



УДК 004.046

© *А. В. Богомолов, С. В. Власьевский, В. А. Луговой, А. В. Левенец, Чье Ен Ун, 2012*

КОДИРОВАНИЕ ИЗМЕРИТЕЛЬНЫХ ДАННЫХ СПОСОБОМ ПРЯМОГО ОТОБРАЖЕНИЯ НА ФУНКЦИЮ

Богомолов А. В. – магистрант кафедры «Автоматика и системотехника», тел. 8-962-222-80-50, e-mail: alex-newborn@yandex.ru (ТОГУ); *Власьевский С. В.* – д-р техн. наук, проф. кафедры «Электротехника, электромеханика и электроника» (ДВГУПС); *Луговой В. А.* – д-р физ.-мат. наук, начальник лаборатории оптоакустических измерений (ФГУП ДФ ВНИИФТРИ); *Левенец А. В.* – канд. техн. наук, доцент кафедры «Автоматика и системотехника», тел. 8-914-191-33-39, e-mail: levalvi@bk.ru; *Чье Ен Ун* – д-р техн. наук, проф., зав. кафедрой «Автоматика и системотехника», тел. 8-924-109-08-98, e-mail: chye@ais.khstu.ru (ТОГУ)

В статье предлагается способ кодирования измерительных данных, который заключается в использовании некоторой функции как области отображения на ней исходных данных. Выбор функции позволяет манипулировать статистическими параметрами выходного потока данных, что может быть использовано для повышения эффективности работы алгоритмов сжатия.

Coding of data is a primary stage of the processing. The way in which data are coded can influence efficiency of mathematical data processing, their compression, storage, etc. The way of data coding is proposed in the paper that consists in using a function as an image domain for the initial data. This makes it possible to manipulate parameters of the output data stream, which can be used to improve the efficacy of compression algorithms.

Ключевые слова: кодирование, измерительные данные, преобразование, сжатие данных.

Одной из первичных процедур обработки данных является их кодирование, необходимое для наиболее эффективного представления данных на различных этапах обработки. Так, например, для повышения эффективности алгоритмов сжатия часто используется предварительная обработка данных, позволяющая увеличить коэффициент сжатия по сравнению с необработанными данными [1]. Кодирование данных в этом случае следует рассматривать как такую предварительную обработку данных, которая позволяет либо адаптировать существующие универсальные алгоритмы сжатия для работы с

измерительными данными, либо формирует требуемые статистические параметры исходных данных для специализированных алгоритмов сжатия [2].

Предлагаемый в данной работе способ кодирования заключается в последовательном переборе входных измерительных данных с одновременным поиском аргумента некоторой кодирующей функции $F_K(x)$, при котором ее значение с заданной точностью совпадает со значением текущего отсчета входных данных. После определения текущего значения аргумента x_k , при котором происходит совпадение значений кодирующей функции и текущего отсчета входной последовательности, вычисляется длина временного интервала между предыдущим и текущим закодированными значениями $[x_{k-1}; x_k]$, причем для первого отсчета берется длина интервала, измеряемая от начала оси координат до этого отсчета. В результате каждому элементу исходной последовательности данных сопоставляется элемент последовательности вычисленных длин интервалов.

Следует отметить, что в общем случае кодирующая функция определяется на некотором интервале $[0; T]$, не связанном с временным интервалом, на котором расположена исходная последовательность данных. Очевидно, что в случае периодичности кодирующей функции $F_K(x)$ в качестве такого интервала следует выбирать ее период. Выбор интервала представления для непериодической функции осуществляется исходя из априорных сведений о свойствах кодируемой последовательности данных. Предлагаемый способ также предполагает, что кодирующая функция обладает большим либо равным динамическим диапазоном по сравнению с динамическим диапазоном кодируемых данных.

Из общих соображений можно утверждать, что предлагаемый способ осуществляет нелинейное преобразование исходных данных, что позволяет предположить возможность формирования закодированной последовательности с заданными статистическими параметрами при соответствующем выборе кодирующей функции.

