

# ОБ ОПТИМИЗАЦИИ ПРОЦЕДУР СЖАТИЯ РЕЧЕВЫХ ДАННЫХ

**И.И.Чижов**  
**Т.Н. Созонова**  
**И. В. Деев**

*Белгородский  
государственный  
университет*

*e-mail: [chizhov@bsu.edu.ru](mailto:chizhov@bsu.edu.ru)*

---

В статье рассмотрены основные современные алгоритмы сжатия речевых данных с потерями, прошедшие международную стандартизацию. В настоящее время авторы подобных методов сжатия данных пытаются найти оптимальное соотношение между минимальным битрейтом и максимальным субъективным качеством восстановленного сигнала.

Ключевые слова: речевой сигнал, сжатие данных, кодек, алгоритм сжатия, интерполяция, среднеквадратическое отклонение.

## **Введение**

Стоит отметить, что речь обладает такими существенными отличиями от обобщенного звукового сигнала, как более узкая полоса частот (около 4 кГц) и наличие значительного числа неинформативных пауз в сигнале, что позволило создать ряд алгоритмов сжатия ориентированных только на речевые сигналы.

На сегодняшний день, наибольшее распространение алгоритмы сжатия речи с потерями получили в системах IP-телефонии. VoIP (Voice-over-IP) – IP-телефония – система связи, при которой аналоговый звуковой сигнал от одного абонента дискретизируется (кодируется в цифровой) вид, компрессируется и пересылается по цифровым каналам связи до второго абонента, где производится обратная операция – декомпрессия, декодирование и воспроизведение аналогового сигнала.

Возможность передачи голосовых сообщений через сеть с пакетной коммутацией впервые была реализована в 1993 году. Данная технология получила название VoIP (Voice over IP). Одним из частных приложений данной технологии является IP-телефония – услуга по передаче телефонных разговоров абонентов по протоколу IP.

Основными преимуществами технологии VoIP является сокращение требуемой полосы пропускания, что обеспечивается учётом статистических характеристик речевого трафика:

- блокировкой передачи пауз (диалоговых, слоговых, смысловых и др.), которые могут составлять до 40-50 % времени занятия канала передачи;
- высокой избыточностью речевого сигнала и его сжатием (без потери качества восстановления) до уровня 20-40 % исходного сигнала.

С другой стороны трафик VoIP критичен к задержкам пакетов в сети, но обладает толерантностью (устойчивостью) к потерям отдельных пакетов. Так потеря до 5 % пакетов не приводит к ухудшению разборчивости речи.

## **Алгоритмы кодирования речевых сигналов**

Источником информационных данных является речевой сигнал, возможной моделью которого является нестационарный случайный процесс. В первом приближении можно выделить следующие типы сигнальных фрагментов: вокализированные, невокализированные, переходные и паузы. При передаче речи в цифровой форме каждый тип сигнала при одной и той же длительности и одинаковом качестве требует различного числа бит для кодирования и передачи. Следовательно, скорость передачи разных типов сигнала также может быть различной, что обуславливает применение кодеков с переменной скоростью. В результате передача речевых данных в каждом направлении дуплексного канала рассматривается как передача асинхронных логически самостоятельных фрагментов цифровых последовательностей (транзакций) с датаграммной синхронизацией внутри транзакции, наполненной блоками различной длины. В основе кодека речи с переменной скоростью лежит классификатор входного сигнала, определяющий степень его информативности и, таким образом, задающий метод кодирования и скорость передачи речевых данных. Наиболее простым классификатором речевого сигнала является Voice Activity Detector (VAD), который выделяет во входном речевом сигнале активную речь и паузы. При этом, фрагменты сигнала, классифицируемые как активная речь, кодируются каким-либо из известных

алгоритмов (как правило, на базе метода Code Excited Linear Prediction – CELP) с базовой скоростью 4-8 кбит/с. Фрагменты, классифицированные как паузы, кодируются и передаются с низкой скоростью порядка 0.1 – 0.2 Кбит/с, либо не передаются вообще. При этом передача минимальной информации о фрагментах пауз предпочтительна. Данная стратегия позволяет оптимизировать скорость кодирования до 2-4 кбит/с при достаточном качестве синтезируемой речи. При этом для особо критичных фрагментов речевого сигнала выделяется большая скорость передачи, для менее ответственных – меньшая. Вместе с тем необходимо отметить, что вокодер вносит дополнительную задержку порядка 15-45 мс, возникающую по следующим причинам:

- использование буфера для накопления сигнала и учёта статистики последующих отсчётов (алгоритмическая задержка);
- математические преобразования, выполняемые над речевым сигналом, требуют процессорного времени (вычислительная задержка).

