УДК 621.391

Е.Г. ЖИЛЯКОВ, д-р. техн. наук, проф., зав. каф. БелГУ (г. Белгород, Россия),
Е.И. ПРОХОРЕНКО, канд. техн. наук, доц. БелГУ (г. Белгород, Россия),
А.В. БОЛДЫШЕВ, аспирант БелГУ (г. Белгород, Россия),
А.В. ЭСАУЛЕНКО, БелГУ (г. Белгород, Россия)

СЖАТИЕ РЕЧЕВЫХ ДАННЫХ КАК СРЕДСТВО ОБЕСПЕЧЕНИЯ СКРЫТНОСТИ РЕЧЕВЫХ СООБЩЕНИЙ

Рассмотрена информационная технология сжатия речевых данных, реализующая обнаружение и кодирование пауз, оптимальное субполосное преобразование и квантование по уровню, с сохранением достаточно высокой степени разборчивости и узнаваемости автора. Также рассмотрена возможность использования применяемых методов сжатия речевых данных как средств обеспечения скрытности речевых сообщений.

Ключевые слова: скрытность речевых сообщений, сжатие речевых данных, оптимальное субполосное преобразование, узнаваемость автора.

Постановка проблемы. В настоящее время в области информационно-телекоммуникационных систем большое внимание уделяется задаче сокращения объема битовых представлений (сжатия) речевых данных. Актуальность этой задачи обусловлена огромной ролью информационного обмена в современном обществе, существенную часть которого составляют речевые данные. Остро встает необходимость в использовании процедур сжатия для компактного хранения данных речевого обмена на жестких носителях, например, при проведении аудиоконференций (протоколирование различных заседаний), которые могут продолжаться длительное время, хранении звукозаписей выступлений лекторов, так же для систем информирования (звукового оповещения) в аэропортах, на авто и ж/д вокзалах и т.д.

При осуществлении сжатия речевых данных сигнал на жестком носителе может храниться в виде некого блока данных. В нем может храниться как преобразованная форма исходного сигнала, например, выходные последовательности КИХ-фильтров (если используется субполосное преобразование), квантованные значения сигнала (если используется квантование по уровню), так и информация о параметрах преобразований – параметры используемого квантователя (количество разрядов квантователя, максимальное по модулю значение вектора исходной последовательности, знаки отсчетов исходной последовательности), или же данные о субполосном преобразовании (количество интервалов, на которые разбивается ось нормированных частот, длительность интервала анализа в отсчетах). Если используется предварительное оценивание информативности сигнала (разделение сигнала на активную речь и паузу), то фрагменты сигнала,

классифицируемые как активная речь, кодируются каким-либо из известных алгоритмов, фрагменты же, классифицированные как паузы анализируются и сами не передаются, а передается минимальная информация о расположении этих фрагментов. На рис. 1 представлен один из вариантов блока кодированных данных:

Квантованные значения исходного речевого сигнала	Информация о расположении пауз в исходном речевом сигнале	Параметры квантователя

Рис. 1. Структура блока кодированных данных

Сигнал, хранимый на жестком носителе в такой форме невозможно воспроизвести и прослушать с помощью стандартных и широко распространенных средств. Для того чтобы воспроизвести его необходимо либо использовать специализированный кодер/декодер, либо вручную произвести декодирование этих данных, но при этом необходимо обладать сведениями о том, с помощью каких методов и алгоритмов было осуществлено кодирование.

Все это ограничивает непосредственный доступ к данным, и можно говорить о том, что сжатие речевых данных может стать средством решения такой немало важной проблемы при хранении и передаче речевых данных, как обеспечение конфиденциальности сообщаемых сведений, которая традиционно решается с использованием современных, хорошо себя зарекомендовавших криптографических методов защиты информации.

Анализ литературы. Предлагаемые во многих источниках, например [1 – 6], решения задачи сжатия порой существенно снижают объем передаваемых данных, однако, как правило, это достигается значительным усложнением аппаратной реализации устройств кодирования и восстановления речевых сигналов, требующих применения высокопроизводительных сигнальных процессоров.

При решении задачи сжатия речевых данных отмечаются два основных момента: необходимость удаления пауз, возникающих между отдельными словами и в режиме диалога, занимающих до 60% длительности исходных звукозаписей, и сокращение объема битовых представлений собственно звуковых данных [4, 5, 7].

Для сжатия участков собственно звуков речи тоже разработаны различные процедуры обработки. Основой этих процедур служат необратимые преобразования исходных данных либо за счет более грубого квантования по уровню, либо путем построения моделей генерации, позволяющих осуществить их воспроизведение [3, 8].