Для первичного исследования предлагаемого способа кодирования в качестве кодирующей функции была использована синусоида с фиксированной частотой и нулевой начальной фазой:

$$F_K(x) = A_0 + (A_{\max} - A_0) \cdot \sin(2\pi x/T),$$

где: T – период синусоиды; $A_0 = (A_{\max} + A_{\min})/2$ – постоянная составляющая; A_{\max} и A_{\min} – максимальное и минимальное значения отсчетов исходной последовательности данных соответственно.

При исследовании предлагаемого способа в качестве исходных данных, подлежащих кодированию, использовались реализации случайного процесса с нормальным распределением.

На рис. 1 показан пример отображения на кодирующую функцию исходной последовательности $\{9, 0, -8, 5, -7, -6, 9\}$. Как было показано выше, кодирование заключается в нахождении интервалов между соседними отсчетами, расположенными на кодирующей функции. Здесь следует отметить, что пе-



риод синусоиды можно без потери общности принять равным единице и оперировать дробными числами. Так, например, расстояние между первым и вторым, а также вторым и третьим отсчетами по графику составляет величину, равную 0,25 периода синусоиды, между третьим и четвертым – примерно 0,38 периода и т. д.

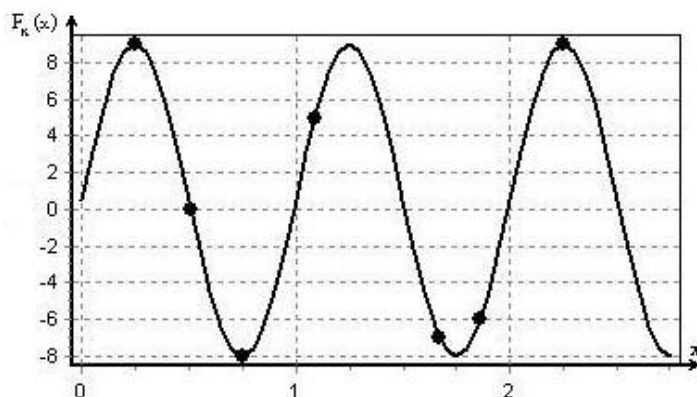


Рис. 1. Кодированная функция с расположенными на ней отсчетами тестовой входной последовательности данных

Таким образом, в результате операции кодирования формируется некоторое множество дробных чисел, поставленных в соответствие с элементами исходной последовательности данных. Можно предположить, что закодированная последовательность, также как и исходная, является случайной, но с измененными статистическими параметрами.

Для исследования предлагаемого способа кодирования представляет интерес оценка статистических параметров выходной последовательности данных. Первичная оценка проводилась по первым четырем статистическим моментам: математическому ожиданию, среднеквадратичному отклонению, коэффициенту асимметрии и коэффициенту эксцесса.

При проведении исследований предложенного способа кодирования оценивалось влияние на статистические параметры как размера выборки, по которому проводится оценка, так и числа усреднений результатов кодирования, необходимого для обеспечения стационарности оценок параметров.

На рис. 2 представлены зависимости статистических моментов от числа усреднений, полученные при объеме выборки исходной последовательности, равной 32. Входные данные представляют собой реализацию гауссовского случайного процесса с нулевым математическим ожиданием и единичной дисперсией.

Полученные данные показывают существенный разброс значений статистических параметров интервальной последовательности данных, полученной при малом числе усреднений, что можно связать со случайным характером распределения отсчетов исследуемого сигнала, что еще раз подчеркивает нелинейный характер предлагаемого кодирования. Тем не менее, следует от-

метить, что при числе усреднений более пятнадцати значения всех четырех моментов становятся относительно стабильными. Аналогичные результаты были получены и для больших размеров выборки, причем графики зависимостей приобретали, как и ожидалось, более гладкий характер.

