

MESSAGING METHOD

This invention relates to a short message format that captures useful information embedded in a data vector of a sequence of symbols or numbers.

- 5 The data vector may represent many different forms of information generated by various electronic and information systems. This short message format is particularly useful when bandwidth limited communication links are used to transmit a data set that can be represented as a set of data vectors that is true for essentially all types of data.

10

Described herein is an algorithm formulated to be useful for data communications problems associated with bandwidth limited communications links.

15 BACKGROUND OF THE INVENTION

- Low data communication bandwidths exist for one or more reasons for many communications systems and the data to be transmitted can thus take inordinate periods of time to transmit because of the low data rate achievable on the channel. Even when communication channels of adequate bandwidth are available, the priority of the data may be such that it is not of adequate benefit to devote that available bandwidth to the task of communicating the data to where it may be of greatest value.

- 20 This highlights the importance of data reduction algorithms for the transmission of data generated by various electronic and information systems.

BRIEF DESCRIPTION OF THE INVENTION

- 30 In this specification a potential solution to this problem is described that uses a sequence of short messages to transmit a data set that can be divided into a

number of data vectors. Each short message captures the essential information embedded in a data vector. A reconstructed data set can then be formed after a number of short messages have been received. The reconstructed data set contains the essential information embedded in the original data set.

5

In a broad aspect of the invention a method of creating a short message for the transmission of information embedded in a data vector, the data vector consisting of data points, and the short message has the form

10

$$\mathbf{M} = [V_1, V_2, \dots, V_N, P_1, P_2, \dots, P_N, \mu, \sigma],$$

where $V_1, V_2, \dots, V_N$ are the values of N peaks, $P_1, P_2, \dots, P_N$ are their corresponding positions, μ and σ are the mean and standard deviation of the data points contained in the data vector, wherein the method includes the following steps:

15

selecting the largest peak in said data vector,
selecting the next largest peak if its position is greater than a
predetermined position separation from each of the said previously
selected peaks, otherwise, select the next largest peak, and repeat the
previous selection steps until the predetermined N peaks have been
20 selected or the list of peaks in said data vector in descending order has
been exhausted.

In yet a further aspect of the invention when values of interest are present in the data vector, more representative mean and standard deviation of the

25

background are calculated with the further step of removal of the data points around the positions of the values of interest.

In an aspect of the invention a plurality of said short messages can be used to reconstruct the value tracks in the data set, the method of reconstruction after
30 receiving a number of successive short messages comprises, forming a matrix

with the number of rows being the number of short messages received and the number of columns being the number of positions, where in each row, the entries whose columns correspond to peaks in the short message are the peak values, while all other entries are set to a predetermined value to represent background values wherein such a matrix contains the value tracks in the data set with all the selected peak values of the original data set.

A specific embodiment of the invention will now be described in some further detail with reference to and as illustrated in the accompanying figures. This embodiment is illustrative of an underwater sensor and communication environment in which the application of the invention is described but it should not restrict the scope of the invention to this application in other environments or applications. Suggestions and descriptions of other embodiments may be included within the scope of the invention but they may not be illustrated in the accompanying figures or alternatively features of the invention may be shown in the figures but not described in the specification.

BRIEF DESCRIPTION OF THE FIGURES

Fig. 1 depicts an embodiment of the flow diagram of the method of the invention.

DETAILED DESCRIPTION OF AN EMBODIMENT OF THE INVENTION

A purposely-designed short message format can be used to transmit a set of data. Each message summarises the information embedded in a data vector in a compact format that is suitable for applications using bandwidth limited communication links. A sequence of such short messages can be used to reconstruct the data set with the relative values preserved. This short message format can be used in conjunction with other data communications

algorithms such as those utilising image compression and reconstruction algorithms, which are the subject of a separate patent application in the name of the same applicant, entitled "Matrix Compression Arrangements" and filed on the same day as this specification, and which is incorporated herein by
5 reference.

This specification provides a design of a very short message that summaries the information provided by a data vector and is so short that it can be used for applications utilising severely bandwidth limited communication links.
10

For any data vector, the short message takes the form

$$\mathbf{M} = [V_1, V_2, \dots, V_N, P_1, P_2, \dots, P_N, \mu, \sigma],$$

15 where $V_1, V_2, \dots, V_N$ are the values of N peaks, $P_1, P_2, \dots, P_N$ are their corresponding positions, μ and σ are the mean and standard deviation of the data points contained in the data vector. The mean μ is an indication of the background level and the standard deviation σ is an indication of the background variability.

20

If there are values of interest present in the data vector, more accurate mean and standard deviation of the background may be calculated with the removal of the data points around the positions of the values of interest.

25 When selecting the N peaks, a minimum position separation should be predetermined so as to avoid two peaks coming from related data points, which may represent the same source represented by the data points or a common signal source if the data is representative of signals, being chosen.

The value of N will depend to some degree on the wanted or needed size of the short message and this may be determined by a number of factors that include the capacity of the communication channel that is available at the time and/or the number of peaks of interest that are required to be communicated in the short message format. The later requirement may sometimes depend on a variable that could change in a short time and be chosen by a processor or which may be determined by a human operator who is capable of determining the peaks of interest.