Данную задержку необходимо учитывать при расчёте сквозных задержек. Проведённый в различных исследовательских группах анализ качества передачи речевых данных через сеть Интернет показывает, что основным источником возникновения искажений, снижения качества и разборчивости синтезированной речи является прерывание потока речевых данных, вызванное:

- потерями пакетов при передаче по сети связи;
- превышением допустимого времени доставки пакета с речевыми данными.

Это требует решения задачи оптимизации задержек в сети и создание алгоритмов компрессии речи устойчивых к потерям пакетов (восстановления потерянных пакетов).

### **Кодеки речевых данных**

Проприетарные:

- G.728.

Гибридный кодек, описанный в рекомендации G.728 в 1992 г, относится к категории LD-CELP – Low Delay – Code Excited Linear Prediction – Кодек с управляемым кодом линейным предсказанием и малой задержкой. Кодек обеспечивает скорость преобразования 16 Кбит/с, вносит задержку при кодировании от 3 до 5 мс и предназначен для использования в системах видеоконференций. В устройствах IP-телефонии данный кодек применяется достаточно редко.

- G.729.

Семейство включает кодеки G.729, G.729 Annex A, G.729 Annex B (содержит VAD и генератор комфортного шума). Кодеки G.729 сокращённо называют CS-ACELP Conjugate Structure – Algebraic Code Excited Linear Prediction – Сопряжённая структура с управляемым алгебраическим кодом линейным предсказанием. Процесс преобразования использует 21,5 MIPS и вносит задержку 15 мс. Скорость закодированного речевого сигнала составляет 8 Кбит/с. В устройствах VoIP данный кодек занимает лидирующее положение, обеспечивая наилучшее качество кодирования речевой информации при достаточно высокой компрессии.

- G.723.1.

Рекомендация G.723.1 описывает гибридные кодеки, использующие технологию кодирования речевой информации, сокращённо называемую – MP-MLQ (Multy-Pulse – Multy Level Quantization – Множественная Импульсная, Многоуровневая Квантизация), данные кодеки можно охарактеризовать, как комбинацию АЦП/ЦАП и вокодера. Как уже упоминалось выше, своим возникновением гибридные кодеки обязаны системам мобильной связи. Применение вокодера позволяет снизить скорость передачи данных в канале, что принципиально важно для эффективного использования как радиотракта, так и IP-канала. Основным принцип работы вокодера – синтез исходного речевого сигнала посредством адаптивной замены его гармонических составляющих соответствующим набором частотных фонем и согласованными шумовыми коэффициентами. Кодек G.723 осуществляет преобразование аналогового сигнала в поток данных со скоростью 64 Кбит/с (ИКМ), а затем при помощи многополосного цифрового фильтра/вокодера выделяет частотные фонемы, анализирует их и передаёт по IP-каналу информацию только о текущем состоянии фонем в речевом сигнале. Данный алгоритм преобразования позволяет снизить скорость закодированной информации до 5,3 – 6,3 Кбит/с без видимого ухудшения качества речи. Кодек имеет две скорости и два

варианта кодирования: 6,3 Кбит/с с алгоритмом MP-MLQ и 5,3 Кбит/с с алгоритмом CELP. Первый вариант предназначен для сетей с пакетной передачей голоса и обеспечивает лучшее качество кодирования по сравнению с вариантом CELP, но менее адаптирован к использованию в сетях со смешанным типом трафика (голос/данные). Процесс преобразования требует от DSP 16,4 – 16,7 MIPS (Million Instructions Per Second) и вносит задержку 37 мс. Кодек G.723.1 широко применяется в голосовых шлюзах и прочих устройствах IP-телефонии. Кодек уступает по качеству кодирования речи кодеку G.729a, но менее требователен к ресурсам процессора и пропускной способности канала.

Бесплатные:

- GSM.
- G.726.

Рекомендация G.726 описывает технологию кодирования с использованием Адаптивной Дифференциальной Импульсно-Кодовой Модуляции (АДИКМ) со скоростями: 32 Кбит/с, 24 Кбит/с, 16 Кбит/с. Процесс преобразования не вносит существенной задержки и требует от DSP 5,5 – 6,4 MIPS. Кодек может применяться совместно с кодеком G.711 для снижения скорости кодирования последнего. Кодек предназначен для использования в системах видеоконференций.

- G.711 uLaw.
- G.711 aLaw.

Рекомендация, утверждённая МККТТ в 1984 г., описывает кодек, использующий ИКМ преобразование аналогового сигнала с точностью 8 бит, тактовой частотой 8 кГц и простейшей компрессией амплитуды сигнала. Скорость потока данных на выходе преобразователя составляет 64 Кбит/с (8 Бит , 8 кГц). Для снижения шума квантования и улучшения преобразования сигналов с небольшой амплитудой, при кодировании используется нелинейное квантование по уровню согласно специальному псевдо-логарифмическому закону A или  $m - \text{Law}$ . Первые ИКМ кодеки с нелинейным квантованием появились уже в 60-х гг. Кодек G.711 широко распространён в системах традиционной телефонии с коммутацией каналов. Несмотря на то, что рекомендация G.711 в стандарте H.323 является основной и первичной, в шлюзах IP-телефонии данный кодек применяется редко из-за высоких требований к полосе пропускания и задержкам в канале передачи (всё-таки 64 Кбит/с это много). Использование G.711 в системах IP-телефонии обосновано лишь в тех случаях, когда требуется обеспечить максимальное качество кодирования речевой информации при небольшом числе одновременных разговоров.

