ФЕДЕРАЛЬНОЕ АГЕНТСТВО

ПО ТЕХНИЧЕСКОМУ РЕГУЛИРОВАНИЮ И МЕТРОЛОГИИ

НАЦИОНАЛЬНЫЙ СТАНДАРТ РОССИЙСКОЙ ФЕДЕРАЦИИ

ГОСТ Р 53556.0— 2009

Звуковое вещание цифровое

КОДИРОВАНИЕ СИГНАЛОВ ЗВУКОВОГО ВЕЩАНИЯ С СОКРАЩЕНИЕМ ИЗБЫТОЧНОСТИ ДЛЯ ПЕРЕДАЧИ ПО ЦИФРОВЫМ КАНАЛАМ СВЯЗИ

Часть III (MPEG-4 AUDIO)

Основные положения

(ISO/IEC 14496-3:2009, NEQ)

Издание официальное

Москва
Стандартинформ
2020

gost.e-ecolog.ru

Предисловие

1 РАЗРАБОТАН Федеральным государственным унитарным предприятием «Ленинградский отраслевой научно-исследовательский институт связи» (ФГУП «ЛОНИИС»)

2 ВНЕСЕН Техническим комитетом по стандартизации ТК 480 «Связь»

3 УТВЕРЖДЕН И ВВЕДЕН В ДЕЙСТВИЕ Приказом Федерального агентства по техническому регулированию и метрологии от 15 декабря 2009 г. № 840-ст

4 Настоящий стандарт разработан с учетом основных нормативных положений международного стандарта ИСО/МЭК 14496-3:2009 «Информационные технологии. Кодирование аудиовизуальных объектов. Часть 3. Аудио»

5 ВВЕДЕН ВПЕРВЫЕ

6 ПЕРЕИЗДАНИЕ. Июль 2020 г.

*Правила применения настоящего стандарта установлены в статье 26 Федерального закона от 29 июня 2015 г. № 162-ФЗ «О стандартизации в Российской Федерации». Информация об изменениях к настоящему стандарту публикуется в ежегодном (по состоянию на 1 января текущего года) информационном указателе «Национальные стандарты», а официальный текст изменений и поправок — в ежемесячном информационном указателе «Национальные стандарты». В случае пересмотра (замены) или отмены настоящего стандарта соответствующее уведомление будет опубликовано в ближайшем выпуске ежемесячного информационного указателя «Национальные стандарты». Соответствующая информация, уведомление и тексты размещаются также в информационной системе общего пользования — на официальном сайте Федерального агентства по техническому регулированию и метрологии в сети Интернет (*gost.e-ecolog.ru

© Стандартинформ, оформление, 2014, 2020

Настоящий стандарт не может быть полностью или частично воспроизведен, тиражирован и распространен в качестве официального издания без разрешения Федерального агентства по техническому регулированию и метрологии

Содержание

1 Область применения 1
2 Термины, определения и сокращения 1
3 Общие положения 5
3.1 Обзор содержания стандарта 5
3.2 Концепция стандарта 5
3.3 Набор инструментов 7
Библиография 15

Введение

Комплекс стандартов ГОСТ Р 53556 «Звуковое вещание цифровое. Кодирование сигналов звукового вещания с сокращением избыточности для передачи по цифровым каналам связи. Часть III (MPEG-4 audio)» состоит из 11 отдельных взаимосвязанных стандартов, перечень которых приведен ниже, разработанных на базе международного стандарта ИСО/МЭК 14496-3:2009 «Информационные технологии. Кодирование аудиовизуальных объектов. Часть 3. Аудио».

В перечень стандартов, входящих в состав национального стандарта ГОСТ Р 53556 «Звуковое вещание цифровое. Кодирование сигналов звукового вещания с сокращением избыточности для передачи по цифровым каналам связи. Часть III (MPEG-4 audio)», входят следующие стандарты:

ГОСТ Р 53556.0 «Звуковое вещание цифровое. Кодирование сигналов звукового вещания с сокращением избыточности для передачи по цифровым каналам связи. Часть III (MPEG-4 audio). Основные положения»;

ГОСТ Р 53556.1 «Звуковое вещание цифровое. Кодирование сигналов звукового вещания с сокращением избыточности для передачи по цифровым каналам связи. Часть III (MPEG-4 audio). Общие требования к кодированию»;

