INTEGRATED DATA COMMUNICATION SYSTEMS WITH DATA COMPRESSION AND ERROR CORRECTING CODES

G.Benelli, V.Cappellini, E.Del Re

Istituto di Elettronica, Facoltà di Ingegneria,
Università di Firenze, Firenze, Italy

ABSTRACT. Two data transformations - data compression and error control coding - are considered for high efficiency integrated data communication systems. Some methods of data compression and error correcting coding are in particular described. Comparison results giving the compression ratio and r.m.s. error for some data communication systems using in different way data compression and channel coding are shown. A new integrated compression-coding stategy - called COSYDAI - is also proposed, in which a suitable synchronization is added to compressed data, to obtain error detection and error correction capabilities.

1. INTRODUCTION

High efficiency communication systems are required in the next future to solve the problems connected to the large amount of data to be transmitted from one place to another through noisy communication channels with severely limited bandwidths. One important data transformation, which can be performed to increase the efficiency of a communication system is the data compression, which reduces the amount of not useful or redundant data. In the literature the data compression techniques are often considered on error-free channels. Neverthless the compressed data are more sensible to the channel errors than the normal data, because at the receiver the signal is reconstructed using a smaller number of bits. For this it can be useful and often necessary to use together data

---

This work has been supported by the Italian National Research Council (CNR) under contr.n. 76.00439.07.

compression the channel coding,which protects the data against disturbances,interferences and noise in the communication channel.
In this paper the performances,obtained through a computer simulation, of communication systems using data compression with or without error correcting codes are presented in the case of channels with memory.As data compression methods using prediction or interpolation and spline functions are mainly considered,while as error control coding binoid codes and generalized Hamming codes for burst error correction are considered. After a new integrated compression-coding strategy,called COSYDAI,is also proposed.For transmission channel with memory,this strategy uses in a suitable way the synchronization information,necessary in the compressed data, to detect and in many cases to correct the channel errors.The COSYDAI strategy permits a higher efficiency than other communication systems using both data compression and channel coding.

2.SOME DATA COMPRESSION AND ERROR CORRECTING CODES TO BE USED IN THE INTEGRATED SYSTEM

Let us consider now two methods of data compression,one with prediction,the other with spline functions,which seem well suitable for applications in integrated digital communication systems having also error control coding transformations.
The zero order prediction (ZOP) algorithm with floating aperture is based on the relation [1]

$$y_{pi} = y_{i-1} \pm \Delta \tag{1}$$

where y_{pi} is the predicted sample at the time $t = t_i$, y_{pi} is the preceeding sample and Δ is the aperture (or tolerance).
The predicted sample y_{pi} is compared with the actual sample y_i and if the resulting difference is in the allowable error tolerance $\pm\Delta$,the actual sample y_i is discarded.Otherwise y_i is considered one of non-redundant sample set to be transmitted.
A simple adaptive modification of this algorithm,we are proposing, can be realized considering two error tolerance values : a smaller value Δ_1 for those parts of the signal in which a higher precision is required and a greater value Δ_2 for those parts in which a lower precision is allowed.The two band tolerances are choosen according to the actual value of the samples.
This algorithm can result in a good performance especially with those signal that the most part of the time are in the amplitude range where a lower precision is required.This compression algorithm has been tested applying it to two types of signals : to an original (not-preprocessed)signal and to a lowpass digitally prefiltered signal to eliminate the high frequency components not of interest of the signal spectrum.A net improvement in the algorithm efficiency resulted in case of prefiltered signal compression.Generally a suitable signal lowpass filtering is desiderable in the

application of this simple compression method, to obtain an overall greater efficiency.

The adaptive compression algorithm using spline functions [2] is based on the approximation of the original signal by means of suitably selected linear segments. In this method it is supposed that the available signal is affected by addittive-type noise and its variance is known or may be evaluated as, for instance, when some time intervals exist in which the signal is absent [2, 3]. The available data samples are of the form

