XVII IMEKO World Congress
Metrology in the 3rd Millennium
June 22–27, 2003, Dubrovnik, Croatia

EFFICIENCY ESTIMATION CRITERIA OF LOSSY COMPRESSION METHODS OF MEASUREMENT SIGNALS

Waclaw Gawedzki, *Jerzy Jurkiewicz*

University of Mining and Metallurgy,
Department of Measurement and Instrumentation, Krakow, Poland

Abstract – The paper presents problems of compressing measurement signals recorded under long-lasting monitoring conditions. The criteria for compression method selection in view of its efficiency (compression ratio) and quality of conversion (coding-decoding procedure), with regard to dynamic properties of the measurement signals are discussed. Attention is given to the interdependence of the achieved compression ratio and the distortion level. A method is proposed for generating synthetic test signals to investigate the efficiency of lossy compression methods, which enables unification of the achieved results.

Keywords: signal compression, quality criteria, test signals

1. INTRODUCTION

Compressing measurement signals of physical quantities makes it possible to considerably reduce the quantity of stored information by:
- removal of redundancy and signal de-correlation in long-lasting recording,
- elimination of insignificant information (e.g. noise).

Information removal by application of lossy compression algorithms results in quality deterioration of the recorded signal. Therefore, analyzing the metrological characteristics of lossy and lossless data compression algorithms, which are essential from the point of view of measurements, should bring an answer to the following questions:
- how high a compression ratio (compression efficiency) is feasible by the various compression algorithms (e.g. [2, 3]) for an assumed distortion level, depending on the class of recorded signals,
- is it possible to establish a classification of measurement signals considering the selection of compression algorithms.

In order to perform this analysis it is necessary to select objective evaluation measures for recording the signals with lossy compression (under widely considered dynamic error definition), as there are no universal and generally accepted measures of signal compression quality. The quality measures quoted in the bibliography [2] (e.g. the signal to noise ratio, or the subjective decoded signal quality scale) are insufficient for an objective evaluation of the quality of measurements conducted with data compression.

2. ESTIMATION CRITERIA OF COMPRESSION METHODS

Fig. 1 shows the block diagram for investigating the efficiency of algorithms for measurement signal compression. The input signal is assumed as a discrete signal *u[n]*, which now occurs most frequently in the digital part of measuring systems. In the process of algorithm efficiency investigation, the signal *u[n]* is first submitted to the compression procedure and then decompressed to the signal $u_r[n]$. In the compression procedure, one can distinguish the processes of de-correlation, quantization and coding. The de-compression procedure consists of decoding and re-correlation to the signal $u_r[n]$. The investigated algorithms can include all mentioned processes or only some of them. For instance, de-correlation processes and quantization with the following re-correlation may not occur, and the whole procedure is then limited to coding and decoding operations. Lossless compression algorithms can be investigated in such a way. In a lossless compression-decompression procedure, the final signal $u_r[n]$ should be identical with the input signal *u[n]*. In the general case of lossy and lossless algorithms, all processes shown in Fig.1 may occur.

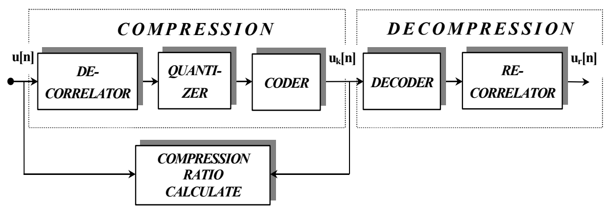

Fig. 1. Block diagram of a system for investigation of compression algorithm efficiency

The procedure for estimating the quality and efficiency of lossy compression methods requires assuming the compression ratio and the corresponding distortion level. As already said, in case of lossy algorithms, the compression ratio can be improved at the sacrifice of increased losses in the compressed signal, and consequently, deteriorating the quality of the reproduced signal $u_r[n]$. Considering the close connection of the compression ratio with the signal conversion procedure of compression-decompression, let us determine the way of estimating the quality of that conversion.