ГОСТ Р 53556.2 «Звуковое вещание цифровое. Кодирование сигналов звукового вещания с сокращением избыточности для передачи по цифровым каналам связи. Часть III (MPEG-4 audio). Кодирование речевых сигналов с использованием гармонических векторов (HVXC)»;

ГОСТ Р 53556.3 «Звуковое вещание цифровое. Кодирование сигналов звукового вещания с сокращением избыточности для передачи по цифровым каналам связи. Часть III (MPEG-4 audio). Кодирование речевых сигналов с использованием линейного предсказания (CELP)»;

ГОСТ Р 53556.4 «Звуковое вещание цифровое. Кодирование сигналов звукового вещания с сокращением избыточности для передачи по цифровым каналам связи. Часть III (MPEG-4 audio). Основные методы кодирования звуковых сигналов (GA): усовершенствованное аудиокодирование (AAC), взвешивающее векторное квантование (TwinVQ), побитовое арифметическое кодирование (BSAC)»;

ГОСТ Р 53556.5 «Звуковое вещание цифровое. Кодирование сигналов звукового вещания с сокращением избыточности для передачи по цифровым каналам связи. Часть III (MPEG-4 audio). Структурированное кодирование звуковых сигналов (SA)»;

ГОСТ Р 53556.6 «Звуковое вещание цифровое. Кодирование сигналов звукового вещания с сокращением избыточности для передачи по цифровым каналам связи. Часть III (MPEG-4 audio). Интерфейс преобразования текста в речь (TTSI)»;

ГОСТ Р 53556.7 «Звуковое вещание цифровое. Кодирование сигналов звукового вещания с сокращением избыточности для передачи по цифровым каналам связи. Часть III (MPEG-4 audio). Параметрическое кодирование звуковых сигналов (HILN)»;

ГОСТ Р 53556.8 «Звуковое вещание цифровое. Кодирование сигналов звукового вещания с сокращением избыточности для передачи по цифровым каналам связи. Часть III (MPEG-4 audio). Параметрическое кодирование высококачественных звуковых сигналов (SSC)»;

ГОСТ Р 53556.9 «Звуковое вещание цифровое. Кодирование сигналов звукового вещания с сокращением избыточности для передачи по цифровым каналам связи. Часть III (MPEG-4 audio). Методы кодирования MPEG-1/2 аудио в MPEG-4. Основные технические требования»;

ГОСТ Р 53556.10 «Звуковое вещание цифровое. Кодирование сигналов звукового вещания с сокращением избыточности для передачи по цифровым каналам связи. Часть III (MPEG-4 audio). Кодирование без потерь информации (DST)».

НАЦИОНАЛЬНЫЙ СТАНДАРТ РОССИЙСКОЙ ФЕДЕРАЦИИ

Звуковое вещание цифровое

КОДИРОВАНИЕ СИГНАЛОВ ЗВУКОВОГО ВЕЩАНИЯ С СОКРАЩЕНИЕМ ИЗБЫТОЧНОСТИ ДЛЯ ПЕРЕДАЧИ ПО ЦИФРОВЫМ КАНАЛАМ СВЯЗИ

Часть III (MPEG-4 AUDIO)

Основные положения

Digital sound broadcasting. Coding of sound broadcasting signals with redundancy reduction for transfer on digital communication channels. Part III (MPEG-4 audio). Main positions

Дата введения — 2013—09—01

1 Область применения

Настоящий стандарт предназначен для применения при проектировании, вводе в эксплуатацию и техническом обслуживании каналов и трактов звукового вещания, организованных в цифровых системах передачи с цифровой обработкой сигналов по системе MPEG-4, организуемых в наземных и спутниковых линиях на магистральной, внутризоновых и местных сетях связи.

2 Термины, определения и сокращения

В настоящем стандарте применены следующие термины с соответствующими определениями:

2.1 **адаптивное предсказание:** Предсказание, в котором оценка производится с учетом спектральных характеристик дискретизируемого сигнала.

2.2 **анимация:** В MPEG-4 — метод кодирования синтетических объектов, заключающийся в однократной передаче базового статического объекта и досылке сообщений, описывающих динамические изменения объекта.

2.3 **анимация лица** (facial animation; FA): Динамические изменения лица.

2.4 **анимации параметры** (animation parameters; AP): Другие параметры анимации.