### **Вычислительные эксперименты**

Вычислительный эксперимент состоял из следующих этапов:

1. Оцифровка и сохранение в формате .wav исходного речевого сигнала;
2. Кодирование исходного сигнала при помощи наиболее современных алгоритмов сжатия речевых данных с потерями на типичных для данного алгоритма битрейтах;
3. Оценка субъективного качества восстановленных сигналов группой экспертов по традиционной 5-ти бальной шкале, где наилучшему качеству звучания соответствует наибольший бал и вычисление средней оценки для алгоритма.

В таблице представлены наиболее типичные результаты экспериментов.

**Результаты оценки субъективного качества восстановленных сигналов**

№	Кодек	Тип кодека	Битрейт	Оценка
1	G.711	ИКМ	64 Кбит/с	4,1
2	G.726	АДИКМ	32 Кбит/с	3,85
3	G.728	LD – CELP	16 Кбит/с	3,61
4	Skype	AMR-WB	14 Кбит/с	4,1
5	G.729	CS – ACELP (без VAD)	8 Кбит/с	3,92
6	G.729	2-х кратное кодирование	8 Кбит/с	3,27
7	G.729	3-х кратное кодирование	8 Кбит/с	2,68
8	G.729a	CS – ACELP	8 Кбит/с	3,7
9	G.723.1	MP – MLQ	6,3 Кбит/с	3,9
10	G.723.1	ACELP	5,3 Кбит/с	3,65

**Литература**

1. Артюшенко, В. М. Цифровое сжатие видеoinформации и звука [Текст] : учеб. пособие / В. М. Артюшенко, О. И. Шелухин, М. Ю. Афонин ; под ред. В. М. Артюшенко. – М. : Дашков и К, 2003. – 426 с.
2. Баранов, Л. А. Квантование по уровню и временная дискретизация в цифровых системах управления [Текст] / Л. А. Баранов. – М. : Энергоатомиздат, 1990. – 304 с.
3. Витерби, Э. Д. Принципы цифровой связи и кодирования [Текст] / А. Д. Витерби, Дж. К. Омура ; пер. с англ. и под ред. К. Ш. Зигангирова. – М. : Радио и связь, 1982. – 536 с. : ил. – (Статистическая теория связи ; вып. 18).
4. Вологдин, Э. И. Слух и восприятие звука [Текст] : учеб. пособие / Э. И. Вологдин. – СПб. : СТ «Факультет ДВО», 2004. – 52 с.
5. Голд, Б. Цифровая обработка сигналов [Текст] : пер. с англ. / Б. Голд, Ч. Рейдер. – М. : Сов. радио, 1973. – 376 с.
6. Грудинин, А. С. Кодирование сигналов звукового вещания в базе дискретного косинусного преобразования [Текст] / А. С. Грудинин, А. М. Синильников // Техника средств связи. – 1986. – Вып. 3. – С. 3-10. – (Сер. ТРПА).
7. Жилияков, Е. Г. О субполосном кодировании сигнала [Текст] / Е. Г. Жилияков,
8. И. Г. Попов, И. И. Чижов // Вестник НТУ (ХПИ) : сб. науч. тр. – Харьков, 2004.
9. № 46. – С. 10-20. – (Тем. вып. «Информатика и моделирование»).
10. Жилияков, Е. Г. Оптимальный синтез квантователя по уровню [Текст] / Е. Г. Жилияков, И. Г. Попов, И. И. Чижов // Вестник НТУ (ХПИ) : сб. науч. тр. – Харьков, 2004. – № 46. – С. 101-106. – (Тем. вып. «Информатика и моделирование»).
11. Ковалгин, Ю. А. Цифровое кодирование звуковых сигналов : учеб. пособие [Текст] / Ю. А. Ковалгин, Э. И. Вологдин. – СПб. : КОРОНА-принт, 2004. – 240 с.

**THE COMPARISON OF THE MODERN METHODS OF SPEECH COMPRESSION**

I.I. Chizhov  
T.N. Sozonova  
I.V. Deev

*Belgorod state university*

*e-mail: [chizhov@bsu.edu.ru](mailto:chizhov@bsu.edu.ru)*

In the article the most part of modern methods of speech compression are described. The main problem of such methods is to find the minimum bitrate where the quality of the reconstructed speech signal is quite good yet.

Key words: speech signal, data compression, codec, compression method, decimation, interpolation, the relative error.