- 10 Referring to Fig.1, the procedure, at a minimum, includes the following steps: selecting the largest peak (14), selecting the next largest peak if it passes the position separation test for each of the previously selected peaks (16), otherwise, go to the next one (20), and so on. Stop if N peaks have been selected or the list of peaks in descending order has been exhausted (18).
- 15 Thus, the message (22) contains at most N peaks. The process is illustrated in the flow chart at Fig. 1.

The step of determining the number of peaks N and the required position separation S (10) is a preliminary step that as described can be determined on an as needs basis or may be predetermined.

The step of determining all the peaks in the data vector and sorting the value of those peaks in descending order to form a peak vector A (12) is also a preliminary step of data preparation.

25

The data vector can represent any data and may in one example represent a signal received at a remote sensor having a variety of characteristic including for example a voltage, current, etc. which fluctuate over time and thus have one or more peaks of which none, one or more may be of particular interest.

After receiving a short message of the form described above, the end user may first assess how much some of the peaks are above the background level with consideration being given to the background variability as well. A large
 5 difference in either and both indicates that a value of interest is present which warrants further investigation.

The end user may also perform a statistical analysis to quantify the significance of the peaks compared to the data contained in the data vector
 10 after assuming a commonly used statistical model for the data. Normal and exponential distributions may be the ones most suitable for modelling the signal data contained in the data vector. The probability density function of a normal distribution with mean μ and standard deviation σ is defined as

$$15 \quad g(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \quad (1)$$

and that for an exponential distribution is

$$h(x) = \frac{1}{\alpha} \exp\left(-\frac{x}{\alpha}\right). \quad (2)$$

20

If $f(x)$ is the probability density function of a probability distribution, then the cumulative probability at p is defined as

$$P(X \leq p) = \int_{-\infty}^p f(x) dx, \quad (3)$$

25

which is the probability of a random variable X with the probability distribution defined by the density function $f(x)$ being less than or equal to p .

Thus, for a given p , this cumulative probability characterises how significant the value p is.

Assuming that values of the data vector fit a normal distribution, then it is possible to fit the model to the values to find μ and σ , which are the maximum likelihood estimation of the mean and standard deviation for the given data. Having calculated the probability density function using equation (1), it is possible to evaluate the cumulative probability at each of the N values utilising equation (3). These cumulative probabilities indicate how significant the values are and provide guidance to an end user as to whether a value of interest might be present.

Assuming that the values of the data vector fit an exponential distribution, it is possible to fit the model to find α in maximum likelihood sense. The end user needs this number α to perform the statistical analysis, so it is needed in the short message

$$\mathbf{M} = [V_1, V_2, \dots, V_N, P_1, P_2, \dots, P_N, \mu, \sigma, \alpha].$$

Then, it is possible to evaluate the cumulative probability for each of the N peaks similarly to the normal distribution case.

The short message formed in this way is so short that it can be transmitted to an end user quickly even via a bandwidth limited communication channels. The end user can examine the peak values taking into account of the mean and variance, which represent the background level and its variability. A statistical analysis to quantify the level of significance of the peaks may also be conducted.

With a number of successive short messages, it is possible to form a matrix with the number of rows being the number of short messages received and the number of columns being the number of positions. In each row, the entries whose columns correspond to peaks in the short message are the peak values, while all other entries are set to an appropriate value as described below and will be called the background value.

Let m and n be the maximum and minimum values respectively from all the short messages. One way to set the background value is

10 $b = \min \{4(m-n), 4n/5\}$. It is then possible to display this matrix as an image, which can be referred to as a reconstructed data set. Value tracks can be visualised in this way with the peak values preserved.

It will be appreciated, by those skilled in the art that the invention is not

15 restricted in its use to the particular applications described. Neither is the present invention restricted in its preferred embodiment with regard to the particular elements and/or features described or depicted herein. It will be appreciated that various modifications can be made without departing from the principles of the invention. Therefore, the invention should be

20 understood to include all such modifications within its scope.

CLAIMS

1. A method for creating a short message for the transmission of information embedded in a data vector, the data vector consisting of data points wherein each point has a value and a position, and the short message having the form

$$\mathbf{M} = [V_1, V_2, \dots, V_N, P_1, P_2, \dots, P_N, \mu, \sigma],$$

where $V_1, V_2, \dots, V_N$ are the values of N peaks, $P_1, P_2, \dots, P_N$ are their corresponding positions, μ and σ are the mean and standard deviation of the data points values contained in the data vector, the method includes the following steps:

a) select the data point with the largest value that has not previously been selected from the data vector,

b) if the selected data points position is greater than a predetermined position separation from each of the positions $\mathbf{M}$, then add value and position of the selected data point to $\mathbf{M}$,

c) repeating steps a) and b) until the said N peaks have been selected or the list of peaks in said data vector in descending order has been exhausted.

2. A method according to claim 1 wherein when values of interest are present in the data vector, an alternative mean and standard deviation of the background are calculated by the further step of:

d) removal of data points around the positions of the values of interest before step b).

3. A method according to claim 1 wherein $V_1, V_2, \dots, V_N$ are values of N peaks of a signal, and $P_1, P_2, \dots, P_N$ are another value associated with each of the N peaks, and μ and σ are the mean and standard deviation of the data vector representing the signal.