2.5 **арифметическое кодирование:** Кодирование без потерь, ставящее в соответствие передаваемой последовательности символов определенный узкий интервал чисел в пределах от 0 до 1.

2.6 **аудио:** Звук (звуковой).

2.7 **байт:** Единица двоичной информации, равная 8 бит.

2.8 **бит:** Единица двоичной информации, соответствующая передаче 0 или 1.

2.9 **буфер:** Устройство памяти большой емкости, позволяющее записывать и хранить большой объем данных.

2.10 **векторное кодирование:** Метод кодирования, при котором по каналу передается адрес в кодовой книге блока, наиболее близкого к передаваемому.

2.11 **взвешивающее векторное квантование с чередованием и преобразованием областей** (transform domain weighted interleave vector quantization; TwinVQ): Инструмент универсального звукового кодирования TwinVQ, в котором используется кодирование спектральных коэффициентов с использованием векторного квантования.

2.12 **виртуальный:** Нереальный, воображаемый, мнимый.

2.13 **виртуальные кодовые книги** (virtual codebooks; VCB): В книгах определены значения спектральных величин, при выходе за пределы которых под влиянием ошибок последние могут быть определены и скрыты.

Издание официальное

2.14 **волновые таблицы** (wavetables): Таблицы, используемые при выборке отсчетов и синтезе музыки в формате цифровых интерфейсов музыкальных инструментов (MIDI).

2.15 **восходящий поток** (upstream): Цифровой поток от приемных терминалов к передающему серверу (поток от сервера — нисходящий).

2.16 **выход за пределы** (escape): Выход за пределы максимальных значений в виртуальных кодовых книгах для различных диапазонов масштабных коэффициентов спектральных данных.

2.17 **гармоники и одиночные линии плюс шум** (harmonics and individual lines plus noise; HILN): Алгоритм параметрического кодирования на очень низких скоростях в MPEG-4.

2.18 **генератор комфортного шума** (comfort noise generator; CNG): Блок генерации фонового шума в паузах разговора в телефонии.

2.19 **группа экспертов по движущимся изображениям** (moving pictures expert group; MPEG): Группа экспертов, сформированная в 1988 г. ИСО и МЭК для выработки стандарта цифрового сжатия движущихся изображений и звуковых сигналов.

2.20 **двоичный формат звуковой сцены** (audio binary format for scenes; AudioBIFS): Набор инструментов для композиции звуковых объектов в диалоговых сценах.

2.21 **двоичный формат описания сцен** (binary format for scene; BIFS): Двоичный формат описания сцен (MPEG-4).

2.22 **двунаправленное предсказание:** Метод кодирования, использующий предсказание «вперед» (по предыдущим кадрам) и «назад» (по последующим кадрам).

2.23 **декодер:** Устройство, преобразующее сигнал в первоначальную форму, которую он имел до прохождения через кодер.

2.24 **демультиплексор:** Устройство, распределяющее потоки данных из единого цифрового потока на множество потоков к приемникам.

2.25 **дескриптор паузы** (silence insertion descriptor; SID): идентификатор (описатель, признак) передаваемого кадра тишины.

2.26 **детектор голосовой активности** (voice activity detector; VAD): Устройство, определяющее наличие речевого сигнала или фоновых шумов.

2.27 **дискретизация:** Операция построения дискретного сигнала по заданному аналоговому сигналу.

2.28 **дискретный сигнал:** Сигнал, принимающий конечные значения в некоторые дискретные моменты времени и не определенный в другие моменты времени.

2.29 **дуплексный канал:** Двунаправленный канал, в котором передача информации происходит в обоих направлениях одновременно.

2.30 **защита от ошибок** (error protection; EP): Свойство метода передачи (кодирования) обнаруживать (исправлять) ошибки в цифровом потоке каналов передачи.

2.31 **импульсно-кодовая модуляция:** Метод модуляции, заключающийся в дискретизации, квантовании и цифровом кодировании исходного сигнала.

2.32 **интернет:** В широком смысле — соединение сетей через маршрутизаторы (interworking); в более узком смысле — глобальная сеть обмена информацией, возникшая в результате объединения национальных, региональных и континентальных сетей.

2.33 **интерфейс:** Точка соединения (стык) между двумя взаимосвязанными системами.

2.34 **канальное кодирование:** Преобразование исходной цифровой последовательности в сигнал, пригодный для передачи по каналу связи.

