Eng. Research Bull.
Faculty of Eng. & Tech.,
Menoufia University.
Vol. VII Part II, 1985.

PP. 209-223

# FURTHER BIT RATE REDUCTION USING HYBRID CODING IN TRANSFORM DOMAIN.

IBRAHIM ZAKARIA MORSI. Ph.D.,
Faculty of Electronic Engineering, Menouf, EGYPT.

## ABSTRACT:

Different linear orthogonal transformations are now used to remove time-domain redundancy in digital singal processing. Coding and transmitting the decorrelated transform domain coefficients, rather than the usually highly correlated time domain signal samples, lead to fewer bits being used than in the case of original PCM transmission, and hence, the bit rate requirement is reduced. In a previous study on Arabic speech transform coding, good results were obtained concerning bit rate reduction, although no attempt has been made to decrease the quantizing distortion by any means.

In the work reported here, conventional methods of improving the quantizers performance, as well as bitrate reduction were combined with the transformation methods. These conventional methods, such as nonlinear coding, and DPCM are applied here to all, or some of, the coefficients in the transform domain. As expected, both contributions resulted in improved performance for the same bit rate requirements, and/or extra saving in channel bit rate for comparable over-all gradings.

## I. INTRODUCTION:

Transform coding is used mainly for bit rate reduction in digital communications, both in image and speech processing. The principles of its use in the rate reduction, stem from the fact that the transform coefficients are normally less correlated than the original time domain data. The exploitation of time domain correlation in channel bit rate reduction has been applied for a relatively long time, in a family of predictive coding schemes. Examples of such coding schemes are Delta modulation, Differential Pulse Code Modulation, (DPCM), and Adaptive DPCM. These techniques have long been successfully used, so that one may question the advantages of transform coding over those presented by predictive coding.

Although the DPCM system, for instance, is rather simple as far as the equipment configuration is concerned, the transmission errors may sharply degrade its performance. On the other hand, the equipment configuration of a transform coding scheme is more complicated even for the simplest orthogonal transform, like Hadamard Transform [1-3]. However, the degradation resulting from channel errors in transform coding schemes is far much less objectionable than the predictive systems. This applies equally well for both subjective and objective assessments.

With the overgrowing advances in micro-electronics and integrated circuit technology, the real-time realization of such orthogonal transforms is feasible, specially for low frequencies, like in the case of speech signals. Thus, the complexity of transform coding schemes should no longer present any obstacle preventing the full exploitation of their outsanding merits.

In this study, the advantages of both predictive and orthogonal transform coding schemes for Arabic speech processing are combined in one scheme. Due to the fact that this combination takes place in the transform domain, this scheme is called 'Hybrid Coding in Transform Domain'.

## II. EXPERIMENTAL SET UP, MATERIAL, AND ASSESSMENT.

Signals used in this practical experimental work were that of normal spoken Arabic speech, as used in some previous studies [4,5]. Relatively long extracts, from talks by professional broadcast programme presenters, lasting for 1-2 minutes were recorded and processed as explained later.

Microprocessor-based real-time Hadamard transforms were used for vector sizes ranging from 4-16 samples, while for larger sizes, a microcomputer simulated transform was used[6].

As far as the processing is concerned, the direct adaptive coding schemes in transform domain were realized in real -time circuitry. For this processing, it was feasible to apply subjective assessment, (panel assessment), for the resulting output. Hybrid Coding in transform domain was assessed abjectively, subject to the well-known criterion of Minimum Square Error (MSE).

Subjective assessments were based on over-all grading of the reconstructed signals, listened to by a panel of Arabic speaking subjects. The scale of over-all grading was that of the 5-grade system used by other researchers[7]. This scale was rated as: Excellent-Good-Fair-Poor-Unsatisfactory. A numerical score was assigned to each grade as: 8 - 6 - 4 - 2 - 0, respectively. In this way, it was possible to get a more-familiar percent assessment, as will be shown. Detailed information concerning material, subjects, and assessment may be found in previous works [4,5,7].