The *u[n]* signal is submitted to the compression-decompression operation, and the $u_r[n]$ signal obtained in effect is being compared with the primary signal $u[n]$ according to a properly formulated error criterion. That criterion is mostly assumed in the form of the error functional F, determined by the signal difference $u[n]$ and $u_r[n]$. The value I of that functional follows generally the relation:

$$I = F\{u[n],\, u_r[n]\} \tag{1}$$

The functional F (1) constitutes the quality estimation measure of recording with signal compression. Frequently applied error measures include integral criteria, integral criteria in generalized form by introducing an additional weight function, or non-integral ones, like the maximum error magnitude criterion. Proper criteria are being selected according to the kind of converted signals and considering the consistence of the physical meaning of the criterion with the measured quantity expressed by the signal $u[n]$. The error criteria can defined both directly in the time domain, as in the frequency domain, and apply Fourier transforms for the $u[n]$ and $u_r[n]$ signals which are compared.

Proper realization of the comparing process and estimation of the processing quality of algorithms requires determination, for each criterion (1) to be applied, of its maximum value, which is closely related to the *a priori* admitted possible losses in the compression process. Standardizing the determined error values for the maximum values secures in that case the comparability of the algorithm quality estimations, considered under different error criteria. The most frequently applied criteria with a useful physical interpretation include the maximum value of the error magnitude criterion and the mean squared error ones, which can be written in normalized form as follows:

$$I_A = \frac{\max\limits_n |\, u[n] - u_r[n]\,|}{\max\limits_n |\, u[n]\,|} \tag{2}$$

$$I_{MSE} = \frac{\sum\limits_n (u[n] - u_r[n])^2}{\sum\limits_n (u[n])^2} \tag{3}$$

3. SYNTHETIC TEST SIGNALS

The efficiency (determined by the compression ratio) and quality (following distortions brought in by the compression) of compression algorithm operation in metrological applications depends not only on their characteristics, but also very much on the characteristics of the data stream (signal). Consequently, the ‘objectivitization’ of research results will require determination of the class of test signals, which can be applied to estimate the compression methods. On the base of an analysis of parameters characterizing the recorded measurement signals, we have to generate synthetic signals (a model), having similar parameters. Then, the synthetic signals can be submitted to compression-decompression operations using selected compression algorithms, which are practical from the metrological point of view. Modifying the parameters of the synthetic signals makes it possible to investigate the operation of data compression algorithms in specific intervals of signal parameter variations, and not at an individual parameter value, which is the case when examining actual signals.

Let us de-correlate the measurements signals using the Karhunen-Loeve transform (KLT) [2,4]. KLT results in a vectorial signal with non-correlated components. The whole energy of the converted signal is accumulated in its several first components, which, by disregarding the remaining ones (lossy compression), enables achieving a considerably improved compression ratio. In view of the fact that the KLT transform is the optimum one, upholding complete de-correlation, it can form the reference point for other transforms when evaluating the degree of signal de-correlation effected by them.

Let us consider mutual relationships between errors (1) introduced in the process of the KLT lossy compression and the resulting compression ratio CR versus data block of size P received arbitrary in the KLT transformation, and number K of the remaining transformation coefficients and eigenvectors (in the lossy compression process). We define the compression ratio CR as the ratio of number of data N (cut to the integer multiple of P) to the number of data after compression (eigenvectors and transformation coefficients necessary for storing):

$$CR = \frac{\lfloor N/P \rfloor \cdot P}{\lfloor N/P \rfloor \cdot K + P \cdot K} \tag{4}$$

There is no unique criterion of selection of data block length P in the KLT transformation, and practical reasons decide on P most often [4]. It should taken into account that if P increases then transformations become more complex and the probability of changing the statistics of data block of length greater than P increases – this results in an increase of the number of significant transformation coefficients and in the reduction of the compression ratio.