2.35 **категория чувствительности к ошибкам** (error sensitivy categories; ESC): Признак, определяющий чувствительность к ошибкам разных частей закодированной информации.

2.36 **квантование:** Замена непрерывного интервала значений сигнала конечным множеством значений.

2.37 **код Хаффмана:** Один из кодов с переменной длиной слова, используемый в цифровом сжатии.

2.38 **кодек:** Совокупность кодера и декодера.

2.39 **кодер:** Устройство, преобразующее исходный сигнал в соответствии с заданным алгоритмом.

2.40 **кодирование методом линейного предсказания** (linear predictive coding; LPC): При кодировании этим методом по линии связи передаются вместо параметров речевого сигнала параметры некоторого фильтра, в определенном смысле эквивалентного голосовому тракту, и параметры сигнала возбуждения этого фильтра.

2.41 **кодирование речевых сигналов с использованием гармонических векторов** (harmonic vector excitation coding; HVXC): Инструмент кодирования речевых сигналов HVXC.

2.42 **кодирование с переменной длиной слова:** Техника кодирования, присваивающая словам с наибольшей вероятностью появления более короткие кодовые комбинации, а более редким символам — более длинные.

2.43 **кодирование с межкадровым предсказанием:** Метод кодирования, при котором по каналу передается разница между текущим кадром и кадром, предсказанным в кодере.

2.44 **кодирование с преобразованием:** Способ кодирования, при котором набор статистически зависимых отсчетов во временной области преобразуется в набор независимых коэффициентов при ортогональных базисных функциях в спектральной области.

2.45 **кодовые неприоритетные слова** (non-PCW): Кодовые слова, заполняющие промежутки между приоритетными кодовыми словами.

2.46 **кодовые слова приоритетные** (priority codawords; PCW): Кодовые слова, которые могут быть декодированы независимо от любой ошибки в других кодовых словах.

2.47 **композиция звуковая:** Создание единой звуковой дорожки из нескольких звуковых подпотоков, нескольких звуковых объектов, объединенных общим замыслом, идеей, местом расположения и т. п.

2.48 **контент:** Содержание, продукт творческой деятельности создателей мультимедийной продукции.

2.49 **копирование спектральной полосы** (spectral band replication; SBR): Метод передачи при кодировании, в котором передается только малое число данных, представляющих параметрическое описание высокочастотной части спектра. При декодировании в верхнюю полосу копируется нижняя; при этом вносятся изменения в соответствии с переданными параметрами для верхней полосы.

2.50 **корректирующий код:** Код, исправляющий ошибки благодаря введению некоторой избыточности в передаваемое сообщение.

2.51 **линейное предсказание с кодовым возбуждением** (code excited linear prediction; CELP): Инструмент кодирования речевых сигналов CELP.

2.52 **маршрутизатор** (router): Сетевой процессор, соединяющий каналы передачи данных без их обработки.

2.53 **маскирование:** Психоакустическое явление, заключающееся в невосприимчивости уха к слабым звукам в присутствии близких по частоте и времени более сильных звуков.

2.54 **масштабируемость:** Способность кодера и декодера формировать и обрабатывать упорядоченный набор из нескольких цифровых потоков.

2.55 **международный фонетический алфавит** (international phonetic alphabet; IPA): Система знаков для записи транскрипции на основе латинского алфавита. Разработан и поддерживается Международной фонетической ассоциацией.

2.56 **модуль прерывистой передачи** (discontinuous transmission; DTX): Модуль, который передает сигнал в периоды речевой активности одним способом, а в паузах — другим, значительно менее затратным.

2.57 **MPEG алгоритмы кодирования;** MPEG: Комплекс алгоритмов кодирования, разработанный группой экспертов по движущимся изображениям — MPEG-1, MPEG-2, MPEG-4 и т. п.

2.58 **мультимедиа:** Комплексное представление информации с использованием текста, графики, изображения, анимации и звука.

2.59 **мультиплексор:** Устройство, объединяющее потоки данных от источников в единый цифровой поток.

2.60 **неравная защита от ошибок** (unequal error protection; UEP): Метод защиты от ошибок, при котором к разным частям потока применяются разные методы в соответствии с их чувствительностью к ошибкам.

2.61 **нисходящий поток:** Поток от сервера к клиентам интерактивного канала взаимодействия.

2.62 **обратная адаптация распределения битов:** Способ распределения битов, при котором одинаковые расчеты выполняются и в кодере, и в декодере.

