

ЦИФРОВОЕ КОДИРОВАНИЕ КЛИППИРОВАННОЙ РЕЧИ С СОХРАНЕНИЕМ РАЗБОРЧИВОСТИ И УЗНАВАЕМОСТИ ДИКТОРА

Е.И. ПРОХОРЕНКО¹
И.А. СИДОРЕНКО²
А.В. БОЛДЫШЕВ³

1) *Белгородский государственный университет*
e-mail: prokhorenko@bsu.edu.ru

2) *Белгородский государственный университет*
e-mail: sidorenko@bsu.edu.ru

3) *Белгородский государственный университет*
e-mail: alexeiboldyshev@mail.ru

В статье рассматривается новая технология сжатия речевого сигнала для цифровых систем передачи информации, основанная на клиппировании речи с сохранением разборчивости и узнаваемости. 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 речи, разборчивость речи, узнаваемость голоса, удаление пауз речи, частотное представление речи, клиппирование, субполосное преобразование, вариационный метод.

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, определяемым такими показателями как разборчивость речи и сохранение тембра речи, обеспечивающего узнаваемость голоса.

Анализ публикаций по проблеме эффективного сжатия речевых данных, например [1,2,3], 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 о только 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]. 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.

В данной статье предлагается алгоритм сжатия речевых данных на основе клиппирования с использованием нового метода частотного анализа и оптимального субполосного преобразования.

Анализ распределения энергии по частотным диапазонам клиппированного сигнала показал значительное увеличение малоэнергетических составляющих спектра по сравнению с исходным сигналом (рис 1). На участке спектра, отмеченного на рисунке 1 как «диапазон 1» область концентрации исходного и клиппированного сигналов совпадают по расположению и мощности. А малоэнергетические составляющие («диапазон 2» и «диапазон 3») спектра клиппированного сигнала превосходят по мощности соответствующие составляющие спектра исходного сигнала.

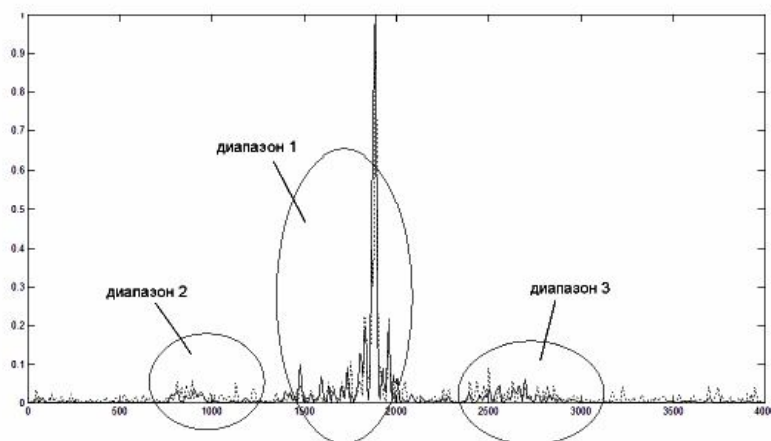


Рис 1. Распределение энергии по частотным интервалам исходного и клиппированного сигнала. Спектр исходного- сплошная линия и клиппированного сигнала- пунктирная линия



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 так же зависит от степени подавления спектральных компонент в полосах непропускания и ширины переходных полос КИХ-фильтров.

В предлагаемой 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)

$$a_{ik}^r = \begin{cases} \frac{\sin[v_2^r(i-k)] - \sin[v_1^r(i-k)]}{\pi(i-k)}, & i \neq k \\ \frac{v_2^r - v_1^r}{\pi}, & i = k \end{cases}, \quad (2)$$

v_1 и v_2 определяются исходя из разбиения области определения спектра $[-\pi, \pi]$ на ряд равновеликих частотных интервалов:

$$V^r = [-v_2^r, -v_1^r) \cup [v_2^r, v_1^r) \quad (3)$$

таких что $v_2^r - v_1^r = \Delta v = \text{const}$

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) [4,5].

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 А,

$L_1^r = \text{diag}(\vec{\lambda}_1, \dots, \vec{\lambda}_J)$ – подматрица собственных чисел матрицы А.

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

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

Данный метод частотного анализа подробно рассмотрен в [5].

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 г-ом 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:

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

где $P_{\text{кл}}$ - значение энергии в г-ом частотном интервале клиппированного сигнала соответственно.

$P_{\text{клmax}}$ - максимальное значение энергии клиппированного сигнала.

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

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

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

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

В результате спектр восстановленного сигнала становится близким к спектру исходного сигнала (рис 2). На участке спектра, отмеченного на рисунке 2 как «диапазон 1» области концентрации энергии восстановленного и исходного сигналов совпадают по расположению и мощности. Малоэнергетические составляющие («диапазон 2» и «диапазон 3») спектра восстановленного сигнала имеют примерно то же значение по мощности, что соответствующие составляющие спектра исходного сигнала.

