DOI: 10.24412/2413-2527-2023-133-43-49

Сжатие спектра речевого потока путем передискретизации звуковых файлов

д.т.н. В. В. Егоров
Санкт-Петербургский государственный университет аэрокосмического приборостроения
Санкт-Петербург, Россия
egorovrimr@mail.ru

д.т.н. С. А. Лобов
Проектно-конструкторское бюро «РИО»
Санкт-Петербург, Россия
lsa_rimr@mail.ru

д.т.н. В. А. Ходаковский
Петербургский государственный университет путей сообщения Императора Александра I
Санкт-Петербург, Россия
hva1104@mail.ru

Аннотация. **Рассматривается задача максимально возможного сжатия цифрового речевого потока для его последующей передачи по узкополосному каналу связи при сохранении разборчивости речи после выполнения на приемной стороне процессов декомпрессии.**

В отличие от известных методов сжатия звука с использованием вокодерных принципов, в работе предлагается использовать:

- **сжатие спектра речи путем передискретизации;**
- **снижение динамического диапазона речевого потока;**
- **передачу в канал связи не отсчетов передискретизированного сигнала, а только значимых действительных и мнимых его компонент;**
- **упаковку 4-битных слов с информацией о действительных и мнимых компонентах сигнала в байтовый поток.**

Декомпрессия принятого потока выполняется в обратной последовательности.

Ключевые слова: **теорема отсчетов, компрессия и декомпрессия речевого потока, дискретизация и передискретизация сигнала.**

Введение

Задачи сжатия потоков данных, в том числе и аудиоданных, остаются актуальными в современных условиях. Вопросам описания методов и алгоритмов сжатия речевой информации посвящено в настоящее много статей, например обширный обзор сделан в [1–4]. Большая часть предлагаемых методов сжатия заключается в использовании вокодерного принципа.

В данной работе предлагается не вокодерный принцип сжатия, а сжатие именно спектра речевого сигнала с целью упаковки его в полосу частот не шире одной и даже менее октавы. В частности, в экспериментах использовалась полоса частот 375–500 Гц.

Постановка задачи сжатия спектра

Как известно, в соответствии с теоремой отсчетов Котельникова функция $s(t)$, имеющая спектр, ограниченный верхней граничной частотой f, может быть полностью восстановлена по ее m отсчетам U, выполненным с равномерным шагом, длительностью $\tau = 1/(2f)$:

$$s(t) = \sum_{k=0}^{m} U_k \times \operatorname{sinc}\left[2 \times \pi \times f \times \left\{t - \frac{k}{2 \times f}\right\}\right]. \qquad (1)$$

Если функция (1) подвергается дискретизации частотой Fd, то непрерывное время t необходимо заменить на дискретный аналог: i/Fd, и в результате получим:

$$S_i = \sum_{k=0}^{m} U_k \times \operatorname{sinc}\left[2 \times \pi \times f \times \left\{\frac{i}{Fd} - \frac{k}{2 \times f}\right\}\right],$$

$$i = 0, 1, \ldots, n-1. \qquad (2)$$

Один из авторов данной статьи в работах [5–7] изложил предположение, что теорема Котельникова может иметь и обратное толкование, когда мы по известной непрерывной функции с ограниченным спектром $s(t)$, хотим получить ее значительно более редкие отсчеты, которые полностью сохраняют всю информацию, содержащуюся в ней.

В такой интерпретации формулы (1) и (2) позволяют синтезировать непрерывный сигнал $s(t)$ по заданному информационному вектору U, а формула (3) восстанавливает информационный вектор U по непрерывному сигналу $s(t)$.

В настоящей статье авторы предлагают еще одну интерпретацию теоремы Котельникова.

Если функция S уже подвергалась дискретизации частотой Fd, и известно, что она имеет спектр, ограниченный некоторой частотой f, причем $f << Fd$, то поставленную выше задачу можно решить с использованием преобразования (3), которое выполняет передискретизацию сигнала S:

$$U_k = \frac{4 \times f}{Fd} \times \sum_{i=0}^{n} S_i \times \operatorname{sinc}\left[4\pi f \times \left\{\frac{i}{Fd} - \frac{k}{2 \times f}\right\}\right],$$

$$k = 0, 1, \ldots, m-1, \qquad (3)$$

где f — половина новой частоты дискретизации;
Fd — частота дискретизации преобразуемого сигнала;
S — преобразуемый сигнал (вектор размерности n);
U — выходной сигнал (вектор размерности m);
$\operatorname{sinc} x = (\sin x)/x$.