$$y_i = s_i + n_i \tag{2}$$

where s_i are the signal samples and n_i the samples of the additive noise, which is supposed to have a Gaussian distribution with zero mean and constant variance. The compression algorithm is based on the approximation of the signal sample s_i by linear or first-order spline functions : suitable time intervals are selected so that in each of them a linear function is found to approximate closely the signal samples. The number and the length of the intervals and the parameters defining the linear approximation in each of them are determined as follow. The noise variance σ_{n_0} is preliminarily evaluated from n_0 consecutive samples of a signal-free time interval. The compression algorithm at n-th step considers n samples of the available signal y_i and determines the best linear approximation $\hat{y}_i$ to them by the least-squares criterion. The mean σ_n^2 of the squares $(y_i-\hat{y}_i)^2$ is evaluated, which can be considered an estimator of the noise variance. Therefore the quantity $\sigma_n^2/\sigma_{n_0}^2$ is an applicable statistics to decide whether the samples y_i are significantly apart from the linear approximation $\hat{y}_i$. This decision must be carried out according to the statistical criterion related to the random variable $\sigma_n^2/\sigma_{n_0}^2$ which has a Fisher distribution with $(n-1, n_0-1)$ degrees of freedom. If the approximation is statistically justified, a new linear approximation $\hat{y}_i$ of n+1 samples y_i (the preceding n samples plus the next one) is evaluated and statistically verified; if the approximation of n samples is not statistically justified, the approximation at the previous step is considered valid, the interval length n-1 and the parameters defining the linear approximation in this interval are transmitted and the procedure starts again for the evaluation of the length and the best linear approximation of a following interval.

It must be pointed out that this method, having the peculiar characteristic of involving the noise level in the data approximation procedure, performs a reduction of noise components (in particular for high frequency noise it performs a sort of smoothing). Therefore in some cases it is not necessary to filter the signal before the application of the compression algorithm based on the spline function approximation. The signal prefiltering, indeed, results in general useful in the application of prediction algorithms to obtain a higher efficiency. Finally it is to point out that in any considered data compression method the synchronization data must

be inserted.Two pratical approaches on this line are :
a)counting of the data taken out before the considered sample ;
b)counting of the sample position in a frame. The first synchronization method permits a higher compression ratio,but in general is more sensible to the channel errors.
In the following we introduce shortly some classes for burst error correction,which have in general simple encoding and decoding operations (particulary when a general purpose computer is used) and which are useful in the integrated systems,using data compression and error correcting codes.Firstly we describe a class of codes, called "Generalized Hamming codes" for burst error correction,obtained through a suitable generalization of the Hamming codes for single error correction [4].The Hamming codes for single error correction defined in a Galois field GF(q) have a parity-check matrix H_1 with m rows and $(q^m-1)/(q-1)$ columns.To obtain a code able to correct all the bursts of length b or less we consider a parity-check matrix H obtained from H_1 by substituting the elements of H_1 with the element itself multiplied by I_b (the bxb identity matrix).These codes have a very simple and fast decoding algorithm and contain,like particular cases,many other subclasses of codes with interesting error detection and correction capabilities.For example we can obtain,with m = 2, the Samoilenko binoid codes[5] which are optimum respect to the Reiger bound[6].We can obtain also some other new classes of binoid codes with a higher code-rate respect the Samoilenko codes and with some error detection capabilities together with error correction capacity [7].Another subclass is some binary burst error correcting codes,which can be decoded in a simple way both by a ge ral computer or by a decoder of the Meggit type[4].

3.A NEW INTEGRATED COMPRESSION-CODING STRATEGY

In this section a new integrated compression-coding startegy is proposed,in which the necessary time synchronization information (t.i.) in data compression algorithms is utilized in a suitable way to detect and correct channel errors.This method,called COSYDAI, (compression with synchronization control and data interpolation) presents,like it is shown in the next section,an higher efficiency than other communication systems using data compression and error correcting codes when the transmission channel is with memory. The compression data algorithms,described in the previous section, are very sensible to the errors in the time information bits and in the most significant bits of the samples.The first time information method (section 2) permits a higher compression ratio CR than the second method.Neverthless it is more sensitive to the channel errors.In the second methos,in fact, it is possible to detect some errors in the time information.In fact the t.i. sequence in the second method is strictly increasing.If an error changes a number t_i to a value t'_i such that the sequence of the t_i is not in-

creasing,then the error can be detected.If $t_i' > t_{i+2}$ or $t_i' < t_{i-2}$ the i-th received time information number t_i' is,with high probability,in error.In these cases,in our simulated systems,we replace t_i' by a value equal to the mean between the preceeding and the following t.i. received number.This value is generally enough near to that exact.
Neverthless inthe case of burst-type errors,this method is not very good (as shown in section 4).In fact several consecutive t.i. are often corrupted by the bursts and therefore the necessary conditions to modify the value of that wrong are not verified and many new errors can be introduced using the previous correction of erroneous t.i.Moreover also the samples between the erroneous t.i. are often altered by the channel noise,for the nature of the burst errors.

