

Е.И. ПРОХОРЕНКО, канд. техн. наук, БелГУ (г. Белгород),
И.А. СИДОРЕНКО, канд. техн. наук, БелГУ (г. Белгород),
А.В. БОЛДЫШЕВ, БелГУ (г. Белгород)

ЦИФРОВОЕ КОДИРОВАНИЕ РЕЧЕВЫХ ДАННЫХ НА ОСНОВЕ КЛИППИРОВАНИЯ И ЧАСТОТНЫХ ПРЕДСТАВЛЕНИЙ

Рассмотрена новая технология сжатия речевого сигнала для цифровых систем передачи информации, основанная на клиппировании речи с сохранением разборчивости речи и узнаваемости голоса. Предложены алгоритмы обработки цифрового речевого сигнала, реализующие обнаружение и удаление шума в паузах клиппированной речи, а также формирование данных об изменении мгновенного спектра речи на основе частотных представлений речевого сигнала.

Ключевые слова: сжатие речевого сигнала, клиппирование, разборчивость речи, узнаваемость голоса, частотное представление речевого сигнала.

Постановка проблемы. Построение эффективных цифровых систем передачи речевых сигналов определяется возможностью снижения затрат как на практическую реализацию алгоритмов обработки речевых данных, так и на сам процесс передачи данных, себестоимость которого пропорциональна объему передаваемых данных. Поэтому, например, операторы сетей мобильной связи заинтересованы в разработке алгоритмов сжатия речи, существенно уменьшающих объем битового представления речевых сигналов с сохранением их разборчивости и узнаваемости. Вот почему проблема сокращения объема данных при передаче речевых сигналов в цифровых системах передачи информации остается, по-прежнему, одной из самых исследуемых. Поиск путей возможного решения проблемы ведется в различных плоскостях и направлен на поиск компромисса между степенью сжатия речевых данных, с одной стороны, и сложностью алгоритма кодирования и обработки речи, с другой стороны. Непременным условием является сохранение качества речевого сигнала на выходе системы передачи информации, определяемое такими показателями как разборчивость речи и сохранение тембра речи, обеспечивающего узнаваемость голоса.

Анализ литературы. Анализ публикаций по проблеме эффективного сжатия речевых данных, например [1 – 9], позволяет определить следующие основные направления исследований:

- разработка эффективных методов компрессии данных, полученных после аналого-цифрового преобразования ограниченного по спектру речевого сигнала (амплитудное компандирование речи, удаление пауз и т.п.);
- поиск новых способов представления речи, основанных на передаче информации не о самой речи, а об её структурных элементах, таких как фонемы или форманты, либо об её спектре (частотное представление речи);

– имитация речевых сигналов посредством конечного набора реализаций шума или случайных сигналов, составляющих "словарь" для последующего синтеза речи в вокодерах;

– использование различных механизмов предсказания на основе математического аппарата Марковских цепей и т.п.

Предлагаемые в указанных источниках решения 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. Вместе с тем, цифровые технологии обработки речи позволяют по новому подойти к реализации механизмов уменьшения избыточности речевых сигналов, известных ещё с "аналоговых" времён.

Одним из наиболее эффективных способов уменьшения избыточности речи по праву считается клиппирование речевого сигнала. С точки зрения цифровых технологий клиппированный речевой сигнал может быть передан одноканальной (однобитовой) последовательностью данных, несущих информацию только о полярности дискретных отсчетов сигнала. Такой сигнал сохраняет достаточную для практики словесную и фразовую разборчивость, вполне приемлемую в тех системах передачи информации, где тембровой окраской голоса (узнаваемость абонента) можно пожертвовать. Этот широко известный факт не находит практического применения именно из-за потери узнаваемости голоса, вызванного сильным зашумлением речи при её предельном амплитудном ограничении. В аналоговую эпоху передачи речи для устранения этого недостатка предлагалось организовывать дополнительный канал, по которому тем или иным способом должна была передаваться информация о поведении огибающей исходного речевого сигнала [1, 2]. На приемной стороне после объединения информации обоих каналов мог быть восстановлен исходный речевой сигнал, достаточного для практического применения качества. Сложность осуществления этой процедуры с использованием аналоговой схемотехники оказалась главной причиной, по которой указанный метод передачи клиппированной речи не нашел практического применения. Другой, не менее сложной, оказалась проблема подавления шума в паузах речи, который резко снижал разборчивость и утомлял слух абонента телефонной сети.