Таким образом, если исходный сигнал S имел n отсчетов при частоте дискретизации Fd, то после передискретизации формируется сигнал U, имеющий m отсчетов при частоте дискретизации $2f$, значит, будет иметь место сжатие информационного объема вектора S до объема вектора U.

Здесь, однако, следует отметить, что степень сжатия как отношение n/m, или Fd/f не может быть слишком высокой, поскольку при очень низкой частоте дискретизации $2f$ будет иметь место высокая потеря информации, которая может привести к потере разборчивости речи. Это следует из теоремы Котельникова, поскольку для

сигнала с верхней граничной частотой f требуется выполнить не менее двух отсчетов на периоде, т. е. использовать частоту дискретизации $2f$. С точки зрения эффективного сжатия речевого потока необходимо выяснить, при какой частоте дискретизации разборчивость речи останется достаточной.

Описание вычислительных экспериментов по сжатию потока

Для вычислительных экспериментов была выбрана среда Mathcad 14, поскольку в ее библиотеке встроенных функций имеются средства для синтеза звуковых файлов PCM (англ. *pulse code modulation* — импульсно-кодовая модуляция) с расширением *.wav. В качестве средства воспроизведения звуковых файлов использовалось свободно распространяемое приложение Audacity 2.4.2, в котором имеется широкий набор средств обработки звука.

Используя указанные средства, было решено провести несколько вычислительных экспериментов с целью выяснения какой может быть наименьшая частота дискретизации речевого потока при выполнении ограничения по разборчивости получаемого речевого сигнала. В результате этих экспериментов было выяснено, что при частоте дискретизации 1 000 Гц разборчивость речевого потока может быть вполне приличной.

Для высококачественного звукового сигнала PCM стандартной частотой дискретизации считается 44 100 Гц с числом уровней квантования амплитуды 16 бит при двух каналах, однако для передачи такого сигнала по радиоканалу потребуется слишком высокая скорость передачи информации, поэтому для экспериментов выбрана одноканальная 8-битная запись с частотой дискретизации 8 000 Гц, что несколько превышает требования Теоремы отсчетов для канала тональной частоты (ТЧ) с диапазоном 300–3 400 Гц.

Для вычислительных экспериментов был выбран короткий фрагмент речевого потока (исходная частота дискретизации $Fd = 8\,000$ Гц, длительность $t = 4{,}8$ с, общее число отсчетов $n = 39\,002$, размерность файла 38,1 Kb). На рисунке 1 приведена его амплитудно-временная зависимость, а на рисунке 2 — спектрограмма.

