УДК 621.395
DOI: 10.30987/conferencearticle_5c19e69ce3af77.81660713

М.В. Илюшин, И.М. Антонов, И.А. Карабцев
(г. Орел, Академия Федеральной службы охраны Российской Федерации)

ЭФФЕКТИВНОЕ СЖАТИЕ РЕЧЕВОГО СИГНАЛА В ВИДЕ ПСЕВДОИЗОБРАЖЕНИЯ

Описан один из перспективных методов эффективного сжатия речевого сигнала, предполагающего его представление в виде псевдоизображения. К полученному псевдоизображению был применен алгоритм сжатия неподвижных изображений JPEG.

The describe one of the promising methods of effective voice compression is presented, which involves its representation in the form of a pseudo-image. The JPEG still images compression algorithm was applied to the resulting pseudo-image.

Ключевые слова: речевой сигнал, псевдоизображение, JPEG.
Keywords: speech signal, pseudo-image, JPEG.

В настоящее время эффективное использование пропускной способности каналов связи и объема памяти в запоминающих устройствах при передаче и хранении речевого сигнала соответственно является важным направлением при разработке перспективных и эксплуатации существующих сетей связи. Указанная тенденция диктует необходимость в применении высокоэффективных алгоритмов кодирования (сжатия) речевого сигнала (РС). При этом не стоит забывать о связи степени сжатия и качества звучания восстановленной речи.

На практике РС после его аналого-цифрового преобразования подвергается различным алгоритмам обработки. В частности, существует множество методов сжатия оцифрованного сигнала, а также методов повышения качества и разборчивости речи.

На сегодняшний день все многообразие форматов и стандартов сжатия РС можно разделить на методы непосредственного кодирования, параметрические и гибридные методы. Методы первой группы обеспечивают приемлемое качество воспринимаемой речи, но низкую степень сжатия; другие, наоборот, предусматривают хорошую степень сжатия, но высокую сложность реализации алгоритма и сравнительно невысокое качество воспиятия речи.

Аналитический обзор результатов исследований в области разработки методов компактного представления РС позволил выявить ряд перспективных методов. Указанные методы основаны на устранении перцептуальной избыточности исходного РС [1, 2] и временной

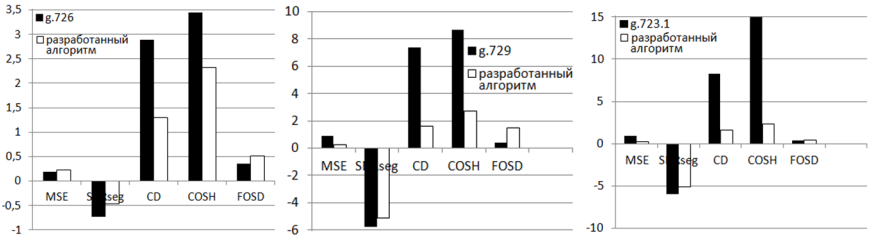
избыточности смежных речевых кадров, имеющих периодическую структуру [3], представлении отсчетов РС в виде псевдоизображения [3, 4].

Реализация алгоритма сжатия РС на основе представления его речевых отсчетов в виде псевдоизображения с дальнейшим применением стандарта *JPEG* [5] показала его состоятельность и конкурентноспособность по сравнению с популярными речевыми кодеками.

Изображение формировалось на основе телевизионной (построчной) развертки последовательности речевых отсчетов фрагмента русской речи блоками размером 8*8 (1 блок = 64 последовательно идущих отсчетов исходного фрагмента РС, картинка = 16*16 блоков). Коэффициент сжатия фрагмента исходного РС был равен 6. Результаты исследований позволили сделать вывод о сохранении в восстановленном фрагменте РС русской речи разборчивости и узнаваемости на достаточно высоком уровне, значения объективных мер искажений между исходным и восстановленным речевыми фрагментами показали способность предложенного алгоритма эффективно преобразовывать РС.

Задачи анализа РС и его дальнейшей эффективной обработки могут быть полезными не только для популярных применений (сжатие РС, идентификация личности по голосу и т.д.), но и для нетрадиционных задач, таких, как установление психо-физического состояния диктора в определенной ситуации (медицина, оборона), поиск людей (система оперативно-розыскных мероприятий), определение качества обслуживания (запись разговоров с оператором) и т.д.

При учете указанных возможных направлений применения методов эффективного представления РС были проведены исследования рассматриваемого алгоритма и получены оценки (рис. 1) качества синтезированного сигнала на скоростях 8, 10, 16 кбит\с в соответствии с различными степенями сжатия и методами временного (*MSE* – среднеквадратичная ошибка, *SNR* – отношение сигна/шум), частотного (*CD* – кепстральное расстояние, *COSH* – расхождение спектров исходного и декодированного сигналов) и психоакустического (*FOSD* – функция ощущения спектральной динамики) анализа.



а) б) в)

Рис. 1. Значения объективных показателей качества восстановленной русской речи некоторых низкоскоростных кодеков и разработанного алгоритма при сжатии x4 (а), x8 (б), x10 (в) раз

Из рис. 1 видно, что разработанный алгоритм показывает лучшие результаты по сравнению с существующими алгоритмами, причем чем выше степень сжатия (ниже скорость кодирования), тем лучше показатели качества восприятия речи. Исключением являются оценки, полученные с помощью функции ощущения спектральной динамики *FOSD* (для случаев а) и б) значения немного превышают оценки для алгоритмов *G*.726 и *G*.729, а в случае сравнения с алгоритмом *G*.723.1 – почти равны). Этот факт можно объяснить сравнительно малым обьемом выборки тестовых речевых фрагментов.