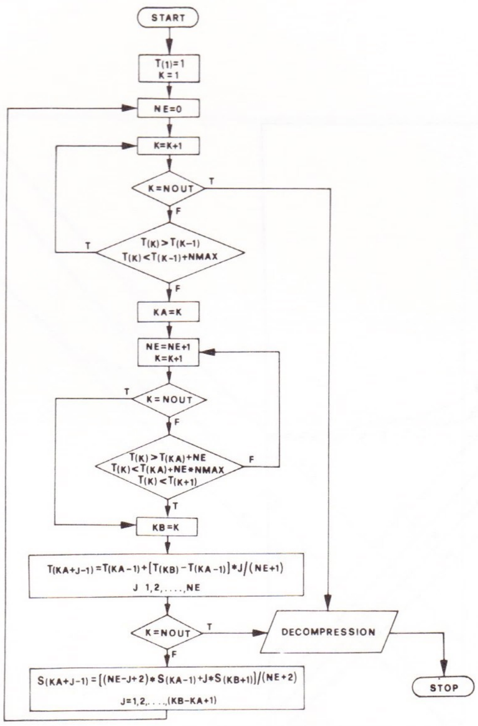

Fig.1- Block-diagram of COSYDAI system

In the COSYDAI strategy the second time information method is utilized. Also the data compression is made so that a fixed number of samples greater than NMAX is not eliminated consecutively. To the receiver, firstly, the t.i. succession is examined to see if the sequence of t.i. is increasing and obeys the previous restrictions. Until this is verified, nearly certainly any error is occurred and therefore nothing is modified. When some consecutive t.i. are detected in error following the previous criterion, they are replaced by new t.i. equidistant between themselves and with values included between the last exact which precedes them and the first we find again exact after that wrong. The values of the samples relative to the mistaken t.i. and that immediately following are modified because they are generally in error: the values that are now assigned to those wrong samples are obtained making a weighed mean of the values of the exact samples which precede them and of that

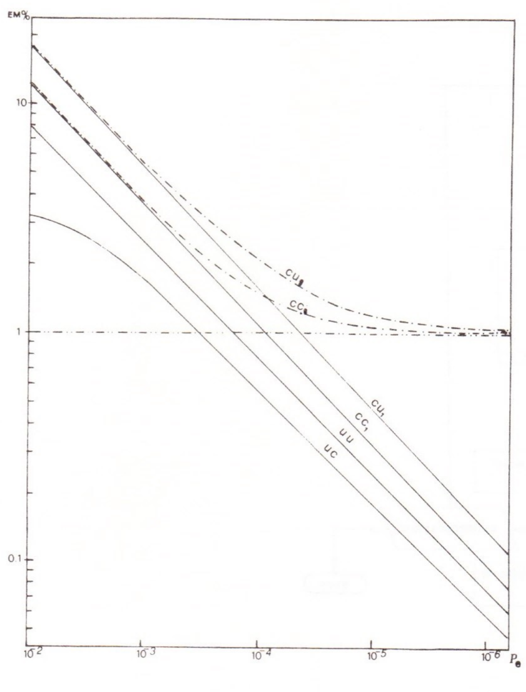

Fig.2-Simulation results regarding some integrated data communication systems

which follow them.The weights are given to the exact samples are inversely proportional to the distance between the sample to replace and that exact.The block-diagram of COSYDAI algorithm is shown in Fig.1.

4.IMPLEMENTATION OF SYSTEMS USING DATA COMPRESSION AND ERROR CORRECTING CODES

In this section the results obtained by a computer simulation of some systems using data compression with or without channel coding are presented.
To characterize the performance of these systems using data compression,we have simulated four structures : uncompressed-uncoded (UU),compressed-uncoded (CU),uncompressed-coded(UC) and compressed-coded (CC) structure.For all these cases,we have obtained the compression ratio CR and the r.m.s. error EM (as percentual of the full scale).