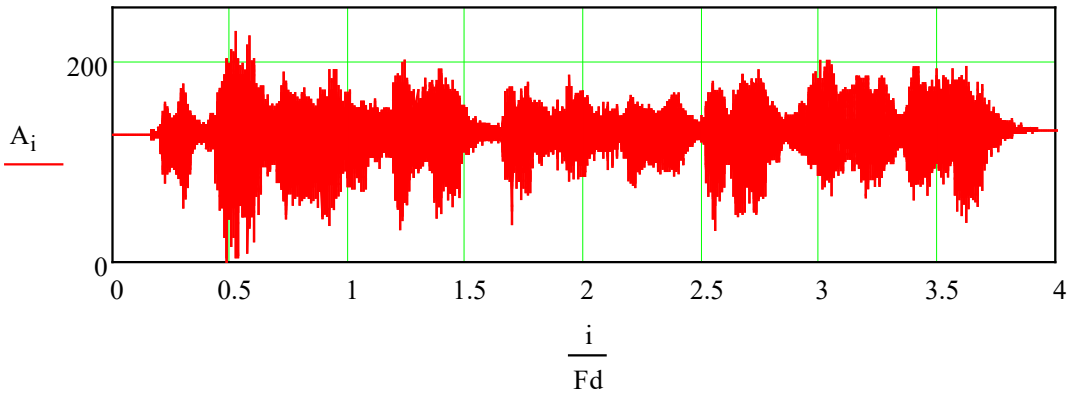

Рис. 1. Исходный 8-битный PCM речевой сигнал

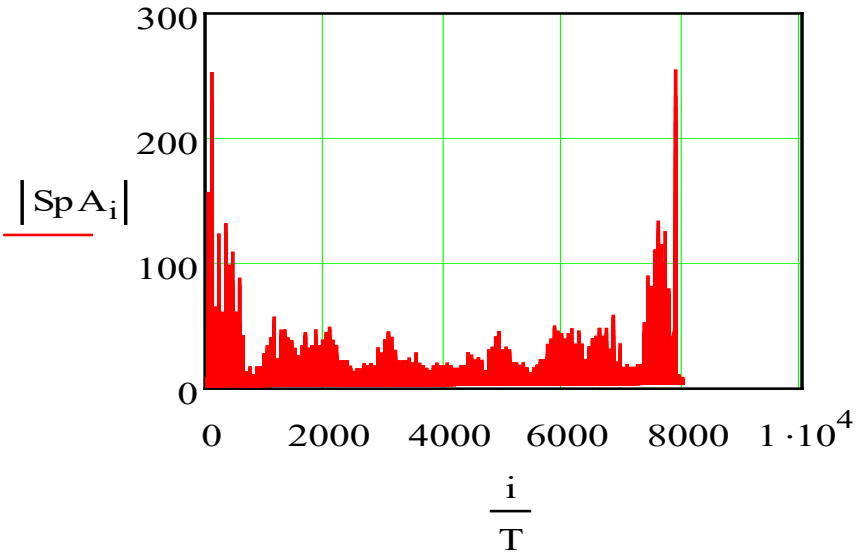

Рис. 2. Спектр 8-битного речевого сигнала

Из анализа рисунка 2 видно, что ширина спектра потока составляет 8 000 Гц, причем спектр имеет симметричный характер относительно частоты 4 000 Гц, что является половиной частоты дискретизации. Характер спектра позволяет сделать вывод о том, что в результате дискретизации имеет место амплитудная модуляция относительно поднесущей частоты, равной половине частоты дискретизации — 4 000 Гц, т. е. речевой поток в данном примере имеет полосу 0–4 000 Гц.

В соответствии с указанным в начале статьи алгоритмом обработки данных, исходный сигнал (рис. 1) подвергался линейному преобразованию вида (3), результат которого приведен на рисунке 3, а его спектрограмма — на рисунке 4. Причем в преобразовании (3) в качестве исходной частоты дискретизации применялась частота Fd = 8 000 Гц, а новая частота дискретизации принималась равной $2f$ = 1 000 Гц.

На рисунке 3 видно, что характер изменения огибающей сигнала полностью соответствует исходному сигналу, но имеет в 8 раз меньшее число временных отсчетов (4 875 против 39 002 в исходном потоке). Для проверки разборчивости полученного речевого потока сигнал с использованием встроенной функции среды Mathcad 14 WRITEWAV(“e:/Prov.wav, 1200, 8”) выводился для прослушивания на звуковую карту. Здесь следует отметить, что частота дискретизации в выходной функции writewav(“__”, 1200, 8) устанавливалась равной 1 200 Гц при числе уровней квантования амплитуды 8 бит.

Результаты оказались очень обнадеживающими, поскольку разборчивость речи оказалось хорошей и можно было даже узнать голос диктора. В результате выполненных преобразований был получен битрейт 8 000 бит/с (полученный файл Prov.wav имел размерность 4,8 Кб против 38,1 Кб у исходного файла).

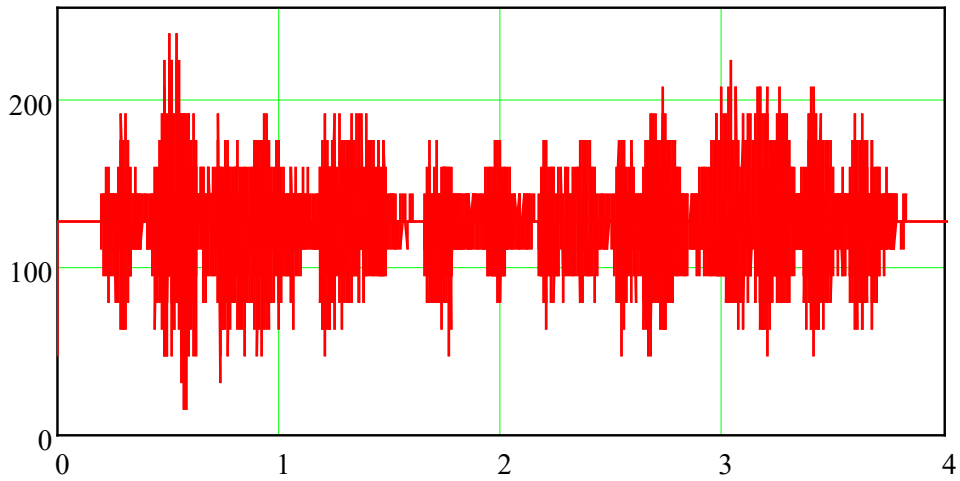

Рис. 3. Сигнал после передискретизации и снижения динамического диапазона до 4 бит (16 уровней), общее число отсчетов $n1$ = 4 875