In Figs. 2 through 4 an exemplifying air temperature measurement signal S1 recorded in a long-term period, and, determined for it by means of the KLT transformation, error characteristic I_{MSE} (3) and compression ratio CR versus block data size P for various values K remaining transformation coefficients and eigenvectors (in the lossy compression) are presented. In Fig. 5 there are presented, after the fusion of the Figs. 3 and 4 characteristics, the relationships between the compression ratio and the I_{MSE} error for different values of K of the remaining transformation coefficients and eigenvectors. Likewise, in Figs. 6 through 9, the relationships identical as for the S1 signal for the second measurement signal S2, recorded as exemplary, of relative air humidity, are presented.

Basing on the characteristics presented in Figs. 3 and 4 for signal S1, and, respectively, in Figs. 7 and 8 for signal S2, it is not possible to directly select optimally a data block size P and the number K of remaining (within the lossy compression) transformation coefficients and eigenvectors. However, interesting results can be obtained analysing the characteristics in Figs. 5 and 9 obtained from the fusion and showing for signals S1 and S2 the relationships between the compression ratio CR and the I_{MSE} error for different values of K. Basing on Figs. 5 and 4 let us see that, e.g., to obtain the compression ratio CR=5 it is better, with regard to the I_{MSE} error, to use a data block of length P=10 and retain K=2 KLT basis vectors and transformation coefficients than

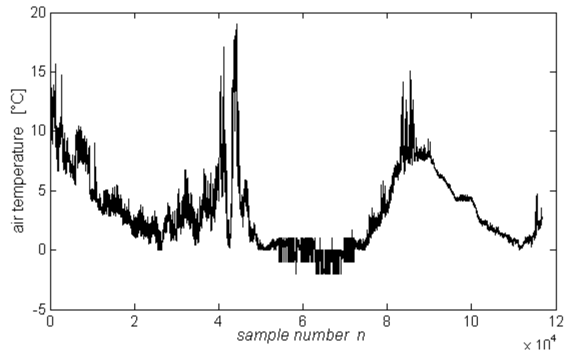

Fig.2. Exemplary measurement signal: S1

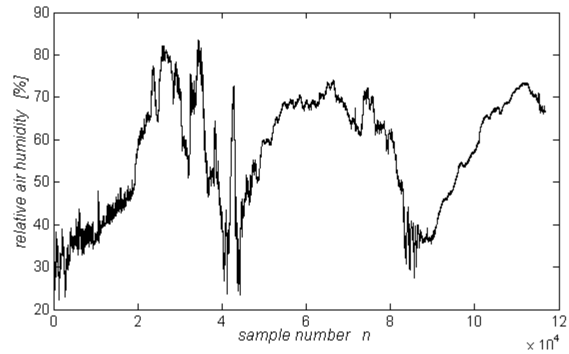

Fig.6. Exemplary measurement signal: S2

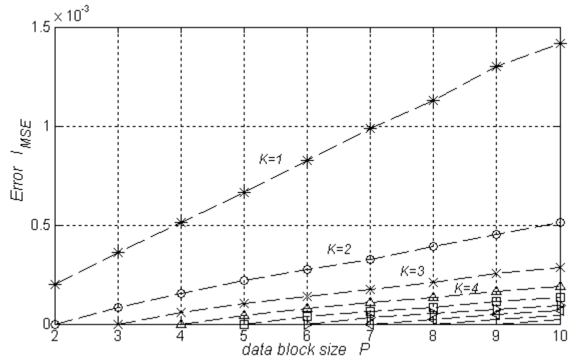

Fig.3. I_{MSE} (3) error for signal S1 versus data block size P in the KLT transformation

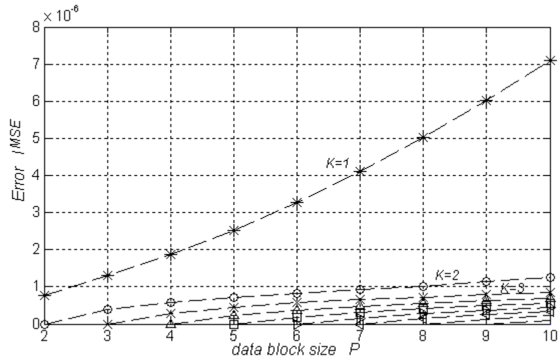

Fig.7. I_{MSE} (3) error for signal S2 versus data block size P in the KLT transformation