Вместе с тем, обе эти проблемы могут быть успешно решены с применением цифровых технологий обработки речи. Для этого необходимо очистить речь от шума в паузах и в общем потоке данных передать сам клиппированный речевой сигнал и информацию об изменении огибающей сигнала, либо его мгновенного спектра.

Цель статьи – разработка новой технологии сжатия речевых данных, основанной на передаче очищенной от шума в паузах клиппированной речи и

дополнительной информации об изменении её мгновенного спектра и обеспечивающей сохранение разборчивости речи и узнаваемость голоса.

Технология кодирования речи основана на методе субполосного преобразования.

Согласно акустической теории речеобразования восприятие звуков определяется областями максимальной концентрации энергии – формантами, т.е. можно сказать, что 95% энергии речевого сигнала сосредоточены в сравнительно узком частотном диапазоне и именно эта составляющая спектра определяет степень разборчивости при воспроизведении речи. В некоторых случаях для достижения большой степени сжатия речевых данных можно пренебречь малоэнергетическими составляющими сигнала, которые определяют эмоциональную окраску речи и узнаваемость голоса диктора.

Таким образом, для решения проблем сжатия возникает необходимость избирательно преобразовывать спектральные компоненты из различных частотных диапазонов. В настоящее время все большее распространение получает подход на основе, так называемого, субполосного кодирования, который реализуется с использованием банков КИХ-фильтров, путем свертки сигнала с несколькими полосовыми фильтрами и децимацией результата.

Такой подход имеет определенные недостатки, поскольку проявляется эффект наложения спектров ("элайзинга"), возникающий при децимации, то есть относительная среднеквадратичная погрешность аппроксимации спектров исходного вектора в соответствующем частотном интервале, в основном, обусловлена малым количеством используемых прореженных данных, что, согласно теории Найквиста, приводит к элайзингу, а также зависит от степени подавления спектральных компонент в полосах непропускания и ширины переходных полос КИХ-фильтров.

В данной процедуре сжатия речевых данных используется новый вариационный метод субполосного преобразования, оптимальный с точки зрения минимума среднеквадратичной погрешности аппроксимации спектров исходного вектора в соответствующем частотном интервале.

Сущность субполосного преобразования заключается в следующем: для

отрезков сигнала вычисляется вектор $\vec{y} = (\vec{y}_1, \vec{y}_2, \dots, \vec{y}_R)$, состоящий из подвекторов $\vec{y}_r = (y_{1r}, y_{2r}, \dots, y_{Jr})$, которые отражают частотные свойства исходного сигнала в некотором частотном интервале (в данном случае ось частот разбивается на R частотных интервалов):

$$\vec{y} = AA\vec{x}, \quad (1)$$

где AA – блочная матрица; \vec{x} – вектор исходного отрезка сигнала длиной N .

Блочная матрица формируется на основе субполосной матрицы $A_r = \{a_{ik}^r\}$ с элементами вида (2)

$$a_{ik}^r = \begin{cases} \frac{\sin[\nu_2^r(i-k)] - \sin[\nu_1^r(i-k)]}{\pi(i-k)}, & i \neq k, \\ \frac{\nu_2^r - \nu_1^r}{\pi}, & i = k, \end{cases} \quad (2)$$

где ν_1^r и ν_2^r определяются исходя из разбиения области определения спектра $[-\pi, \pi]$ на ряд равновеликих частотных интервалов:

$$V^r = [-\nu_2^r, -\nu_1^r) \cup [\nu_2^r, \nu_1^r), \quad (3)$$

таких, что $\nu_2^r - \nu_1^r = \Delta\nu = \text{const}$.