## III. ENERGY DISTRIBUTION IN TRANSFORM DOMAIN.

Adaptive coding of signals in the transform domain, is based on processing each transform coefficient in a manner which may differ than other coefficients. An obvious criterion to assess a particular coefficient is, of course, its role in the signal reconstruction at the receiving end. This means, that for an adaptive coding scheme, the energy distribution in the transform domain should be carefully studied, as a prerequisite.

In a previous study [5], it has been noticed that speech energy tends normally to concentrate at lower sequency coefficients, in the transform domain. In this study, the characteristics of transform coefficients are studied further, in view of application of new technique.

Figure (1) shows the relative energy, that is a variance, of the coefficient, normalized by the variance of the original samples, for different sizes of transform. The graphs concerning vector sizes of 32-128 samples are the results of simulated transforms on real data, while those for vector sizes of 4 - 16 samples represent the real-time transformations for real-time speech signals.

From these graphs, it seems that, in general, energy content associated with each transform coefficient decreases sharply with increasing sequency. Moreover, this decrease in energy content with increasing sequency is most noticeable as the vector size increases. Thus, a larger vector size will remove more time - domain redundancy, leading to the possibility of coding only fewer number of coefficients, therefore, more saving in the over-all bit-rate may be achieved.

In addition to the apparent feature of decreased energy content with increasing sequency, the graphs also show the 'abruptancy' noted in a previous study and analysis [5,8]. This results from the clustering of coefficients in subgroups (subvectors), divided in binary boundaries [8]. This specific feature of abruptancy makes it sensible to think in treating the coefficients which belong to the same subvector together in the same way, concerning the number of allocated coding bits, and the coding step sizes.

## IV. ENERGY PACKING EFFICIENCY OF TRANSFORMS.

As the main advantage of transformation is the transmission of a limited number of coefficients, rather than the whole sequence of time domain samples, a critical and important criterion used in conjunction with the transform coding is the 'Energy Packing Efficiency'. An 'Energy Compaction Measure' for transform image processing was suggested by Pratt[9]. In that study, this measure was set as the percentage of energy content in largest quarter of transform coefficients, as compared to the total energy content in the whole vector. The value of a quarter was rather an arbitrary value, and suggested what energy compaction efficiency was met if a maximum bit-rate reduction of a factor of 4 were attempted. Results presented there, with the suggested study were for data modelled as a Markov process, and not for actual data.

In this work, the energy packing efficiency was calculated for real speech signals. Instead of calculating this measure as the energy percent in the largest quarter of coefficients, regardless their actual positions in the vector, it is rather computed for different proportions of coefficients, in a manner of increasing sequency. Figure (2) shows the energy packing efficiency for different vector sizes. Once again, it is clear from this figure, that the efficiency increases with the vector size, for the same proportion of coefficients. The effect of abruptancy is also clear from the figure.

## V. ADAPTIVE CODING IN TRANSFORM DOMAIN.

One obvious way of putting more emphasis on the coefficients which have more energy, is to code them more finely than the coefficients having less energy. Therefore, in a simple adaptive coding scheme, one would think of assigning number of bits to any transmitted coefficient, in proportionality to a messure of its energy content. The variance of each coefficient is taken to act as the required energy measure of this coefficient.

As a matter of practicality, the limited total number of bits allowed for coding, will make this process of assigning only approximate. In the tests involving this study an over -all bit-rate average was initially preset, and bits were adaptively allocated to low sequency coefficients subject, nearly, to their variances. To permit fine coding for these coefficients, some high sequency coefficients were sometimes discarded. Both uniform and non-uniform coding schemes were considered. For non-uniform coding, Max-type based [10], coders were applied.

Table (1) and Figure (3) show the results of the subjective assessment for these adaptive coding schemes at an arbitrary average bit-rate of 2 bits per sample (i.e. 16 kbits/second, for a sampling frequency of 8 KHz). Recalling the results of some coding schemes considered in a previous study [5], curves showing characteristics of non-adaptive coding schemes are included, for the sake of comparison.

Figure (3) reveals that, an average increase in performance over the non-adaptive scheme is quite apparent, specially in the case of non-uniform coding based on the coefficient variance. The average increase in numerical grading assessment ranges from about 45 % to about 115 %, taking into consideration, the whole range of transform sizes (4-128). It may be impressive to point out here, that these results were obtained at an average over-all bit-rate of 16 kbits/second (2 bits/sample).