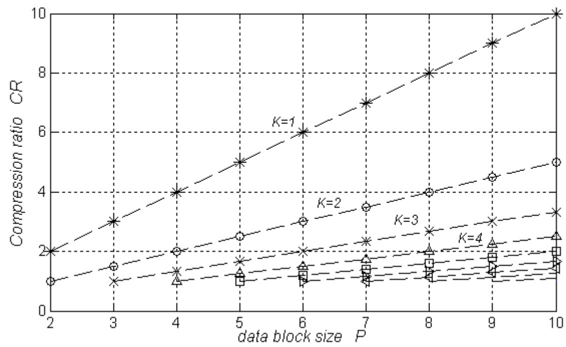

Fig.4. Compression ratio CR (4) for signal S1 versus data block size P in KLT

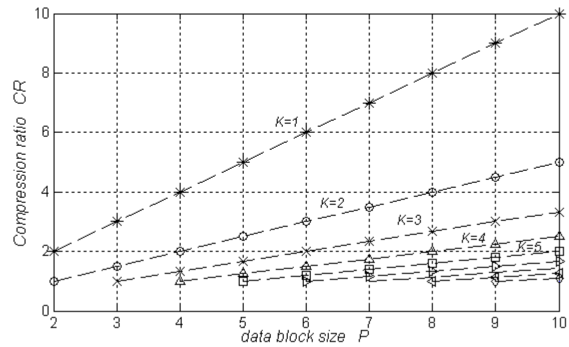

Fig.8. Compression ratio CR (4) for signal S2 versus data block size P in KLT

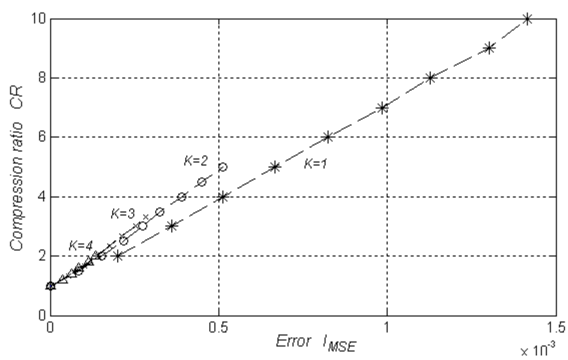

Fig.5. Compression ratio CR (4) for signal S1 versus the I_{MSE} (3) error

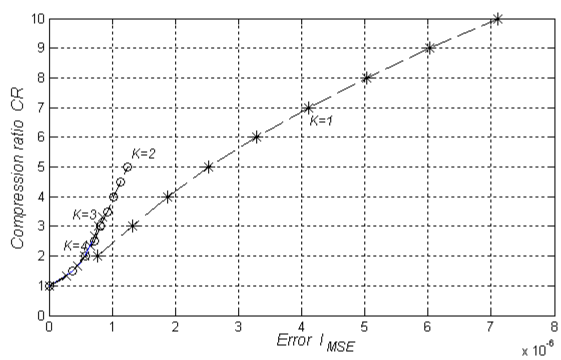

Fig.9. Compression ratio CR (4) for signal S2 versus the I_{MSE} (3) error

P=5 and K=1. Similarly, to obtain the compression ratio CR=3 it would be better, with regard to the I_{MSE} error, to use a data block of length P=9 and retain K=3 KLT basis vectors and transformation coefficients than P=6 and K=2 or P=3 and K=1. Basing on characteristics form Figs. 8 and 9, analogical conclusion can be drawn for measurement signal S2. Using relationships from Figs. 5 and 9, a general conclusion can be drawn that it is better to use data blocks of bigger sizes P and, what follows (to preserve the assumed error) adequately greater values of K. However, with increasing P and K the values of the I_{MSE} error decrease at slower and slower step, and the complexity of transformations and probability of changing data block statistics increase.