Матрица A обладает тем свойством, что значения ее собственных чисел с номерами меньшими $m = 2 \left\lfloor \frac{N}{2R} \right\rfloor$ при упорядочивании по возрастанию близки к единице, а с номерами большими $J = 2 \left\lceil \frac{N}{2R} \right\rceil + 4$ стремятся к нулю (квадратная скобка означает операцию взятия целой части содержимого) [3,4].

Таким образом, существует возможность создания блочной матрицы:

$$AA = \begin{pmatrix} \sqrt{L_1^1} (Q_1^1)^T \\ \sqrt{L_1^2} (Q_1^2)^T \\ \dots \\ \sqrt{L_1^R} (Q_1^R)^T \end{pmatrix}, \quad (4)$$

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

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

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

Данный метод частотного анализа подробно рассмотрен в [4].

Для отрезков сигнала осуществляется вычисление распределения энергии по частотному интервалу и упорядочивание полученных значений по убыванию, далее определяются номера интервалов, суммарная энергия в которых составляет 95% от общей.

Для данных интервалов вычисляются коэффициенты:

$$B_1^r = \sqrt{\frac{P_r}{P_{\max}}}, \quad (6)$$

где P_r – значение энергии в r -м частотном интервале исходного сигнала соответственно; P_{\max} – максимальное значение энергии исходного сигнала.

Для исходного сигнала осуществляется клиппирование с последующим сохранением полученных значений и вычисленных ранее коэффициентов.

При восстановлении для клиппированного сигнала вычисляются значения энергии (5) и коэффициенты:

$$B_2^r = B_1^r \sqrt{\frac{P_{\text{кл max}}}{P_{\text{клr}}}}, \quad (7)$$

где $P_{\text{клr}}$ – значение энергии в r -м частотном интервале клиппированного сигнала соответственно; $P_{\text{кл max}}$ – максимальное значение энергии клиппированного сигнала.

Значения коэффициентов B_2 умножаются на значение вектора субполосного преобразования для соответствующего частотного интервала

$$\hat{y} = \vec{y} B_{2r}. \quad (8)$$

Значения подвекторов субполосного преобразования \vec{y}_r , соответствующие интервалам, суммарная энергия которых составляет 5% от общей обнуляются. Далее осуществляется обратное субполосное преобразование:

$$\hat{x} = A A' \vec{y}. \quad (9)$$

Для проверки работоспособности алгоритма были проведены вычислительные эксперименты для большого количества сигналов. Процедура сжатия осуществлялась в 3 этапа: кодирование пауз [5, 6], клиппирование с вычислением коэффициентов (6), арифметическое кодирование. Полученные после процедуры сжатия данные сохранялись на жестком носителе, затем рассчитывались коэффициенты сжатия по формулам (10) и (11):

$$K1 = \frac{L1}{L2}, \quad (10)$$

$$K2 = \frac{V1}{V2}, \quad (11)$$

где $L1$ – длина исходного сигнала в отсчетах; $L2$ – длина сигнала с предварительно удаленными паузами, в отсчетах; $V1$ – объем исходного файла; $V2$ – объем файла, полученного в результате сжатия.

Типичные примеры результатов экспериментов представлены в табл.

Таблица

Тип сигнала	Параметры сигнала	Длина исходного сигнала (L1 отчетов)	Длина сигнала с предварительно удаленными паузами (L2 отчетов)	Объем исходного wav файла, Кб (V1)	Коэф. сжатия после удаления пауз (K1)	Объем файла, полученного в результате сжатия, Кб (V2)	Коэф. сжатия (K2)
диалог	$Fd = 8\text{кГц}$, количество бит на 1 символ равно 8	145125	92672	283	1.57	13.2	13.9
стих		82944	66650	162	1.25	9.45	17.14
слитная фраза		35327	19006	69	1.86	2.72	25.37
лекция		376426	164352	735	2.29	23.4	31.4

Таким образом, на основе приведенных данных можно утверждать, что данный подход к сжатию данных позволяет существенно сократить объем речевых данных при их хранении на жестких носителях информации, при сохранении относительно высокого качества воспроизведения восстановленного сигнала.