2.63 **обратный канал:** Восходящий поток интерактивного канала взаимодействия.

2.64 **перемежение:** Чередование. Перестановка местами соседних символов в последовательности некоторым детерминированным образом — последовательным или псевдослучайным.

2.65 **переупорядочивание кодовых слов Хаффмана** (huffman codaword reordering; HCR): Инструмент защиты от ошибок в MPEG-4 для AAC, при котором определяются и используются приоритетные и неприоритетные кодовые слова.

2.66 **перцептуальный:** Предоставленный непосредственному восприятию, наблюдению.

2.67 **побитовое арифметическое кодирование** (bit-sliced arithmetic coding; BSAC): Арифметическое кодирование с нарезанием бит (инструмент универсального звукового кодирования BSAC).

2.68 **порог маскирования:** Граница восприятия вблизи маскирующего тона.

2.69 **потоковая передача данных:** Метод последовательной передачи данных, при котором данные размещаются в полезной части пакета и могут быть дополнительно описаны в заголовке пакета.

2.70 **предсказание:** Оценка значения текущего отсчета сигнала на основе одного или нескольких предшествующих отсчетов.

2.71 **преобразование текста в речь** (text-to-speech; TTS): Метод и интерфейс преобразования текста в речь при кодировании в MPEG-4.

2.72 **профиль:** Конфигурация, совокупность используемых средств и инструментов обработки цифрового потока, определяющая, какие из этих инструментов используются вместе для различных приложений.

2.73 **прямая адаптация распределения битов:** Способ распределения битов в кодере, при котором кодер производит все расчеты и посылает результаты декодеру.

2.74 **прямая передача потока** (direct stream transfer; DST): В MPEG-4 — это инструмент кодирования звука без потерь.

2.75 **психоакустическая модель слуха:** Устройство в кодере цифрового сжатия звукоданных, вычисляющее пороги маскирования при различных комбинациях воздействующих сигналов.

2.76 **реверсивное кодирование с переменной длиной** (reversible variable length coding; RVLC): Инструмент защиты от ошибок в MPEG-4 для усовершенствованного звукового кодирования (AAC).

2.77 **саундтрек:** Звуковая дорожка — запись музыкального оформления какого-либо материала, например фильма, мультфильма или компьютерной игры; означает также оригинальную звуковую дорожку отдельно от того материала, к которому был написан.

2.78 **сервер:** Компьютер или другое устройство, подключенное к сети и предоставляющее определенные услуги другим устройствам в сети — клиентам.

2.79 **сеть:** Совокупность транспортных потоков, передаваемых в единой системе доставки.

2.80 **синусоидальное кодирование** (sinusoidal coding; SSC): Инструмент параметрического кодирования в MPEG-4.

2.81 **старший значащий бит** (most significant bit; MSB): Разряд двоичного слова с наибольшим весом.

2.82 **структурированное аудио** (structured audio; SA): В MPEG-4 — это инструмент синтеза, обеспечивает общие способы описания и нормативного создания синтетического звука, такие как SAOL, SASL и волновые таблицы MIDI.

2.83 **структурированный оркестровый язык** (structured audio orchestra language; SAOL): В MPEG-4 — язык описания музыкальных звуков.

2.84 **структурированный партитурный язык** (structured audio score language; SASL): В MPEG-4 — язык описания партитуры музыкальных инструментов, составляющих звуковой объект.

2.85 **тип объекта:** В MPEG-4 — параметр, определяющий структуру цифрового потока для одиночного объекта.

2.86 **универсальный звук** (general audio; GA): Принятая в MPEG-4 совокупность алгоритмов кодирования широкого круга звуковых объектов, базирующаяся на AAC.

2.87 **упреждающая коррекция ошибок** (forward error correction; FEC): Метод коррекции ошибок в MPEG-4.

2.88 **уровень обработки звуковых данных** (layer): Алгоритм, использующийся при обработке звуковых данных; определяется степенью сжатия и сложностью кодера и декодера.

2.89 **усовершенствованное звуковое кодирование** (advanced audio coding; AAC): Улучшенное звуковое кодирование — алгоритм цифрового сжатия звуковых сигналов.

2.90 **устойчивость к ошибкам** (error resilient; ER): Свойство метода передачи (кодирования) поддерживать допустимое качество передачи (воспроизведения) при наличии ошибок в канале передачи.