На рис. 2 показаны значения объективных показателей качества восприятия речи различных языков для разных коэффициентов сжатия.

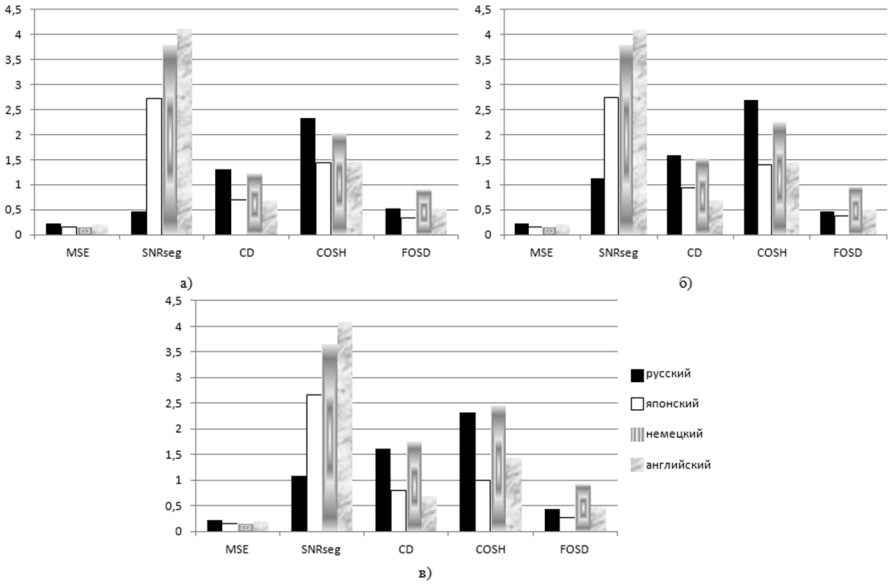


Рис.2. Значения объективных показателей качества восприятия речи разных языков при сжатии разработанным алгоритмом в 4 (а), 8 (б) и 10 раз (в)

Исходя из сравнительной характеристики видно, что при различной степени сжатия разработанный алгоритм показывает пропорциональные результаты на всех 4-х языках.

В заключение необходимо отметить, что в основе существующих стандартов сжатия подвижных изображений лежит алгоритм сжатия неподвижных изображений *JPEG*. В этой связи перспективным направлением исследований является разработка алгоритма совместного кодирования (декодирования) речевой и видеоинформации.

Предлагаемый алгоритм кодирования (декодирования) речевой и видеоинформации предполагает представление аудиоданных в виде псевдоизображений, инкапсулированных в блок кадра неподвижного изображения для совместного кодирования, например, алгоритмом *JPEG* (рис. 3).

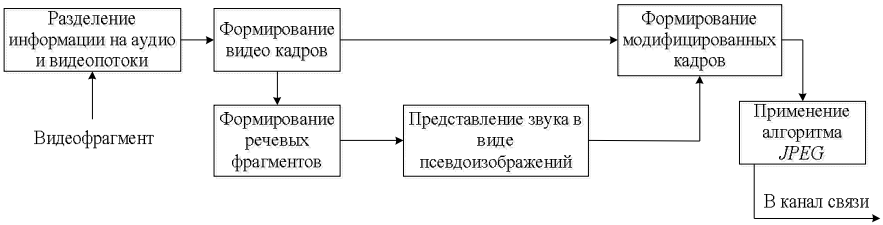


Рис. 3. Структурная схема совместного кодирования речевой и видеоинформации

При реализации указанного подхода необходимо исследовать степень влияния выбранного варианта развертки и матрицы квантования на значения объективных показателей качества восприятия неподвижного изображения.

Список литературы

1. *Лившиц, М.З.* Широкополосный CELP-кодер с мультиполосным возбуждением и многоуровневым векторным квантованием по кодовой книге с реконфигурируемой структурой / М.З. Лившиц, М. Парфенюк, А.А. Петровский // Цифровая обработка сигналов, 2005. – № 2. – С. 20–35.

2. *Илюшин, М.В.* Кодирование широкополосного речевого сигнала с адаптацией к психоакустическим особенностям восприятия синтезированной речи человеком / М.В. Илюшин, К.С. Беспалов, А.П. Бочарников. // Научно-технический журнал "Вестник Рязанского государственного радиотехнического университета" № 4 (выпуск 42). – 2012. – Ч.1. – С. 8–13.

3. *Гаврилов, И.А.* Сжатие аудиосигналов на основе межаудиокадровой обработки и псевдоизображений / И.А. Гаврилов, Х.Х. Носиров, М.Р. Мансурова. // Электросвязь. – 2010. – № 2. – С. 64–66.

4. *Носиров, Х.Х.* Фрактально-спектральный метод сжатия широкополосных аудиосигналов / Х.Х. Носиров, И.А. Гаврилов, Т.Г. Рахимов. // Электросвязь, 2010. – № 2. – С. 64–66.

5. *Илюшин, М.В.* К вопросу об эффективном сжатии речевого сигнала / М.В. Илюшин, М.В. Стремоухов, П.К. Литвин. // Современные технологии в науке и образовании – СТНО-2018 : сб. тр. междунар. науч.-техн. форума: в 10 т. / под общ. ред. О.В. Миловзорова. – Рязань, 2018.– Т.1. – С. 205–209.

Материал поступил в редколлегию 11.10.18.