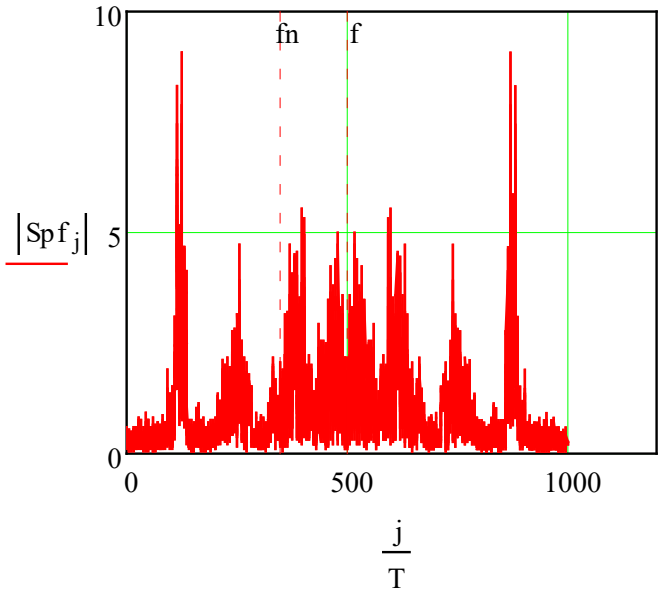

Рис. 4. Спектр сигнала после передискретизации частотой 1 000 Гц

Из анализа спектрограммы на рисунке 4 видно, что ширина спектра потока снизилась до 1 000 Гц, причем спектр имеет симметричный характер относительно частоты 500 Гц, что является половиной новой частоты дискретизации. Характер спектра позволяет сделать вывод о том, что в результате передискретизации имеет место амплитудная модуляция относительно поднесущей частоты, равной половине частоты дискретизации — 500 Гц, а полоса спектра модулирующего сигнала составила 500 Гц против исходных 4 000 Гц, т. е. имеет место сжатие спектра в 8 раз.

Поскольку разборчивость оказалась хорошей, авторами было решено продолжить эксперименты по сжатию в соответствии с назначенным алгоритмом. Необходимо было выделить значимую часть спектра, приведенного на рисунке 4. Было принято решение в качестве нижней границы полосы анализа принять частоту fn = 375 Гц, а в качестве верхней границы fv = 500 Гц.

В результате из общей размерности вектора спектральных чисел, равного 4 875, были выделены 732 действительных и 732 мнимых спектральных числа, причем их динамический диапазон снижен с 8 до 4 бит. Выделенные 732 комплексных спектральных числа соответствовали частотам обрабатываемого диапазона 732–500 Гц, а остальные спектральные числа обнулялись, что позволило сжать поток еще в 6,6 раза, т. е. общий коэффициент сжатия составил 52 раза.

Поток байтов представлял из себя 732 4-битных слова с действительными компонентами спектра и столько же слов с мнимыми компонентами, что в результате кодирования реализовано в виде 732 8-битных слов.

На рисунке 5 приведена спектрограмма действительных спектральных чисел, а на рисунке 6 — мнимых спектральных чисел.

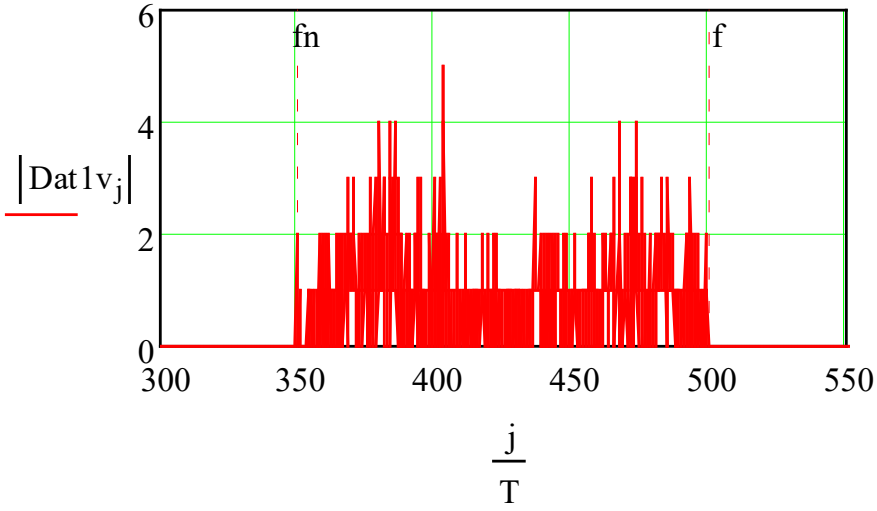

Рис. 5. Действительные спектральные числа, содержащие основную информацию о принятом речевом потоке