2.91 **формат загружаемых звуков уровня 2** (downloaded sounds level 2; DSL-2): Формат по стандарту ассоциации изготовителей MIDI DSL-2, в котором используются волновые таблицы.

2.92 **формат звукового потока с низкой избыточностью** (low overhead audio stream; LOAS): Формат, обеспечивающий малые издержки и описанный в MPEG-4.

2.93 **формат звукового транспортного мультиплексирования с низкой избыточностью** (low-overhead MPEG-4 audio transport multiplex; LATM): Формат MPEG-4, обеспечивающий малые издержки уплотнения и механизма транспортирования в аудиоприложениях, которые не требуют сложного кодирования.

2.94 **формат интерфейса доставки мультимедиа** (delivery multimedia integration framework; DMIF): Формат, описывающий возможности транспортного уровня и связи между транспортом, мультиплексором и функциями демультиплексора в кодерах и декодерах.

2.95 **циклический избыточный код** (cyclic redundancy check; CRC): Код, используемый в методе обнаружения ошибок в передаваемом сообщении, заключающемся в сравнении остатков от деления блоков кодовой последовательности на фиксированный делитель, производимого на передающей и на приемной стороне.

2.96 **цифровое аудиовещание** (digital audio broadcasting; DAB): Звуковое вещание, при котором используются цифровые технологии.

2.97 **цифровой интерфейс музыкального инструмента** (musical instrument digital interface; MIDI): Формат для записи, передачи и синтеза музыки, разработанный ассоциацией изготовителей MIDI.

2.98 **цифровое кодирование:** Замена квантованных значений сигнала набором двоичных символов.

2.99 **цифровое сжатие:** Процесс устранения избыточности из передаваемого цифрового сообщения.

2.100 **цифровой сигнал:** Сигнал, принимающий в дискретные моменты времени дискретные значения, выражающиеся двоичными числами.

2.101 **элементарный поток** (elementary stream; ES): Последовательность данных, которая исходит из источника и приходит через один канал к одному получателю.

2.102 **ДИКМ:** Дифференциальная импульсно-кодовая модуляция.

3 Общие положения

3.1 Обзор содержания стандарта

Комплекс стандартов ГОСТ Р 53556 является новым видом звукового стандарта, который интегрирует много различных аспектов звукового кодирования: естественный звук — с синтетическим звуком, низкую скорость передачи данных — с высококачественной передачей звука, речь — с музыкой, комплексные саундтреки — с простыми и традиционный контент — с контентом интерактивной и виртуальной реальности. Стандартизируя индивидуально-сложные инструменты кодирования и новую, гибкую оболочку для звуковой синхронизации, смешивая их с загруженной компоновкой телевизионной программы, таким образом в стандарте MPEG-4 аудио создана новая технология для нового, интерактивного мира цифрового звука.

Данный стандарт не предназначен для функционирования отдельного приложения, например телефонной связи в режиме реального времени или высококачественного звукового сжатия. ГОСТ Р 53556 — стандарт, применимый к каждому приложению: использование звукового сжатия, синтеза, манипуляции, или воспроизведение звука. Отдельные стандарты, входящие в комплекс взаимосвязанных стандартов, определяют современные инструменты кодирования в нескольких областях. Однако ГОСТ Р 53556 — больше чем только сумма его частей, поскольку инструменты, описанные здесь, интегрированы с остальной частью стандарта MPEG-4. При этом появляются новые возможности для аудиокодирования, основанные на объектах, так называемое объектное кодирование, интерактивная звуковая система, использование динамических звуковых дорожек и других видов новых средств массовой информации.

Так как в системе кодирования звуковых сигналов используется набор отдельных инструментов, то для охвата потребности широкого диапазона приложений необходима функциональная совместимость систем, которые приведены в ГОСТ Р 53556. Например, система речевой связи в режиме реального времени, использующая набор инструментов кодирования речи MPEG-4, может использовать инструментальные средства совместно с другими системами, которые работают в различных областях, но используют тот же самый инструмент — с индексацией голосовой почты и информационно-поисковой системой.

3.2 Концепция стандарта

ГОСТ Р 53556 не стандартизирует методы кодирования звука. Выбор лучшего метода создания полезных битовых потоков возлагается на авторов контента. В связи с несовершенством современных методов автоматического преобразования естественного звука в синтетический или мультиобъектный непосредственные решения принимаются в интерактивном режиме. Этот процесс похож на современные методы создания звуковых дорожек (основанных на формате MIDI [1] и многоканальном объединении).