Существующие методы сжатия звуковых данных с использованием грубого квантования по уровню основываются на психоакустической модели,

что приводит к необходимости применения так называемых субполосных преобразований отрезков (векторов) отсчетов речевых сигналов, позволяющих получить другие векторы, подвекторы которых отражают частотные свойства исходного вектора в выбранных диапазонах оси частот. Именно компоненты этих подвекторов подвергаются квантованию по уровню с различными шагами, чем достигается учет частотно-избирательных свойств человеческого слуха. Но следует помнить, что при использовании различных методов и алгоритмов сжатия необходимо обращать внимание на сохранение качества речевого сигнала на выходе системы передачи информации, которое определяется такими показателями как разборчивость речи и сохранение тембра речи, обеспечивающего узнаваемость голоса.

Целью статьи является анализ возможности использования предлагаемых методов сжатия речевых данных как средства обеспечения скрытности речевых сообщений при их передаче или хранении.

Субполосное преобразование. В основе предлагаемой процедуры сжатия речевых данных используется новый вариационный метод оптимального субполосного преобразования, подробно описанный в [9].

Сущность субполосного преобразования заключается в следующем: для отрезков сигнала вычисляется вектор $\vec{yy} = (\mathbf{r}_{y_1}, \mathbf{r}_{y_2}, ..., \mathbf{r}_{y_R})$, состоящий из подвекторов $\vec{y}_r = (y_{1r}, y_{2r}, ..., y_{Jr})$, которые отражают частотные свойства исходного сигнала в некотором частотном интервале (в данном случае ось частот разбивается на R частотных интервалов):

$$\vec{yy} = AA \, \mathbf{r}_x, \tag{1}$$

где $\mathbf{r}_x$ – вектор исходного отрезка сигнала длиной N; AA – блочная матрица.

Блочная матрица AA формируется на основе субполосной матрицы $A_r = \{a_{ik}^r\}$ и имеет вид

$$AA = \begin{pmatrix} \sqrt{L_1^1}(Q_1^1)^T \\ \sqrt{L_1^2}(Q_1^2)^T \\ \mathbf{L} \\ \sqrt{L_1^R}(Q_1^R)^T \end{pmatrix}, \tag{2}$$

где $Q_1^r = (\vec{q}_1, ..., \vec{q}_J)$ – подматрица собственных векторов матрицы A_r, $L_1^r = diag(\vec{\lambda}_1, ..., \vec{\lambda}_J)$ – подматрица собственных чисел матрицы A_r.

Субполосная матрица A_r имеет элементы вида (3)

$$a_{ik}^r = \begin{cases} \dfrac{\sin[\upsilon_2^r(i-k)] - \sin[\upsilon_1^r(i-k)]}{\pi(i-k)}, & i \neq k, \\ \dfrac{\upsilon_2^r - \upsilon_1^r}{\pi}, & i = k, \end{cases} \tag{3}$$

где υ_1 и υ_2 определяются исходя из разбиения области определения спектра $[-\pi, \pi]$ на ряд равновеликих частотных интервалов

$$V^r = [-\upsilon_2^r, -\upsilon_1^r) \cup [\upsilon_2^r, \upsilon_1^r) \tag{4}$$

таких, что $\upsilon_2^r - \upsilon_1^r = \Delta\upsilon = \text{const}$.

Матрица A_r обладает тем свойством, что значения ее собственных чисел с номерами меньшими $m = 2[N/2R]$ при упорядочивании по возрастанию близки к единице, а с номерами большими $J = 2[N/2R] + 4$ стремятся к нулю (квадратная скобка означает операцию взятия целой части содержимого).

Значения энергии сигнала в заданном частотном интервале вычисляются с использованием полученных подвекторов субполосного преобразования:

$$P_r = \sum_{i=1}^{J} (y_{ir})^2, \ r = 1, 2, ..., R. \tag{5}$$

Также имеет место обратное субполосное преобразование:

$$\hat{\vec{x}} = AA'\overrightarrow{yy}. \tag{6}$$

Кодирование пауз. Предлагаемый метод основан на учете отличий в распределении энергетических составляющих звуков речи и сигнала паузы в частотной области [4, 6]. Формулируется следующая гипотеза:

H_0: энергия исходного отрезка $\vec{x}_i$, $i = 1, 2, ..., N$ в r-м частотном интервале $(\nu_{2r} - \nu_{1r})$ обусловлена внешними шумами.

Положим

$$S_r^i = \frac{P_r^i}{P_r^{\Pi}}, \ r = 1, 2, ..., R, \tag{7}$$

где P_r^{Π} – доля энергии паузы, P_r^i – доля энергии отрезка сигнала (энергия вычисляется с использованием выражения (5)).