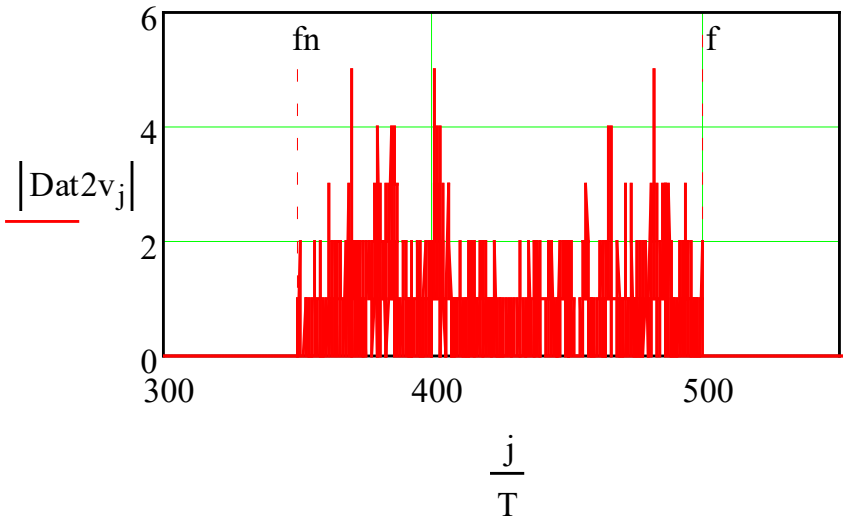

Рис. 6. Мнимые спектральные числа, содержащие основную информацию о принятом речевом потоке

В результате упаковки получился поток из 732 байтов, которые содержат информацию о звуковом файле длительностью 4,875 с, то есть битрейт составил 732 × 8 ⁄ 4,875 = 1 200 бит/с. Данные передавались в файл с использованием функции Mathcad 14:

$$\text{WRITEBIN}(\text{“e/Frazy\_5.dat,”byte”},0):=U. \qquad (4)$$

Битрейт 1 200 бит/с является очень неплохим результатом и может быть использован в цифровых узкополосных каналах связи.

Следует также отметить, что поток из 732 байтов, содержащих всю информацию о звуковом фрагменте, звучащем 4,875 с, может быть подвергнут сжатию методом оптимального кодирования по Хаффману или Шеннону, что дополнительно обеспечит сжатие не менее чем в 2 раза.

На рисунке 7 представлена гистограмма вероятностей значений уровней в файле (4).

Из графика видно, что 36 % значений содержат информацию об уровне «128», что для 8-уровневого PCM соответствует «0» и еще по 18 % приходится на два ближайших уровня, т. е. около 70 % из переданных значений соответствуют всего трем уровням из 256, что и говорит о возможности сжатия такого цифрового потока.

Описание экспериментов по распаковке

Поток 8-битных слов распаковывался в два 4-битных потока, затем по действительной и мнимой компонентам формировался комплексный спектр, который дополнялся нулями. В результате был восстановлен полный спектр, состоящий из 4 875 спектральных чисел. Указанный восстановленный спектр приведен на рисунке 8.

Путем применения обратного преобразования Фурье восстановленный спектр преобразовывался во временной сигнал, представленный на рисунке 9.

В соответствии с алгоритмом обработки после восстановления сигнала во временной области производилось восстановление исходной частоты дискретизации 8 000 Гц с коэффициентом 1,4, т. е. с Fd = 10 400 Гц с использованием преобразования (1) (рис. 10).

