predictors have the same total number of nonzero matrix elements and therefore result with the same complexity. A group of such quantizers designed with $b = 6$ and at $FR = 50$ is shown on Figure 7 together with full quantizers with $b = 1$ for reference. It is obvious that the resulting spectral distortion of the optimal sparse quantizers is much lower then it is the case of quantizers with diagonal prediction matrices. This difference is around 1 bit if expressed in ΔB.

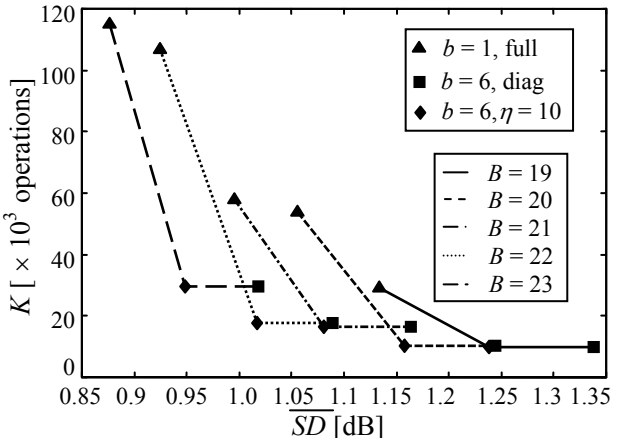

Figure 7. Number of operations vs. $\overline{SD}$ for quantizers with sparse, diagonal and full predictors (*FR* = 50 fr/s)

In this paper, it has been demonstrated based on various simulation results that the proposed technique offers a good solution if complexity of the SP-VQ quantization scheme is a matter of concern. The actual implementation was outside the scope of this paper. Nevertheless, this technique has been implemented in real hardware and the obtained results substantiate the effectiveness of the approach.

4. Conclusion

Although reduction of the LSF quantizer complexity can be achieved in various ways, this paper demonstrates how this can be accomplished by modifying prediction in the SP-VQ scheme by using large number of sparse switched prediction matrices. It was shown that proposed quantizers with optimal sparse matrix structures result with lower spectral distortion (or require 1 bit less) then those designed with diagonal predictors. By variation of several design parameters, it is possible to realize the quantizer that offers the best compromise between complexity and quantization quality for the actual application. Although this technique was tested on the LSF process, it may be applied to any vector process with similar correlation properties.

5. References

[1] T. Eriksson, J. Linden, J. Skoglund, "Interframe LSF quantization for noisy channels", *IEEE Transaction on Speech And Audio Processing*, vol. 7, no. 5, pp. 495-509, Sept. 1999.

[2] M. Yong, G. Davidson, A. Gersho, "Encoding of LPC spectral parameters using switched-adaptive interframe vector prediction", *Proc. ICASSP*, vol.1, pp. 402-405, 1988.

[3] S. Wang, E. Paksoy, A. Gersho, "Product code vector quantization of LPC parameters", in *Speech and Audio Coding for Wireless and Network Applications*, editors: B. S. Atal, V. Cuperman, A. Gersho, Kluwer Acad. Pub., pp. 251-258, 1993.

[4] D. Petrinović, D. Petrinović, "Switched-adaptive interframe vector prediction with binary-tree searched predictors", *Proc. of 2000 IEEE Workshop on Sig. Proc. Sys. Design & Implementation* (*SIPS*), Lafayette, SAD, pp. 733-742, 2000.

[5] D. Petrinović, D. Petrinović, "Sparse vector linear predictor matrices with multidiagonal structure ", *Proc. of EUROSPEECH '99*, Budapest, Hungary, pp. 733-742, 2000.

[6] D. Petrinović, D. Petrinović, "Sparse vector linear prediction with near-optimal matrix structures", *Proc. of International. Workshop on Image and Signal Processing and Analysis*, Pula, Croatia, pp. 235-240, 2000.