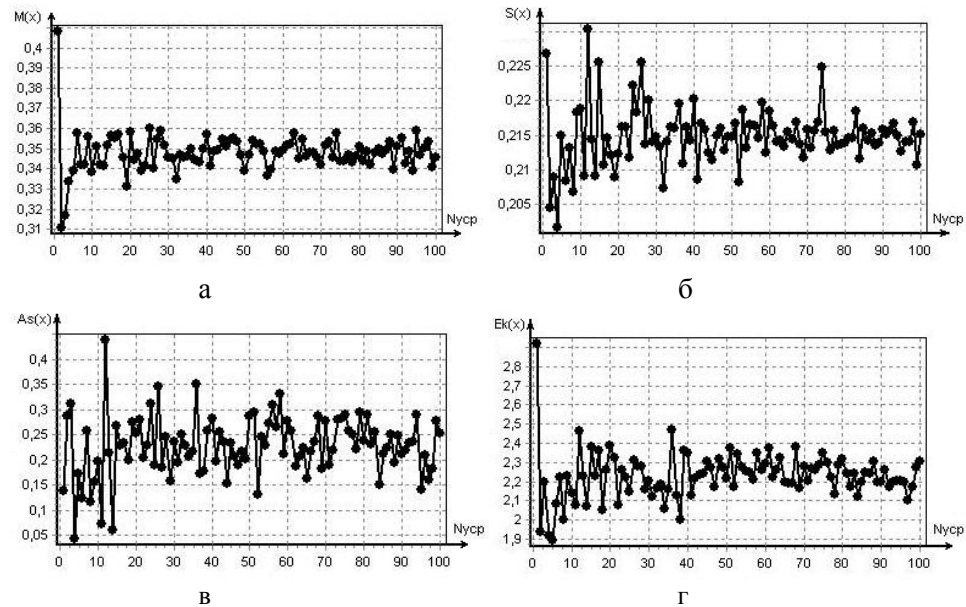


Рис. 2. Зависимость первых статистических моментов закодированной последовательности от числа усреднений $N_{уср}$:
а) математическое ожидание; б) среднеквадратичное отклонение;
в) коэффициент асимметрии; г) коэффициент эксцесса

Оценка поведения статистических моментов закодированной последовательности данных для различных объемов выборки проводилась при числе усреднений 20. Результаты проведенных исследований приведены на рис. 3.

Полученные экспериментальные данные показывают существенный разброс значений статистических параметров интервальной последовательности данных при малых объемах выборок, что можно объяснить случайным характером исходной последовательности данных. Некоторая стабилизация параметров происходит при увеличении объема выборки до 700 и более, однако следует отметить общую нестационарность поведения параметров, что может быть объяснено как неидеальностью используемого генератора случайных чисел, так и проявлением нелинейных свойств предлагаемого способа кодирования.

Анализируя полученные данные, можно сделать вывод о том, что у выходного потока данных по сравнению с входным потоком статистические параметры существенно изменяются. Интересным фактом является существенное отличие математического ожидания от нуля и значительное сниже-



ние величины среднеквадратического отклонения, что позволяет говорить о возможности более эффективного сжатия таких данных, в том числе неспециализированными алгоритмами общего назначения.