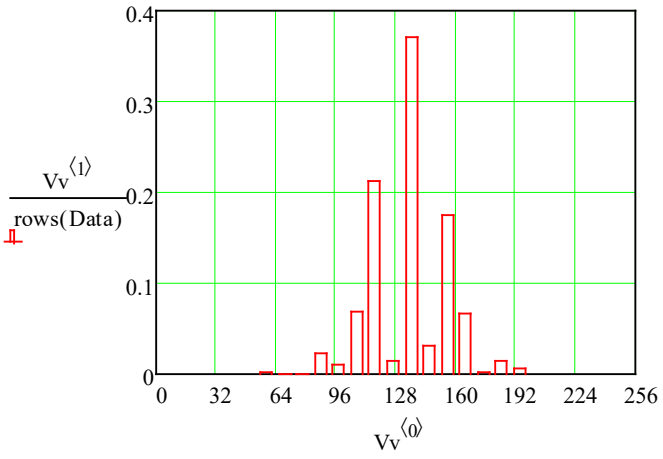

Рис. 7. Гистограмма вероятностей значений уровней в сжатом файле

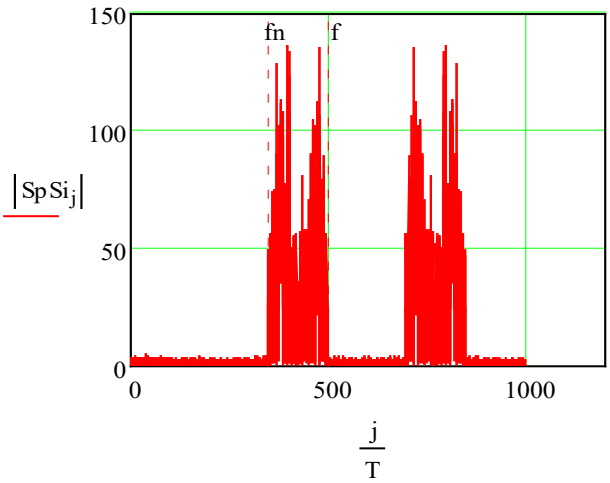

Рис. 8. Восстановленный спектр fn = 375 Гц, f = 500 Гц

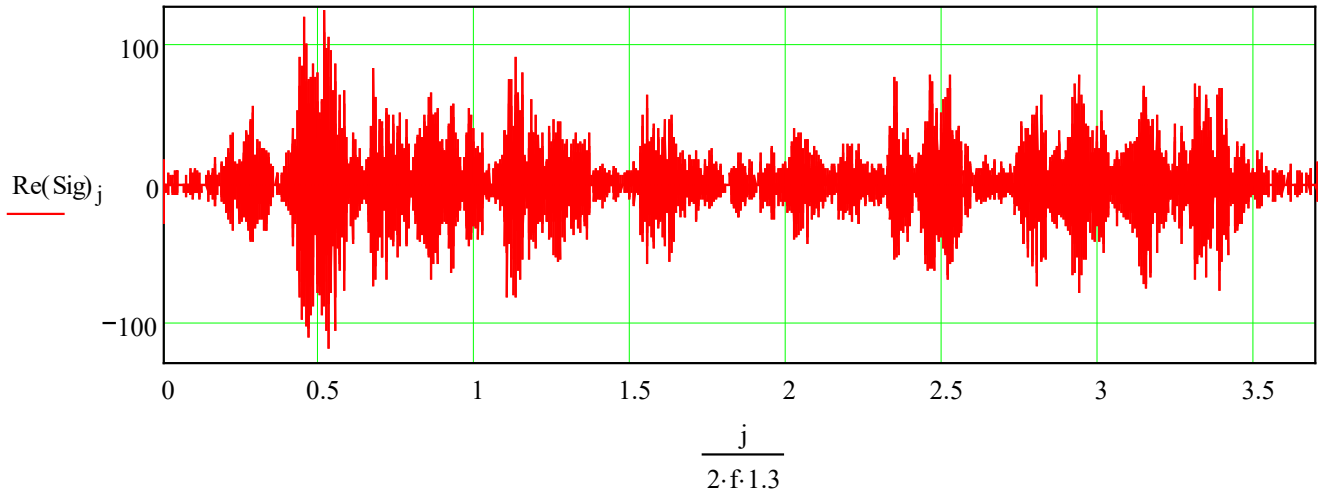

Рис. 9. Восстановленный сигнал при частоте дискретизации 1 300 Гц

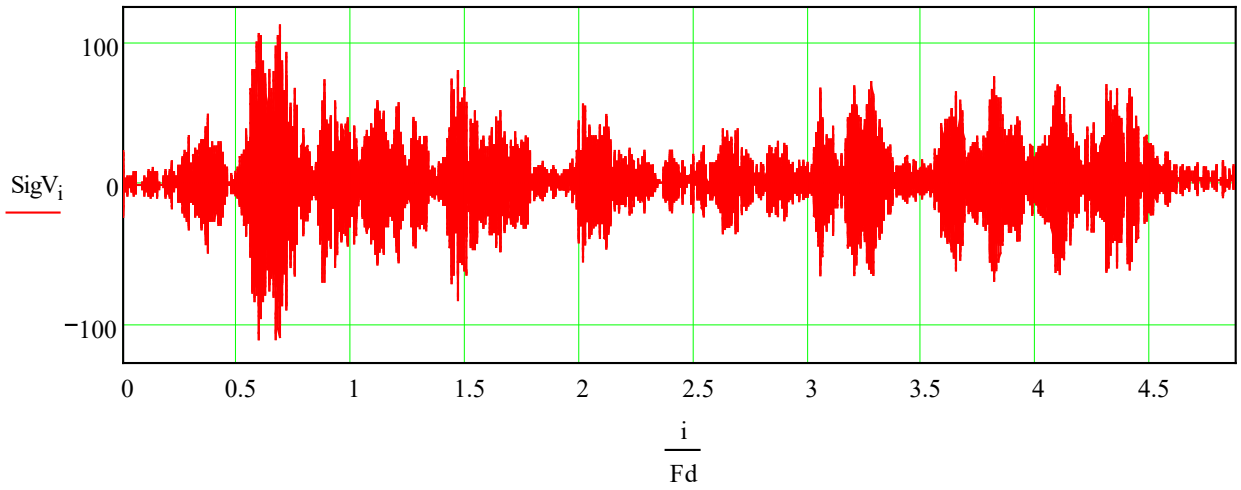

Рис. 10. Восстановленный сигнал после возврата к исходной частоте дискретизации с повышением до 10 400 Гц для восстановления исходного темпа речи

Затем для прослушивания результатов восстановления вызывалась функция записи звукового файла

WRITEWAV(“e:/Prov_8_V.wav”,Fd*1.4,8):=SigV+2^7.

Результаты прослушивания восстановленного звукового фрагмента позволили сделать убедительный вывод о достаточной разборчивости речи для ее восприятия.

Заключение

Задача, поставленная авторами, о возможности существенного сжатия звуковых файлов за счет передискретизации оказалась успешно реализуемой. В ходе экспериментов достигнуто сжатие в 8 раз за счет снижения частоты дискретизации и в 6,7 раза за счет передачи только значимых комплексных коэффициентов преобразования Фурье, что обеспечило общий коэффициент сжатия речи более 50 и возможность устойчивой передачи речи по узкополосному каналу связи с канальной скоростью 1 200 бит/с без использования методов оптимального кодирования битовых потоков.

В ходе проведенных экспериментов удалось выяснить, что можно одновременно заметно сжать полосу формируемого речевого потока и снижать динамический диапазон с 45 до 18 дБ, т. е. вместо 8-битного сигнала PCM использовать 4-битное квантование уровней сигнала.

Анализ результатов эксперимента показал, что за счет использования оптимального кодирования можно еще повысить коэффициент сжатия звуковых файлов.