Для увеличения вероятности правильного обнаружения границы пауза/звук целесообразно с порогом сравнивать максимальное значение из

отношений вида (7), так как энергия сигнала, соответствующего звуку может быть сосредоточена в сравнительно узком диапазоне частот.

Если имеет место

$$\max(S_r^i) \geq h, \tag{8}$$

то отвергается нулевая гипотеза H_0, т.е. отрезок речевого сигнала принимается за звук.

Если же выполняется неравенство

$$\max(S_r^i) \leq h, \tag{9}$$

то H_0 считается справедливой и данный отрезок речевого сигнала принимается за паузу.

В выражениях (8) и (9) h – это порог, который определяется адаптивно. Причем, отрезки сигнала, на которых величина решающей функции не превышала установленный порог, как правило, являются паузами малой длительности между фонемами или слитно произнесенными словами.

Информация об удаленных паузах хранится в "карте пауз", которая содержит сведения о начале паузы и ее длительности, она используется при декодировании сигнала для восстановления паузы. Причем, имеется возможность выбора параметров восстановления – длительность пауз, тип заполнения пауз ("тишина" или комфортный шум).

На рис. 2 представлена функциональная схема системы сжатия речевых данных на основе кодирования пауз, оптимального субполосного преобразования и квантования.

При поступлении сигнала на вход системы первоначальным этапом сжатия является обнаружение и кодирование пауз, при этом возможно оперировать следующими параметрами: N – длительность отрезков сигнала, на которые разбивается исходный сигнал, R – количество частотных интервалов, на которые разбивается отрезок сигнала. На выходе блока обнаружения пауз формируется последовательность речевых данных, не содержащая пауз, а также служебная информация ("карта пауз") о том, где паузы находятся и какой они длительности: $N_{\text{нач}}$ – номер начального отсчета и N_n – длина, т.е. количество отсчетов для каждой паузы. Применение данного метода обнаружения и кодирования пауз позволяет сократить объем данных в 2 раза, при сохранении разборчивости и узнаваемости говорящего.

Затем этот сигнал подвергается субполосному преобразованию, исходными параметрами для этого блока служат: N – длительность отрезков сигнала, на которые разбивается сигнал "без пауз", R – количество частотных интервалов, на которые разбивается отрезок сигнала, а также блочная матрица AA. Сигнал теперь представляется в виде набора векторов субполосного преобразования, и выделить необходимую информацию без параметров матрицы AA представляется невозможным.

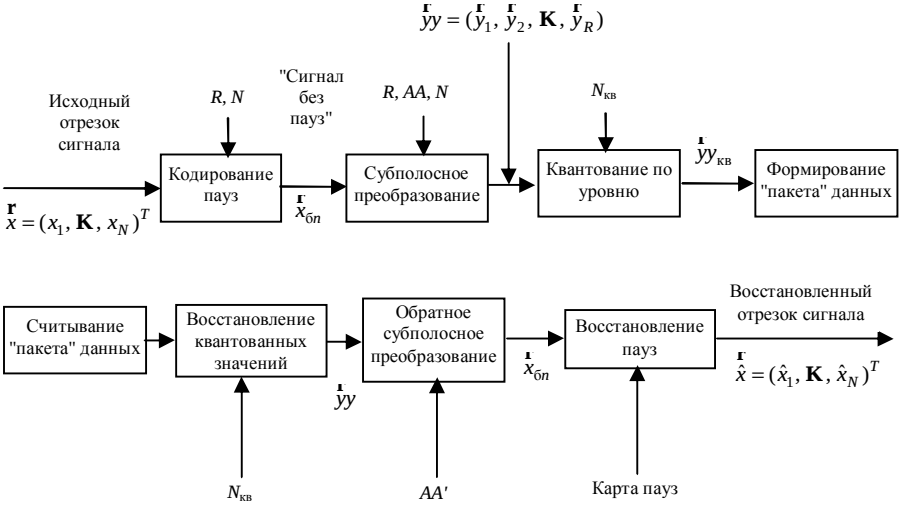

Рис. 2. Функциональная схема системы сжатия/восстановления речевых данных

В ряде работ [10, 11] описывалась возможность удаления подвекторов субполосного преобразования, к которым относятся малоэнергетические составляющие спектра речевого сигнала, без существенной потери разборчивости и качества воспроизведения, что существенно сокращает объем речевых данных. Для увеличения степени закрытия данных, можно также осуществить перестановку подвекторов субполосного преобразования, а информацию об их позициях кодировать и передавать вместе со служебной информацией.