## VI. CORRELATION IN THE TRANSFORM DOMAIN.

To assess the redundancy removal property of any transform, the correlation of transform coefficients should be computed. The lower the value of this correlation, the more the redundancy removed due to the application of this transform. Figure (4) shows the correlation factor for one dimensional Hadamard transform of different sizes, as applied to Arabic speech signals. Although the ideal purpose of the transformation is to result in decorrelated coefficients, in the transform domain, a degree of correlation is still noticeable. From the fundamental discussions on the Hadamard transformation based on Hadamard matrix [3,8,11,12], it could be said generally, that this particular transform is 'differential'. This is true in a sense, because all the coefficients (except the first one) are the direct results of subtracting half the time domain samples from the other half, with variations in the time intervals between samples and in the original arrangements, resulting the unique sequency of each coefficient. Hence, correlation in one-dimensional Hadamard transform domain is relatively low.

As mentioned before, the energy compaction increases with tranform size. The increase in transform size results hardware complexity of its own, which makes it practical to limit the vector size within realistic values (8-16). However, the advantages of larger transform sizes could still be exploited, using 'temporal' direction correlation.

Although speech signal is a single-dimensional stream of data, a quasitemporal dimension can be considered through the sequencing of different transform vectors. This is approximately similar to the two-dimensional block transformation applied to image signal processing [8]. To apply this technique to speech data, consecutive vectors are considered in sequence, as illustrated in Figure (5). Assuming a reasonable value (N) to represent the vector size, then (K,N) time-domain samples are grouped in (K) vectors. Each vector is then transformed separately, to result (K) transformed vectors, each of (N) coefficients. Redundancy associated with corresponding coefficients in consecutive vectors is assessed by computing the correlation function for each coefficient, along different numbers of vectors.

Figure (6) shows correlation, in 'temporal' dimension, between corresponding transform coefficients, for different depths, k=2,.....,K. In general, and as shown in the figure, correlation in this quasi-temporal dimension is much higher than in the single dimension (horizontal direction). Such high value of correlation, makes it obvious to try to apply DPCM technique in the transform domain, to further reduce the channel over-all bit-rate [13]. The figure shows that the correlation is higher for the low sequency coefficients than the higher sequency. Thus, DPCM is better applied to these low frequencies, in order to achieve reasonable degree of redundancy removal, and keeping the equipment configuration at reasonable degree of complexity.

Improvement in DPCM performance over the linear PCM coding depends, strongly on the correlation. It has been shown [14], that such an improvement in Signal-to-Noise

(S/N) ratio is related to the sample-to-sample correlation $R_{(T)}$, as:

$$\text{S/N Improvement} = 10 \log \left[ \frac{1}{1-R_{(T)}} \right] \quad \ldots \quad \ldots \quad (1)$$

This relation was calculated, subject to an assumption that the distribution of the error signal can be approximated by the Laplace density function:

$$P_{(e)} = \frac{1}{\sqrt{2}\,\sigma_e} \exp \left( - \frac{\sqrt{2}}{\sigma_e} |e| \right) \quad \ldots \quad \ldots \quad (2)$$

Figure (7) represents the plot of Equation (2), which shows substantial improvement when the correlation coefficient is near its maximum value of unity.

## VII. DPCM IN THE TRANSFORM DOMAIN. (HYBRID CODING).

As mentioned before, there is a high degree of correlation between corresponding coefficients in successive transform vectors. This correlation is more noticeable in the lower sequency group of coefficients. DPCM was applied to these low sequency coefficients. Instead of coding a transform coefficient, only the difference between the coefficient and the value of the corresponding coefficient in the preceeding vector is coded. Of course, there will be a large number of different combinations for which one coefficient is selected to be coded in the DPCM scheme. As this new scheme combines the usual transform coding with predictive coding schemes, it is called Hybrid Coding.

Figure (8) shows a block diagram of the equipment configuration. $H_i$ is the Hadamard transform coefficient which is selected to be DPCM coded. Several sequencies (values of i), were considered, one at a time, in computer simulated experiments, and the results were as shown in Table (2).