Выводы. В результате проделанной работы была предложена новая технология сжатия речевых данных, основанная на новом вариационном методе субполосного преобразования, оптимального с точки зрения минимума среднеквадратичной погрешности аппроксимации спектров исходного вектора в соответствующем частотном интервале.

Список литературы: 1. Бухвинер В.Е. Управляемое компандирование звуковых сигналов. – М.: Связь, 1978. – 208 с. 2. Козленко Н.И. Помехоустойчивость дискретной передачи непрерывных сообщений. – М.: Радиотехника, 2003. – 352 с. 3. Гантмахер Ф.Р. Теория матриц. – М.: Физматлит, 2004. – 560 с. 4. Жилияков Е.Г. Частотный анализ речевых сигналов // Научные ведомости Белгородского государственного университета. – 2006. – № 2 (31). – Вып. 3. – С. 201–208. 5. Прохоренко Е.И. Метод обнаружения пауз в речевых сигналах // Вестник московской академии рынка труда и информационных технологий. – 2006. – № 4 (26). – С. 13–20. 6. Жилияков Е.Г., Белов С.П. О сжатии речевых сигналов // Вестник Национального технического университета "Харьковский политехнический институт". Тематический выпуск: Информатика и моделирование. – Х.: НТУ "ХПИ". – 2005. – № 56. – С. 32–40. 7. Сергиенко А.Б. Цифровая обработка сигналов. – СПб.: Питер, 2005. – 280 с. 8. Шелухин О.И., Лукьянцев Н.Ф. Цифровая обработка и передача речи. – М.: Радио и связь, 2000. – 320 с. 9. Орищенко В.И., Санников В.Г., Свириденко В.А. Сжатие данных в системах сбора и передачи информации / Под ред. В.А. Свириденко. – М.: Радио и связь, 1985. – 184 с.

УДК 621.391

Цифрове кодування мовних даних на основі кліппірування і частотних представлень
/Е. І. Прохоренко, І.А. Сидоренко, А.В. Болдишев // Вісник НТУ "ХПІ". Тематичний випуск:
Інформатика і моделювання. – Харків: НТУ "ХПІ". – 2008 – № 49. – С. 139 – 145.

Розглянута нова технологія стиснення мовного сигналу для цифрових систем передачі інформації, заснована на кліппірованій мові із збереженням розбірливості мови і впізнанності голосу. Запропоновані алгоритми обробки цифрового мовного сигналу, що реалізують виявлення і видалення шуму в паузах кліппірованої мови, а також формування даних про зміну миттєвого спектру мови на основі частотних представлень мовного сигналу. Табл.: 1. Бібліогр.: 9 назв.

Ключові слова: стиснення мовного сигналу, кліппірування, розбірливість мови, впізнанність голосу, частотне представлення мовного сигналу.

UDC 621.391

Digital coding of speech data on the basis of speech clipping and frequency representations
/E.I. Prokhorenko, I.A. Sidorenko, Boldyshev A.V.// Herald of the National Technical University
"KhPI". Subject issue: Information Science and Modelling. – Kharkiv: NTU "KhPI". – 2008. – № 49. –
P. 139 – 145.

The new technology of compression of a speech signal for digital systems of transfer of the information, based on clipping speeches with preservation of legibility speech and recognition voice is considered. Algorithms of processing of the digital speech signal, realizing detection and removal of noise in pauses of clipped speech, and also formation of data about change of an instant spectrum of speech on the basis of frequency representations of a speech signal. Tabl.: 1. Bibliogr.: 9 titles

Keywords: compression of a speech signal, legibility of speech, frequency presentation of speech, clipping of speech.

Поступила в редакцію 10.10.2008