Литература

1. Подвальный, С. Л. Обзор методов и алгоритмов сжатия речевой информации в системах цифровой радиосвязи / С. Л. Подвальный, А. Д. Рощупкин // Вестник Воронежского государственного технического университета. 2017. Т. 13, № 2. С. 7–13.

2. Рабинер, Л. Р. Цифровая обработка речевых сигналов = Digital processing of speech signals / Л. Р. Рабинер, Р. В. Шафер; пер. с англ. под ред. М. В. Назарова и Ю. Н. Прохорова. — Москва: Радио и связь. Редакция литературы по электросвязи, 1981. — 496 с.

3. Flanagan, J. L. Source-System Interactions in the Vocal Tract / J. L Flanagan // Annals of the New York Academy of Science. 1968. Vol. 155, Is. 1. Pp. 9–17. DOI: 10.1111/j.1749-6632.1968.tb56744.x.

4. Wang, C. Robust Pitch Tracking for Prosodic Modeling in Telephone Speech / C. Wang, S. Seneff // Proceedings of the 2000 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), (Istanbul, Turkey, 05–09 June 2000). — Institute of Electrical and Electronics Engineers, 2000. — Vol. 3. — Pp. 1343–1346. DOI: 10.1109/ICASSP.2000.861827.

5. Ходаковский, В. А. О теореме отсчетов и ее применении для синтеза и анализа сигналов с ограниченным спектром / В. А. Ходаковский, В. Г. Дегтярев // Известия Петербургского университета путей сообщения. 2017. Т. 14, № 3. С. 562–573.

6. Ходаковский, В. А. Теорема отсчетов и обратное ее толкование для анализа сигналов с ограниченным спектром // Проблемы математической и естественнонаучной подготовки в инженерном образовании: Сборник трудов IV Международной научно-методической конференции (Санкт-Петербург, Россия, 03 ноября 2016 г.) / под ред. В. А. Ходаковского. — Санкт-Петербург: ПГУПС, 2017. — Т. 2. — С. 135–147.