Figure (9) shows the plottings of results in Table (2). From this figure, the objective assessment criterion of Signal-to-MSE was used. As a result of these experiments, it is clear that coefficients in the same binary divided (bound) subvector have, on average, the same effect on the over-all performance of the coder. To relate these objective assessments with the subjective assessments, derived for non-adaptive and direct adaptive schemes, the objective SNR criteria for linear PCM coding are marked on Figure (9). These markers are based on the well-known standard analysis concerning the calculation of rms Signal-to-quantizing noise ratio in linear quantization PCM. Such analysis [15], computes this ratio as:

$$\text{rms signal-to-quant. noise} = 1.225 \ (M-1) \quad \ldots \quad \ldots \quad (3)$$

where $M$ = number of quantizing levels.
This, again leads to the short from equation:

$$SNR = 1.76 + 6.n \quad dB \quad \ldots \quad \ldots \quad \ldots \quad (4)$$

where $n = \log_2 M$, is the number of coding bits.

Figure (9) shows that for hybrid coding of the whole low sequency subvector, the output at 2 bits/sample is comparable to 6-bit PCM linear quantization, which was normally graded as Excellent [5].

Applying DPCM to more than one coefficient is, of course, a promising means of improving the characteristics of the system, either in reducing the bit-rate, or increasing the fidelity. However, such an application would require an N-deep delay unit for each coefficient coding (where N is the transform size). This may increase the complexity of the hardware of the circuit configuration to a prohibitive limit.

For quick comparison, Figure (9) shows also the result of applying the DPCM coding to the whole ($H_2$-$H_4$) subgroup. The figure shows a very good response at a vector size of 4, since in this case, all the transform coefficients will be DPCM coded. For higher vector sizes, the SNR decreases, but not in the same manner as in the case of single DPCM coded coefficient. It is clear that the over-all performance of more than one DPCM coded coefficient is much better than a single one. A total improvement ranging from 3 dB over the single coding of $H_2$, to 11 dB over the single coding of any coefficient in the $(H_5-H_8)$ subvector, is apparent.

## VIII. CONCLUSIONS.

Adaptive coding of transform coefficients of Arabic speech signals was extensively studied in this work. The detailed characteristics of transform domain coefficients were analyzed in order to exploit their major features. The first adaptive coding scheme was that based on coding coefficients with a degree of accuracy related to their energy contents. This scheme is a direct adaptive one.

The second scheme is more complicate, but is much more efficient. It exploits the remaining redundancy in the transform domain, and hence, tries to idealize the transform. Correlation among respective transform coefficients in consecutive vectors is allowed to initiate a DPCM application to the more correlated coefficients. Although the combination of conventional predictive coding schemes with the more recent transform coding scheme may complicate the equipment configuration, the substantial improvement in performance justified such a complexity.

## IX. REFERENCES.

1. W.K. Pratt, J. Kane, and H,C. Andrews, "Hadamard Transform Image Coding", Proc. IEEE, Vol. 57, PP. 58-68, Jan. 1969.
2. I.Z. Morsi, "Design of a real-time Hadamard transformer for bit-rate reduction in digital signal coding", Accepted for the Int. 84 AMSE Conf., "Modelling and Simulation", Athens, 27-29 June 1984.
3. R, Walker, "Hadamard transformation: A real-time transformer for broad cast standard PCM television", BBC Research Report 1974/7 , 1974.

(4) I.Z. Morsi, "Subjective assessment of bit-rate requirements for digital processing of Arabic speech", Proc. IASTED Int. Conf. TELECON'84, Halkidiki, Greece, Aug. 27-30, 1984, PP. 318-321.

(5) I.Z. Morsi, "Walsh-Hadamard transform coding of Arabic speech signals for bit-rate reduction", Proc. IASTED Int. Conf. TELECON'84, Halkidiki, Greece, Aug. 27-30, 1984, PP. 449.

6) I.Z. Morsi, "Micro-computer simulation of Walsh-Hadamard transformation for some special low frequency signals", Proc. Ninth Int. Cong. for Statistics, Comp. Science, Soc. & Domog. Research, Ain Shams Univ., Mar. 31-Apr. 10, 1984, Vol. 6, PP. 325-334.