Fig. 10 presents two plots of the error I_{MSE} (3) in function of the number K of first eigenvectors used to approximate the re-correlation process, for two exemplifying signals S1 and S2 (the lower the K value the higher the compression algorithm efficiency). It can be easily noted that the limit value of efficiency (curves 1 and 2), attainable at a given signal distortion level (I_{MSE}) for other lossy algorithms (effecting incomplete de-correlation), changes depending on the signal type. Fig. 10 shows the additional *Limit* curve determined for the u_0 signal at the maximizing I_{MSE} criterion:

$$u_o = \arg \max_{u \in U} I_{MSE} \tag{5}$$

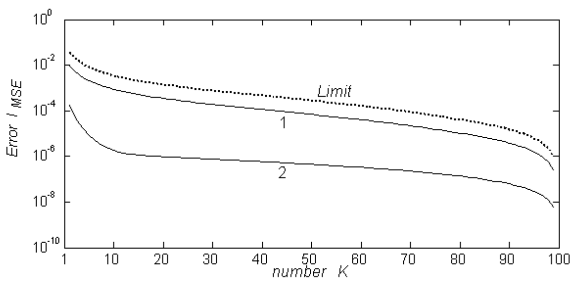

Fig. 10. I_{MSE} (3) error versus number K of first eigenvectors (for $P=100$)

The test signal u_0 can be determined [1] by simulation with additional limitations imposed on the signal ($u \in U$), e.g. limitation of the signal amplitude, of its variability rate, etc. Applying the u_0 signal for investigation of lossy compression algorithms renders the achieved results independent from the shape of individual signals, and makes them valid for the whole class of measurement signals.

4. ACKNOWLEDGMENT

This work is supported by the State Committee of Scientific Research, Poland (grant # 8 T10C 016 21).

REFERENCES

[1] E. Layer, W. Gawędzki, “Dynamic of Measurement Systems. Investigation and Estimation”. PWN Warszawa, Poland, 1991 (in polish).

[2] N.S. Jayant, P. Noll, “Digital Coding of Waveforms. Principles and Applications to Speech and Video”. *Prentice-Hall, Inc. Englewood Cliffs*, New Jersey 1984.

[3] R.M. Gray, “Fundamentals of Data Compression”. *Intern. Conference on Information., Communication and Signal Processing*, Singapore, September 1997.

[4] K. Sayood, “Introduction to Data Compression” *Morgan Kaufmann Publishers,* 2 Edition, 2000 U.S.A.

[5] MATLAB & SIMULINK for Windows - *User’s Guide The MathWorks, Inc*.

[6] W. Gawędzki, “An Adaptive Algorithm for the Measurement Data Compression”. *Proceedings of the XVI IMEKO World Congress*, Vienna Sept. 25-28, 2000.

[7] W. Gawędzki, “Analysis of Measurement Signals Compression Algorithm Properties in the Time Domain”. *IEEE Instrumentation & Measurement Technology Conference IMTC/2001*, pp.725-728, Budapest, Hungary, 21-23 May, 2001

[8] W. Gawedzki, J. Jurkiewicz, “Comparative analysis of a Codec with Adaptive Difference Quantization (ADQ) for Compression of Measurement Signals”. *Proceedings of 12th IMEKO TC4 International Symposium – Electrical Measurements and Instrumentation*. Part 1, pp. 181-184, Zagreb, Croatia, September 25-27, 2002

[9] J. Jurkiewicz, W. Gawędzki, “Compression of Digital Measurement Signals by Reverse Scaling Reconstruction Method”. *Proceedings of 12th IMEKO TC4 International Symposium – Electrical Measurements and Instrumentation*. Part 1, pp.190-193, Zagreb, Croatia, September 25-27, 2002

AUTHOR(S): Ph.D. Waclaw Gawedzki, Ph.D. Jerzy Jurkiewicz, Department of Measurement & Instrumentation, University of Mining & Metallurgy, Al. Mickiewicza 30, 30-059 Krakow, Poland, Phone: (+4812) 6172828, Fax: (+4812) 6338565, E-mail: waga@uci.agh.edu.pl