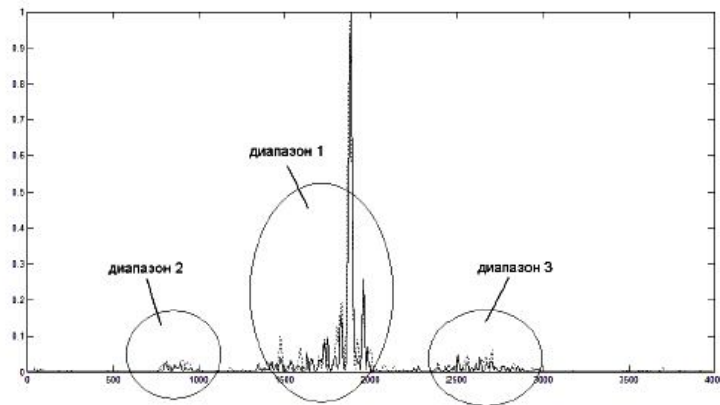


Рис 2. Распределение энергии по частотным интервалам исходного и восстановленного сигнала. Спектр исходного сигнала показан пунктирной линией, клипированного сигнала- сплошной линией

В предлагаемой процедуре сжатия осуществляется предварительное кодирование пауз сигнала по методу, рассмотренному в [6]. Кроме того, степень сжатия можно увеличить применив арифметическое кодирование к значениям клипированного сигнала [7].

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

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

где $L1$ –длина исходного сигнала в отсчетах,
 $L2$ - длина сигнала с предварительно удаленными паузами, в отсчетах.

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

где $V1$ -объем исходного wav,
 $V2$ - объем файла, полученного в результате сжатия.

Типичные примеры результатов экспериментов представлены в таблице 1.

Таблица 1

Тип сигнала	Параметры сигнала	Длина исходного сигнала ($L1$ отсчетов)	Длина сигнала с предварительно удаленными паузами ($L2$ отсчетов)	Объем исходного wav файла, Кб ($V1$)	Кэф. сжатия после удаления пауз ($K1$)	Объем файла полученного в результате сжатия, Кб ($V2$)	Кэф. сжатия ($K2$)
диалог стих	$Fd= 8кГц$, Количество бит на 1 символ= 8,	145125	92672	283	1.57	13.2	13.9
		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 данный подход к сжатию данных позволяет существенно сократить объем речевых данных при их хранении на жестких носителях информации, при сохранении высокого качества воспроизведения восстановленного сигнала.

Литература

1. В.Е. Бухвинер. Управляемое компандирование звуковых сигналов.- М.:Связь, 1978.208 с.
2. Н.И. Козленко. Помехоустойчивость дискретной передачи непрерывных сообщений.- М.: Радиотехника, 2003.- 352с.
3. Орищенко В.И. Сжатие данных в системах сбора и передачи информации. В.И.Орищенко, В.Г.Санников. В.А. Свириденко; Под ред. В.А. Свириденко.- М.: Радио и связь, 1985.-184с., ил
4. Ф.Р. Гантмахер. Теория матриц.- 5-е изд.-М.: ФИЗМАТЛИТ, 2004.-560 с.
5. Е.Г. Жилияков. Частотный анализ речевых сигналов. Научные ведомости Белгородского государственного университета. Белгород, 2006. – №2(31), выпуск 3. – С. 201-208.
6. Е.И. Прохоренко. Метод обнаружения пауз в речевых сигналах. Вестник московской академии рынка труда и информационных технологий. – Москва, 2006. №4(26). – С.13-20.
7. Д. Сэломон. Сжатие данных, изображений и звука. – Москва: Техносфера, 2004. – 368 с.

DIGITAL CODING OF CLIPPED SPEECH WITH PRESERVATION OF LEGIBILITY AND RECOGNITION OF THE ANNOUNCER

E.I. PROKHORENKO¹⁾

I.A. SIDORENKO²⁾

A.V. BOLDYSHEV³⁾

1) Belgorod state university

e-mail: prokhorenko@bsu.edu.ru

2) Belgorod state university

e-mail: sidorenko@bsu.edu.ru

3) Belgorod state university

e-mail: alexeiboldyshev@mail.ru

In the article new technology of compression of vocal signal is examined for the digital systems of information transfer, based on clipping speech with the maintainance of legibility and knowableness. The algorithms of the digital vocal signal processing, realizing discovery and delete of noise, are offered in the pauses of clipped speech, and also forming of information about the change of instantaneous spectrum speech on the basis of frequency presentations of vocal signal.

Keywords: compression of speech, legibility of speech, knowableness of voice, delete of pauses of speech, frequency presentation of speech, clipping, subbar transformation, variation method.