(7) D.J. Goodman, and J.S. Goodman, "Intelligibility and ratings of digitally coded speech", IEEE Trans. Acous. Speech, Signal Proc., Vol. ASSP-26, No. 5, Oct'78, PP 403-409.

(8) I.Z. Morsi, "Hadamard transform coding of television signals", Ph.D. Thesis, Bradford University, UK, 1980.

(9) W.K. Pratt, "Spatial transform coding of color images", IEEE Trans. Commun. Tech., Vol. COM-19, PP. 980-992, Dec. 1971.

(10) J. Max, "Quantization for minimum distortion", IRE Trans. Inf. Theory, Vol. IT-6, PP. 7-12, March 1960.

(11) R.E.A. Paley, "On orthogonal matrices", J.Math. Phys., Vol. 12, PP. 311-320, 1933.

(12) H.Y.L. Mar, and C.L. Chemg, "Fast Hadamard transform using the H Digram", IEEE Trans. Computer, Vol. C-22, PP.957-959, Oct. 1973.

(13) M. Ishii, "Picture Bandwidth compression by DPCM in the Hadamard transform domain". FUJITSU Sc. & Tech. J., PP. 51-65, Sep. 1974.

(14) J.O'Neal, "Prdictive quantizing systems (DPCM) for the transmission of TV signals", Bell Sys. Tech.J., 45, No.5, May-June 1966, PP. 689-721.

(15) J.A. Betts, "Signal processing, modulation, and noise", Unibooks, Londom 1975.

Table (1). Subjective Assessment of Adaptive Coding schemes in the transform domain.

| Vector size | Zonal % | Threshold & | Uniform coders | | Non-Uniform coders | |
|---|---|---|---|---|---|---|
| | | | % | Grade | & | Grade |
| 4 | 48 | 54 | 64 | Fair-Good | 78 | Good |
| 8 | 42 | 47 | 60 | Fair | 75 | Good |
| 16 | 37 | 42 | 54 | Fair | 69 | Fair-Good |
| 32 | 29 | 33 | 48 | Poor-Fair | 64 | Fair-Good |
| 64 | 23 | 26 | 45 | Poor-Fair | 60 | Fair-Good |
| 128 | 19 | 22 | 42 | Poor-Fair | 57 | Fair |

Table (2). Objective Assessment for DPCM in the transform domain of some coefficients. (S/N ratio, in dB)

| Coeff. to be DPCM coded. \ Vector size | 4 | 8 | 16 | 32 | 64 | 128 |
|---|---|---|---|---|---|---|
| $H_2$ only | 39 | 38 | 37 | 36 | 35 | 34 |
| $H_3$ or $H_4$ | -- | 37 | 36 | 34.7 | 33 | 31 |
| Any one of ($H_5$-$H_8$) | -- | -- | 31.7 | 30 | 28.1 | 26 |
| All of $H_2$, $H_3$, and $H_4$ | 40 | 39.6 | 39.2 | 38.6 | 37.8 | 37 |

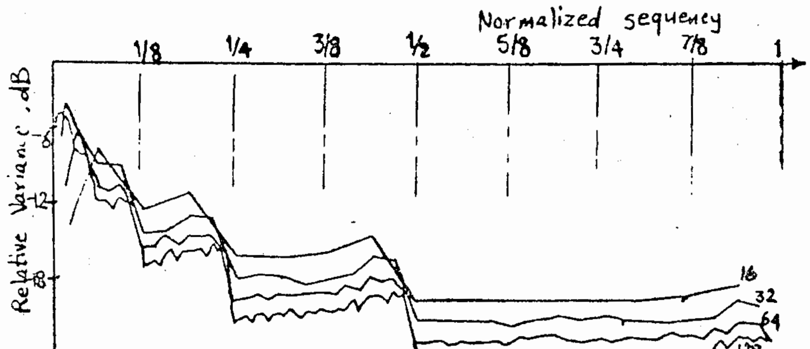


Figure (1). Relative energy content in different coefficients.