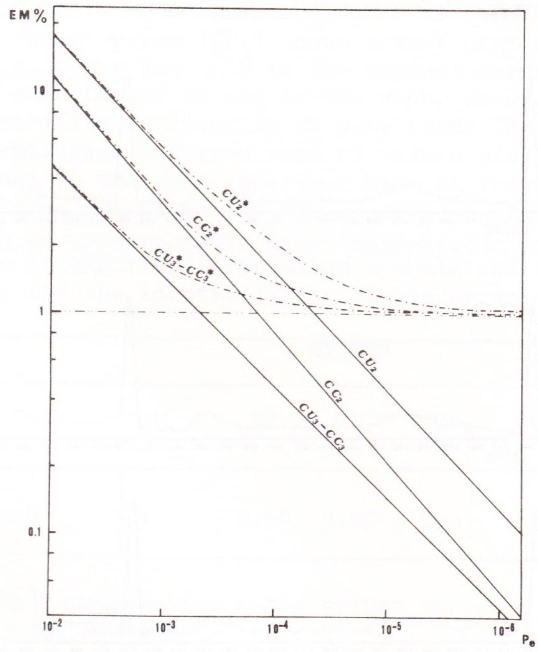

Fig.3-Simulation results regarding some other integrated data communication systems.

In the systems CU and CC,EM value includes the distortion introduced by the compression error and channel noise.To show the different influence of these two error types,we have also computed the r.m.s. error due only to the channel noise;it is obtained through the comparison between the reconstructed data at the receiver from the compressed vector with and without channel errors. In the following,the results considering only the distortion introduced by the channel noise are denoted without asterisk.
The channel noise is simulated by the Gilbert channel model,using a Markov chain with two states [8],which describes approximately the behaviour of some channels with memory,like telephone channels.The utilized code is a Samoilenko binoid code (90,80) defined in a Galois field GF(31).In the binary transmission,as we have simulated,the code is of the type (450,400) and it is able to correct all the bursts of length 21 bits or less.

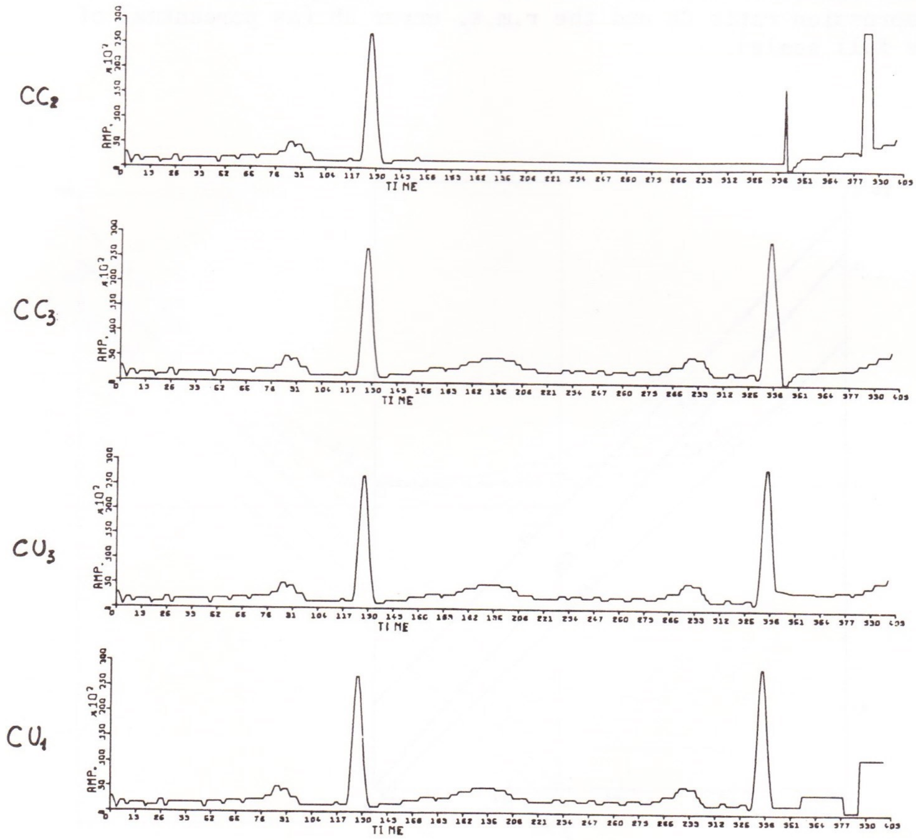

Fig.4-Examples of reconstructed ECG analog waveforms for several transformations :reconstructed ECG with CU_1,CU_3,CC_3(closely similar to the original signal),CC_2.