Следующим этапом является квантование по уровню, полученных на предыдущем этапе, значений векторов субполосного преобразования, при этом главным параметром является количество уровней квантования, которое определяется числом разрядов m:

$$N_{кв} = 2^m - 1. \qquad (10)$$

Чем больше $N_{кв}$, тем на большее число ступеней разбивается шкала квантователя и тем с большей точностью воспроизводится исходная последовательность при восстановлении. Величина шага квантования Δ определяется максимальным из абсолютных значений вектора квантуемой последовательности. При квантовании используется способ округления к ближайшему двоичному уровню с порогом округления 0,5 (рис. 3).

При пороге округления 0,5 квантованный сигнал симметричен относительно оси времени, в нем присутствуют только нечетные гармоники искажений. Этот вариант квантования наиболее распространен.

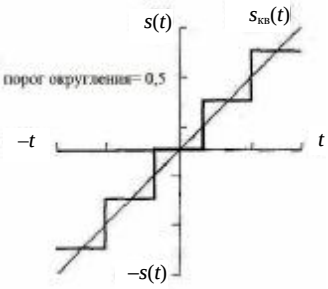

Рис. 3. Характеристика квантователя

Квантованная последовательность на выходе квантователя $s_{i\ кв}$ имеет вид:

$$s_{i\ кв} = [|y_i| / \Delta + 0{,}5], \tag{11}$$

где шаг квантования $\Delta = 1/N_{кв}$.

Квантование осуществляется для каждой субполосы, полученной на предыдущем этапе, отдельно, т.е. шаг квантования вычисляется для каждого подвектора $\vec{y}$ исходной последовательности $\vec{yy}$.

Для осуществления квантования каждого из подвекторов выполняется следующая последовательность действий:

1. Определяются знаки отсчетов исходной последовательности $zn_i = sign(y_i)$.

2. Определяется максимальное по модулю значение подвектора исходной последовательности $y^r_{\max} = \max(|\vec{y}|)$ соответствующего r-го частотного интервала.

3. Вычисляется шаг квантования для данного подвектора $\Delta^r = \frac{1}{N_{кв}}$, где количество уровней квантования $N_{кв}$ определяется заданным числом разрядов m (10).

4. Вычисляется квантованная последовательность для данного подвектора как

$$s_{i\ кв} = [|y^r_i| / (y^r_{\max}\Delta^r) + 0{,}5]. \tag{12}$$

Таким образом, на выходе квантователя формируется массив данных, содержащий: N бит значений знаков отсчетов исходной последовательности

$\vec{yy}$; R максимальных значений подвекторов $\vec{y}$; mN бит значений квантованной последовательности $s_{i\,\text{кв}}$ и значение, определяющее количество разрядов квантования m. При таком алгоритме квантования имеется возможность устанавливать любое число разрядов квантования m, в зависимости от необходимой степени сжатия и требуемого качества воспроизведения восстановленного сигнала.

В совокупности все предложенные методы позволят достичь высокой степени сжатия при сохранении требуемого качества речи для различных прикладных областей применения.

На рис. 4 представлена структура блока кодированных данных, полученных при использовании предлагаемых методов.

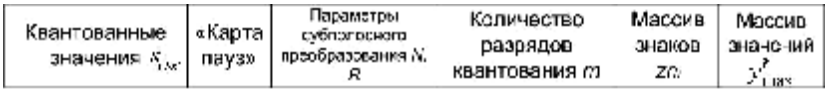

Рис. 4. Структура блока кодированных данных, полученных при использовании предлагаемых методов

Что же касается скрытности данных, полученных в результате сжатия предложенными методами, то, чтобы злоумышленник при перехвате сообщения смог выделить из нее необходимые данные, ему необходимо знать структуру сформированного блока (порядок записи информации в нем) и обладать сведениями о примененных методах и алгоритмах сжатия.

При осуществлении декодирования данных на попытку подобрать все параметры для выполненных преобразований, а именно параметры квантователя, субполосного преобразования, порядок следования подвекторов субполосного преобразования (если используется их перемешивание), а также место и длительность пауз в сообщении, будет затрачено длительное время. В частности, что касается субполосного преобразования, то без точного представления о математическом аппарате, используемом в данном методе, злоумышленник не сможет восстановить компоненты вектора в речевой сигнал.

Выводы. При использовании различных методов и алгоритмов сжатия необходимо обращать внимание на сохранение качества речевого сигнала на выходе системы передачи информации, которое определяется разборчивостью речи и сохранением тембра речи, обеспечивающего узнаваемость голоса.