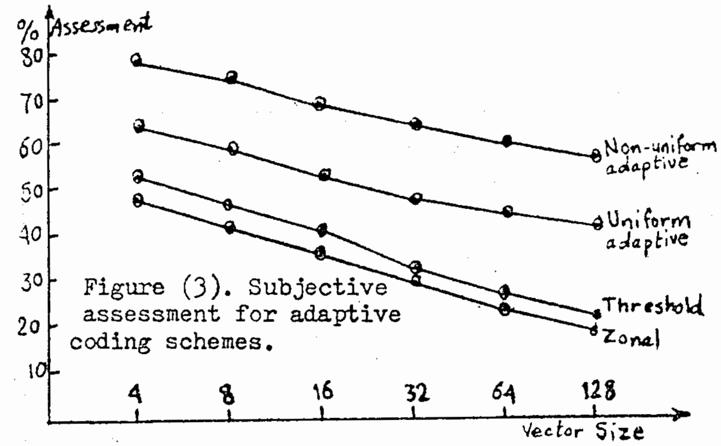


Figure (3). Subjective assessment for adaptive coding schemes.

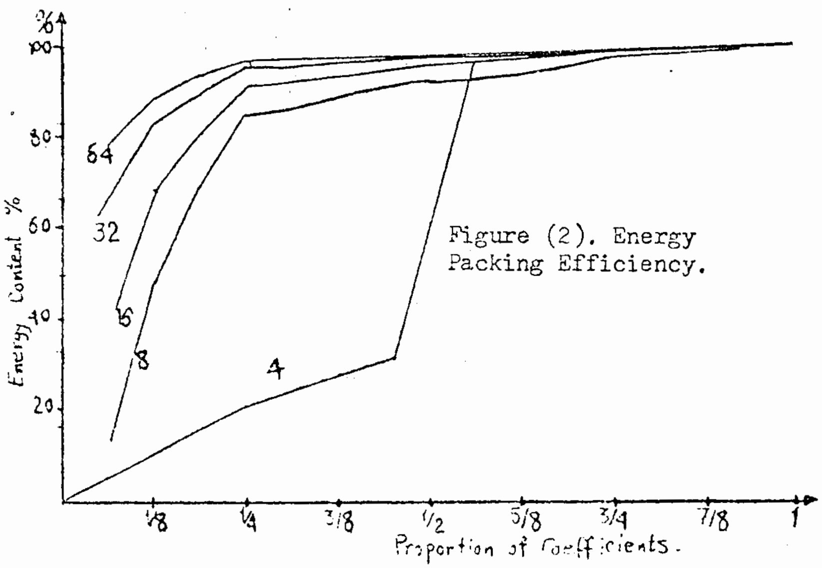


Figure (2). Energy Packing Efficiency.

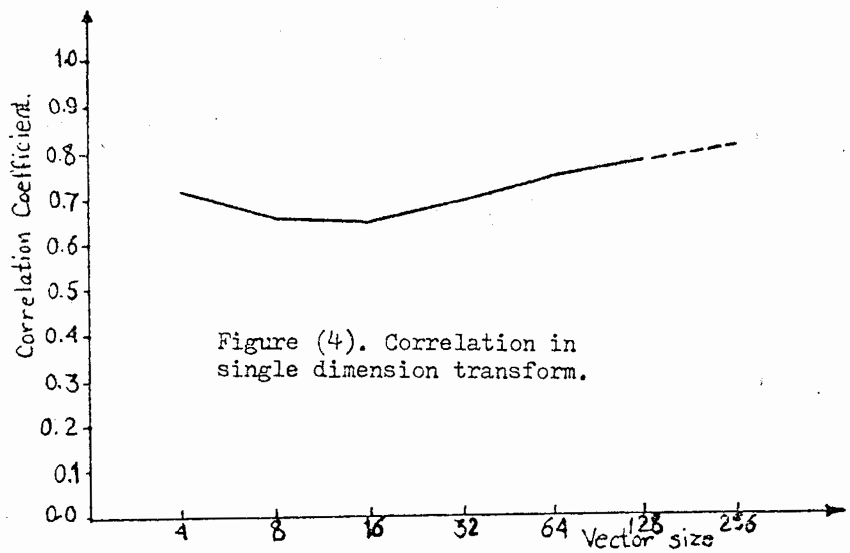


Figure (4). Correlation in single dimension transform.