In our simulated system,the utilized signal,which is an electrocardiogram (ECG),is firstly processed using a low-pass filter and after compressed by a ZOP algorithm with floating aperture.For a tolerance Δ= 2.18% respect to the full scale,we have obtained an r.m.s. error EM = 1%.It is clear that in the cases CU and CC it is impossible to go below this value.
The obtained r.m.s. error EM,expressed like percentual of the full scale,versus the channel error probability P_e is shown in Fig.2 and 3 for all the simulated systems.In these figures we have denoted with the index 1 the system using the first t.i. method, with the index 2 the second t.i. method and with the index 3 the modified system of Fig.1. In the Table 1 the compression ratio CR for the different systems is shown.
For high P_e the EM error is principally determined by the channel noise,while the influence of the compression distortion is negligeable.By reducing P_e,the importance of the compression errors becames more and more high.
It is important to note that using the second method for time information (CU_2) ,the improvement is very low respect to the case CU_1.At the same time the CR value is reduced to 2.6.In the case CC_2,particulary for low P_e,we have an improvement for EM respect to CC_1,but the compression ratio is reduced to 2.2.
The third system (CU_3) shows a very high efficiency;in fact the r.m.s. error due only to the channel error,is lower respect to the case UU and UC and to the other cases.This follows from the possibility to identify in many cases the errors in the time synchronization and from this to reduce also the influence of the errors in the most important bits of the samples.
The case CC_3 generally has an efficiency similar to CU_3.In fact when an uncorrectable burst happens,the code can correct some errors in the time information symbols and some information necessary for the identification of the error positions is lost.

	SYSTEM							
	UU	UC	CU_1	CU_2	CU_3	CC_1	CC_2	CC_3
CR	1	0.89	3.07	2.6	2.6	2.73	2.16	2.16

TABLE 1-Compression ratio values for some simulated data communication systems.

In this case a higher EM can results.Fig.4 shows an example of reconstructed ECG anolog waveforms for several transformations above considered and considering a particular channel error structure.
From these results it is clear the great importance of the time

synchronization in the transmission of compressed data on a noisy channel.By a properly choose of the synchronization information, in fact,it is possible to reduce the influence of the channel errors.Naturally to reduce for all the systems the error EM,some code,able to correct longer bursts,can be used.This naturally reduces also the compression ratio.Therefore for the systems CC, it is necessary with reference to the considered signal and channel, a compromise between the EM and CR values.

REFERENCES

1. J.E.Medlin,"Sampled-Data Prediction for Telemetry Bandwidth Compression",IEEE Trans.Space Elect.Telem.,vol AES-3,n.5,p.784-795,September 1967.
2. V.V.Shakin,P.Breuer,"Adaptive Least-Squares Spline Fitting the Vectorial Signal",Proceed.Florence Conf.on Digital Signal Proc. September 1975.
3. G.Benelli,C.Bianciardi,V.Cappellini,E.Del Re,"High Efficiency Digital Communication using Data Compression and Error Correcting Coding",Proceed.Europ.Conf.on Electrotechnics,EUROCON'77, Venice,May 1977.
4. G.Benelli,C.Bianciardi,V.Cappellini,"Generalized Hamming Codes for Burst-Error-Correction",Alta Frequenza,vol.44,November 1975.
5. S.I.Samoilenko,"Binoid Error Correction Codes",IEEE Trans.Inf. Theory,vol-IT-19,January 1973.
6. W.W.Peterson,E.J.WeldonJr.,"Error Correcting Codes",Cambridge Mass.,M.I.T. Press 1972.
7. G.Benelli,C.Bianciardi,V.Cappellini,"Some Burst -Error-Correction Codes",IEEE Trans.Inf.Theory,vol IT-21,November 1975.
8. E.N.Gilbert,"Capacity of Burst-Noisy-Channel",Bell System Tech.Journal,vol 39,September 1960.