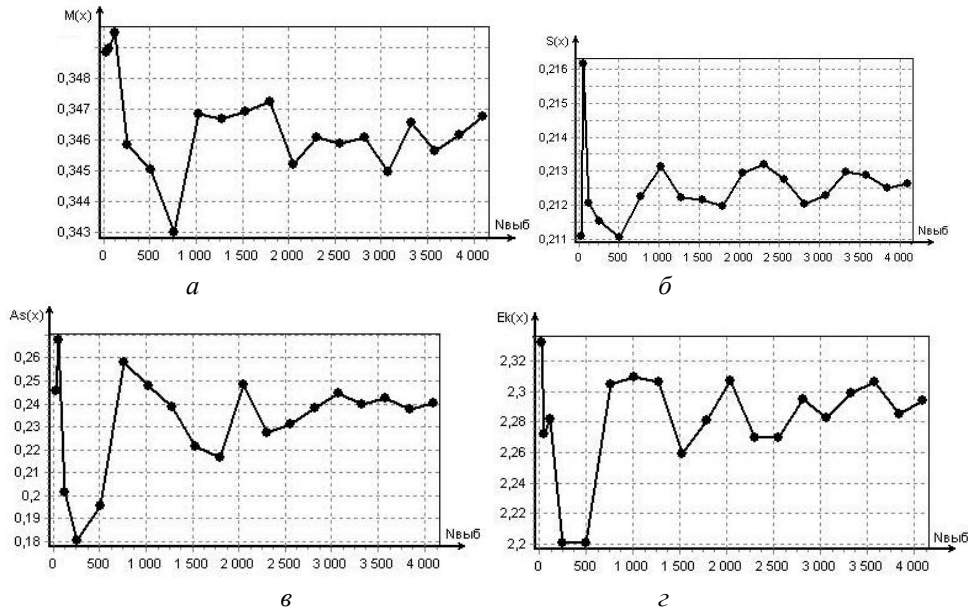


Рис. 3. Зависимость первых статистических моментов закодированной последовательности от объема выборки $N_{\text{выб}}$:
а) математическое ожидание; б) среднеквадратическое отклонение;
в) коэффициент асимметрии; г) коэффициент эксцесса

Следует отметить, что для ряда практических приложений существенным фактором являются жесткие требования к вычислительным затратам, необходимым для реализации способов обработки данных. В этом случае наиболее естественным выходом является применение табличного способа вычисления необходимых функций. Недостатком такого метода являются более высокие требования к объему запоминающего устройства и необходимость учета погрешностей восстановления, возникающих из-за необходимости округлять точное значение функции до ближайшего табличного значения.

В качестве одного из способов снижения объема требуемой памяти можно предложить ограничение разрядности таблиц функций. Следует отметить, что такое решение приведет к еще большему увеличению погрешностей восстановления исходных данных. Для оценки влияния ограничения разрядности значений функций на погрешность восстановления было проведено исследование, в ходе которого изменялось число уровней равномерного квантования кодирующей функции и оценивались погрешности восстановления исходного процесса. При кодировании каждый из отсчетов исходной после-

довательности данных округлялся до ближайшего уровня квантования кодирующей функции.

В ходе исследований была проведена оценка абсолютных и относительных погрешностей восстановления и оценка их основных статистических параметров (математическое ожидание, среднеквадратичное отклонение, коэффициент асимметрии и коэффициент эксцесса). Кроме того, оценивалось влияние на погрешность восстановления как объема выборки исходной последовательности данных, так и числа уровней квантования кодирующей функции.

Результаты исследования показали, что при малых размерах выборки (до 256 отсчетов) и малом числе уровней квантования (до 8) относительная погрешность восстановления имеет весьма существенные величины и составляет десятки процентов, что вполне предсказуемо. Тем не менее, при увеличении числа уровней квантования до 16 погрешность восстановления существенно снижается, и ее величина составляет единицы процентов.

При больших размерах выборки (более 256 отсчетов) влияние числа уровней квантования на погрешность также значительно, но можно отметить более резкий спад величины погрешности. Так, для выборки размером 2048 отсчетов математическое ожидание погрешности восстановления при числе уровней квантования 4 составляет 32,2 %, а при числе уровней квантования 16 – уже 1,1 %, что вполне приемлемо для некритичных измерений.

Следует отметить, что для обеспечения требуемого уровня погрешности восстановления и стационарности поведения ее статистических параметров для малых объемов выборок следует выбирать большее число уровней квантования, чем для больших выборок, для которых число уровней квантования можно достаточно серьезно ограничить.

Таким образом, проведенные исследования показали принципиальную пригодность предлагаемого способа для задач кодирования измерительных данных. Дальнейшее исследование способа следует связать с решением задачи выбора вида кодирующей функции, необходимой для получения заданных статистических характеристик закодированные последовательности данных.

Библиографические ссылки

1. *Ватолин Д., Ратушняк А., Смирнов М., Юкин В.* Методы сжатия данных. Устройство архиваторов, сжатие изображений и видео / Д. Ватолин, А. Ратушняк, М. Смирнов, В. Юкин. – М.: ДИАЛОГ-МИФИ, 2003.
2. *Левенец А. В., Нильга В. В., Чье Ен Ун.* Алгоритм структурного упорядочивания измерительных данных // Информатика и системы управления. – 2010. – № 3 (25).