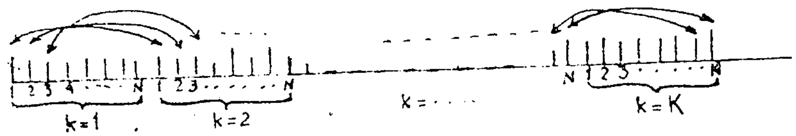


Figure (5). Consecutive vectors grouped for temporal correlation analysis.

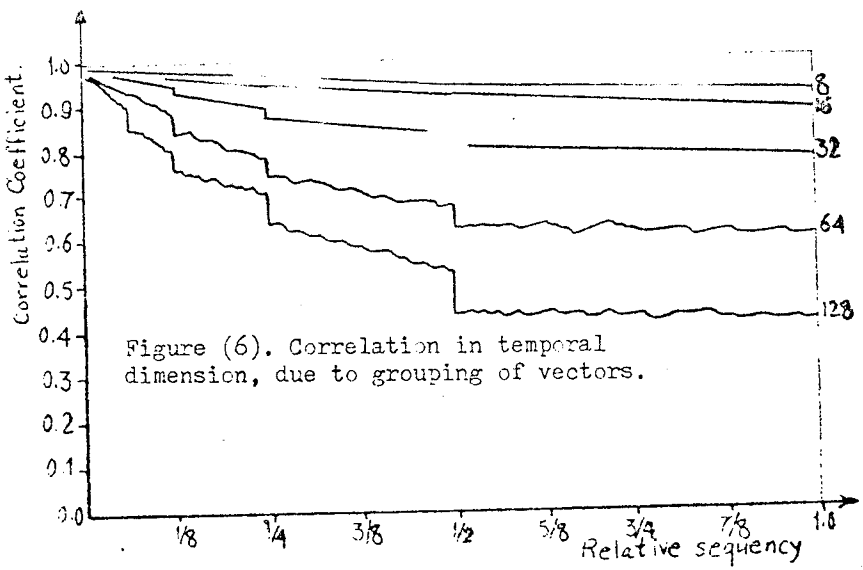


Figure (6). Correlation in temporal dimension, due to grouping of vectors.

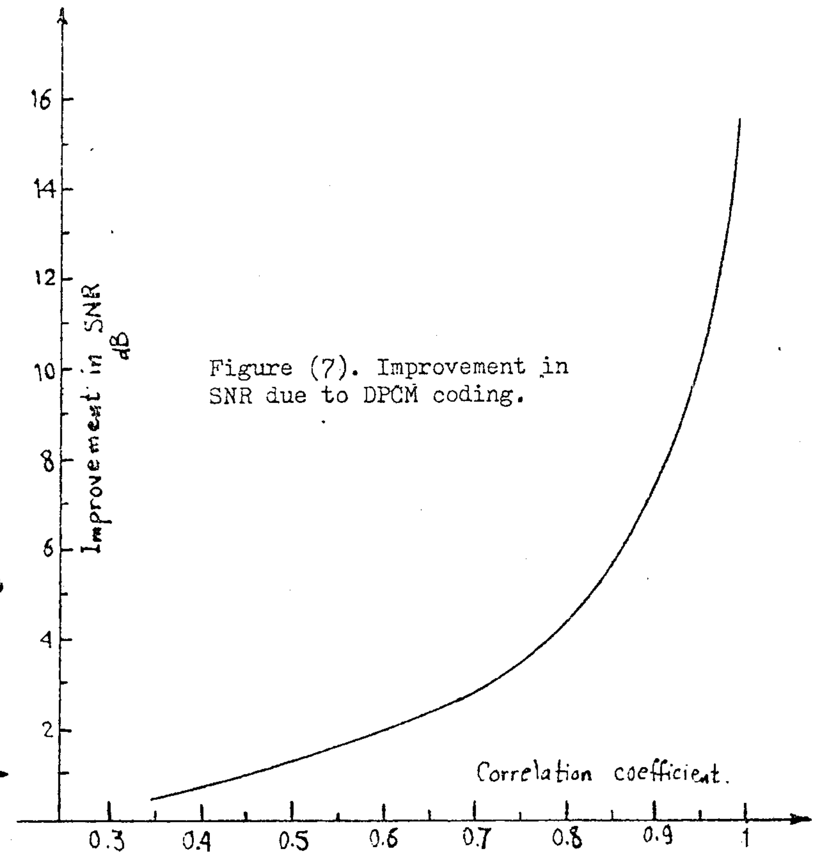


Figure (7). Improvement in SNR due to DPCM coding.