7. Ходаковский, В. А. Синтез многополосного фильтра с требуемой частотной характеристикой / В. А. Ходаковский, Т. В. Ходаковский // Интеллектуальные технологии на транспорте. 2015. № 1. С. 38–42.

DOI: 10.24412/2413-2527-2023-133-43-49

Compression of the Speech Stream Spectrum by Resampling Sound Files

Grand PhD V. V. Egorov
Saint Petersburg State University of Aerospace Instrumentation
Saint Petersburg, Russia
egorovrimr@mail.ru

Grand PhD S. A. Lobov
Design and Construction Bureau «RIO»
Saint Petersburg, Russia
lsa_rimr@mail.ru

Grand PhD V. A. Khodakovsky
Emperor Alexander I St. Petersburg State Transport University
Saint Petersburg, Russia
hva1104@mail.ru

Annotation. **In this paper, we consider the problem of the maximum possible compression of the digital speech stream for its subsequent transmission over a narrow-band communication channel while maintaining speech intelligibility after decompression processes are performed on the receiving side. In contrast to the known methods of sound compression using vocoder principles, it is proposed to use: compression of the speech spectrum by resampling; reduction of the dynamic range of the speech stream; transmission to the communication channel of not samples of the oversampled signal, but only significant real and imaginary components of it; packing 4-bit words with information about the real and imaginary components of the signal into a byte stream. Decompression of the received stream is performed in the reverse sequence.**

Keywords: **sampling theorem, compression and decompression of the speech stream, sampling and resampling of the signal.**

REFERENCES

1. Podvalny S. L., Roshupkin A. D. Review of Methods and Algorithms of Speech Information Compression in Digital Communication Systems [Obzor metodov i algoritmov szhatiya rechevoy informatsii v sistemakh tsifrovoy radiosvyazi], *Bulletin of Voronezh State Technical University [Vestnik Voronezhskogo gosudarstvennogo tekhnicheskogo universiteta]*, 2017, Vol. 13, No. 2, Pp. 7–13.

2. Rabiner L. R., Schafer R. W. Digital processing of speech signals [Tsifrovaya obrabotka rechevykh signalov]. Moscow, Radio and Communications Publishing House, 1981, 496 p.

3. Flanagan J. L. Source-System Interactions in the Vocal Tract, *Annals of the New York Academy of Science*, 1968, Vol. 155, Is. 1, Pp. 9–17.
DOI: 10.1111/j.1749-6632.1968.tb56744.x.

4. Wang C., Seneff S. Robust Pitch Tracking for Prosodic Modeling in Telephone Speech, *Proceedings of the 2000 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), Istanbul, Turkey, June 05–09, 2000. Volume 3*. Institute of Electrical and Electronics Engineers, 2000, Pp. 1343–1346. DOI: 10.1109/ICASSP.2000.861827.

5. Khodakovskiy V. A., Degtyarev V. G. On Sampling Theorem and Its Application for The Puposes of Synthesis and Analysis of Band-Limited Signals [O teoreme otschetov i ee primenenii dlya sinteza i analiza signalov s ogranichennym spektrom], *Proceedings of Petersburg Transport University [Izvestiya Peterburgskogo universiteta putey soobshcheniya]*, 2017, Vol. 14, No. 3, Pp. 562–573.

6. Khodakovsky V. A. The Sampling Theorem and Its Reverse Interpretation for the Analysis of Signals with a Limited Spectrum [Teorema otschetov i obratnoe ee tolkovanie dlya analiza signalov s ogranichennym spektrom], *Problems of Mathematical and Natural Science Training in Engineering Education: Proceedings of the IV International Scientific and Methodological Conference [Problemy matematicheskoy i estestvennonauchnoy podgotovki v inzhenernom obrazovanii: Sbornik trudov IV Mezhdunarodnoy nauchno-metodicheskoy konferentsii], Saint Petersburg, Russia, November 03, 2016. Volume 2*. St. Petersburg, PSTU, 2017, Pp. 135–147.

7. Khodakovskiy V. A., Khodakovskiy T. V. Synthesis of Multi-Band Digital Filter with Demand of Frequency Characteristic [Sintez mnogopolosnogo filtra s trebuemoy chastotnoy kharakteristikoy], I*ntellectual Technologies on Transport [Intellektualnye tekhnologii na transporte]*, 2015, No. 1, Pp. 38–42.