Предлагаемый комплексный подход к сжатию речевых данных позволяет добиться не только высокого коэффициента сжатия, при сохранении достаточно высокой степени разборчивости восстановленного сообщения, но и не позволяет произвести декодирование за короткий промежуток времени без наличия полных сведений о примененных методах сжатия. Следовательно,

предлагаемая процедура сжатия речевых данных, без дополнительных затрат, может обеспечить некоторый уровень конфиденциальности информации.

Список литературы: **1.** *Бухвинер В.Е.* Управляемое компандирование звуковых сигналов / *В.Е. Бухвинер.* – М.: Связь, 1978. – 208 с. **2.** *Козленко Н.И.* Помехоустойчивость дискретной передачи непрерывных сообщений / *Н.И. Козленко.* – М.: Радиотехника, 2003. – 352 с. **3.** *Шульгин В.И.* Основы теории связи. Часть 1. Теория и практика кодирования. Учебное пособие, Харьков: "ХАИ", 2005. **4.** *Жиляков Е.Г.* О сжатии речевых сигналов / *Е.Г. Жиляков, С.П. Белов* // Вестник НТУ "Харьковский политехнический институт". – Харьков, 2005. – № 56. – С. 32-40. **5.** *Сергиенко А.Б.* Цифровая обработка сигналов / *А.Б. Сергиенко* – СПб.: Питер, 2005. **6.** *Шелухин О.И.* Цифровая обработка и передача речи / *О.И.Шелухин, Н.Ф. Лукьянцев.* – М.: Радио и связь, 2000. **7.** *Жиляков Е.Г.* Метод обнаружения пауз в речевых сигналах / *Е.Г. Жиляков, С.П. Белов, Е.И. Прохоренко* // Системы синхронизации, формирования и обработки сигналов для связи и вещания. Материалы научно-технического семинара. – Белгород, 2006. – С. 94-98. **8.** Применение цифровой обработки сигналов. Под ред. Э. Оппенгейма. – М.: Мир, 1980. – 550 с. **9.** *Жиляков Е.Г.* Частотный анализ речевых сигналов // Научные ведомости Белгородского государственного университета. – Белгород, 2006. – № 2 (31). – Вып. 3. – С. 201-208. **10.** *Прохоренко Е.И.* Цифровое кодирование клиппированной речи с сохранением разборчивости и узнаваемости диктора / *Е.И. Прохоренко, И.А. Сидоренко, А.В. Болдышев* // Научные ведомости БелГУ. Серия: информатика и прикладная математика. Белгород, 2008. **11.** *Прохоренко Е.И.* Цифровое кодирование речевых данных на основе клиппирования и частотных представлений / *Е.И. Прохоренко, И.А. Сидоренко, А.В. Болдышев //* Вестник НТУ "Харьковский политехнический институт". – Харьков: НТУ "ХПИ", 2008. – № 49. – С. 184 – 189.

УДК 621.391

Стиснення мовних даних як засіб забезпечення скритності мовних повідомлень / Жиляков Е.Г., Прохоренко Е.І., Болдишев А.В., Есауленко А.В. // Вісник НТУ "ХПІ". Тематичний випуск: Інформатика і моделювання. – Харків: НТУ "ХПІ". – 2009. – № 43. – С. 75 – 83.

Розглянута інформаційна технологія стиснення мовних даних, що реалізовує виявлення і кодування пауз, оптимальне субсмугове перетворення і квантування по рівню, із збереженням достатньо високого ступеня розбірливості і впізнанності автора. Також розглянута можливість використання вживаних методів стиснення мовних даних, як засобів забезпечення скритності мовних повідомлень. Іл.: 4. Бібліогр.: 11 назв.

Ключові слова: скритність мовних повідомлень, стиснення мовних даних, оптимальне субсмугове перетворення, впізнанність автору.

UDC 621.391

Compression of vocal data as backer-up secrecy of vocal reports / Zhilyakov E.G., Prokhorenko E.I., Boldyshev A.V., Esaulenko A.V. // Herald of the National Technical University "KhPI". Subject issue: Information Science and Modelling. – Kharkov: NTU "KhPI". – 2009. – №. 43. – P. 75 – 83.

Information technology of compression of vocal data, realizing a discovery and encoding of pauses, optimum subbar transformation and quantum on a level, is considered, with a maintainance there is an enough high degree of legibility and knowableness of author. Possibility of the use of the applied methods of compression of vocal data is also considered, as backer-ups secrecy of vocal reports. Figs: 4. Refs: 11 titles.

Key words: secrecy of vocal reports, compression of vocal data, optimum subband transformation, knowableness of author.

Поступила в редакцию 09.10.2009