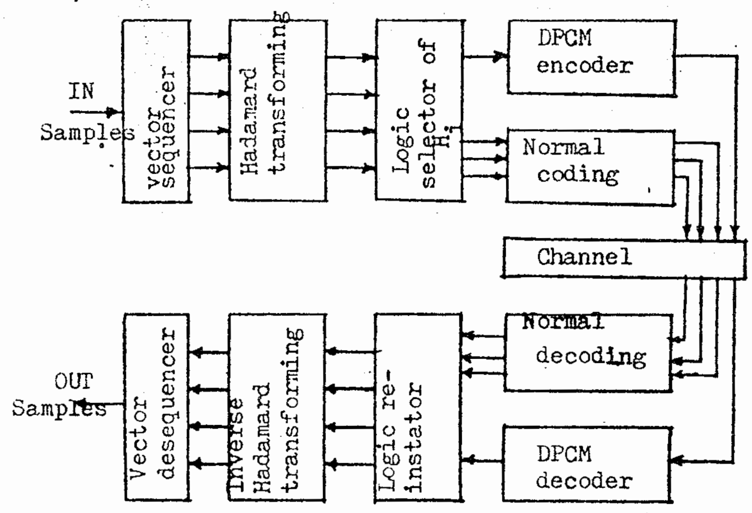


Figure (8). Block diagram of DPCM system in transform domain, for selected coefficient, $H_i$.

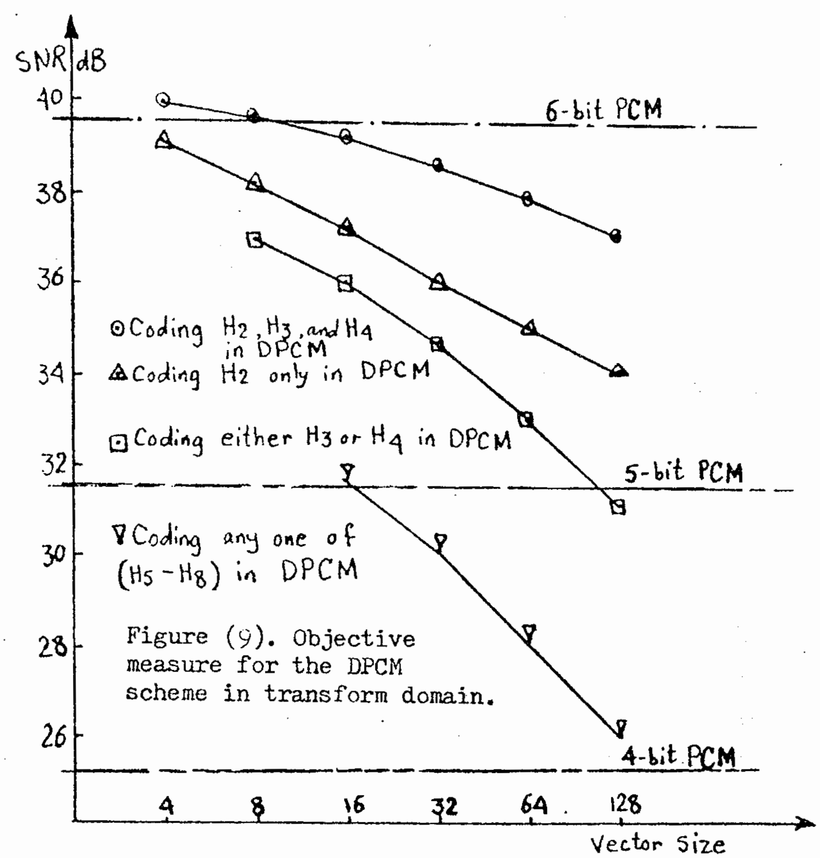


Figure (9). Objective measure for the DPCM scheme in transform domain.