Краткий обзор понятий, используемых в ГОСТ Р 53556, приведен ниже.

3.2.1 Средства хранения и транспортирования звуковых сигналов

Во всех инструментах звукового кодирования функции стандарта кодирования заканчиваются в точке создания доступных модулей, которые содержат сжатые данные. Технические требования к системе MPEG-4 содержат методы преобразования индивидуально закодированных доступных модулей в элементарные потоки.

В связи с наличием широкого диапазона приложений, в которых используется технология MPEG-4, в настоящее время не существует стандартного единого механизма транспортирования этих элементарных потоков по каналу связи из-за чрезмерных требований к доставке, которые можно было бы описать единственным решением. Стандартизирован интерфейс (формат интерфейса доставки мультимедиа или DMIF), описывающий возможности транспортного уровня и связь между транспортом, мультиплексором и функциями демультиплексора в кодерах и декодерах. Использование DMIF и спецификаций системы MPEG-4 требует предоставления транспортных функций намного более сложных, чем в предыдущих стандартах MPEG.

Формат звукового потока с низкой избыточностью LATM и формат звукового транспортного мультиплексирования с низкой избыточностью LOAS предназначены для того, чтобы снизить затраты на мультиплексирование и механизм транспорта сигналов для приложений естественных звуков, которые не требуют сложного кодирования, основанного на объектах, или других функций, предоставляемых системой MPEG-4.

Для того чтобы потребитель на удаленной стороне канала мог динамически управлять потоковым сервером контента MPEG-4, последний определяет потоки обратного канала, которые могут нести пользовательскую интерактивную информацию.

3.2.2 Кодирование с низкой скоростью передачи данных

Предыдущие стандарты MPEG аудио сосредоточились прежде всего на прозрачном (необнаруживаемом) или почти прозрачном кодировании высококачественного аудио с любой требуемой для этого скоростью передачи данных. MPEG-4 не только предоставляет новые и улучшенные инструменты для этой цели, но также стандартизирует инструменты, которые могут использоваться для передачи звука при низких скоростях передачи данных, приемлемых для Интернета, цифрового радио или других систем передачи с ограниченной шириной полосы. Новые инструменты, описанные в MPEG-4, являются современными инструментами, поддерживающими кодирование с низкой скоростью передачи данных речи и другого аудио.

3.2.3 Объекто-ориентированное кодирование различными инструментами

Ранее принятые стандарты MPEG предоставляли отдельный набор инструментов с различными конфигурациями для использования в разных приложениях. ГОСТ Р 53556 предоставляет несколько наборов инструментов, у которых нет специфической связи друг с другом: каждый имеет различную целевую функцию. Конфигурация (профиль) ГОСТ Р 53556 определяет, какие из этих инструментов используются вместе для различных приложений.

В предыдущих стандартах MPEG передается отдельная часть контента (возможно, многоканальная или многоязычная). ГОСТ Р 53556 поддерживает намного более гибкое понятие звуковой дорожки. При этом множественные инструменты могут использоваться для передачи нескольких звуковых объектов. При использовании множественных инструментов вместе система звуковой композиции обеспечивает создание отдельной звуковой дорожки из нескольких звуковых подпотоков. Для создания отдельной звуковой дорожки из нескольких объектов могут использоваться: взаимодействие с пользователем, возможности терминала и конфигурации громкоговорителя (динамика). Это дает ГОСТ Р 53556 существенные преимущества в качестве и гибкости по сравнению с предыдущими звуковыми стандартами.

3.2.4 Возможность воспроизведения синтетического звука

При кодировании естественного звука существующий звуковой сигнал сжимается сервером, затем передается и распаковывается в приемнике. Этот вид кодирования является предметом многих существующих стандартов для звукового сжатия. ГОСТ Р 53556 стандартизирует новую парадигму, в которой синтетические звуковые описания, включая синтетическую речь и синтетическую музыку, передаются потребителю, а затем в приемнике синтезируется звук. Такие технологии открывают новые возможности для передачи данных с низкой скоростью и высококачественным кодированием.

3.2.5 Повышение устойчивости инструментов кодирования к ошибкам

Повышение устойчивости к ошибкам всех инструментов кодирования в ГОСТ Р 53556 обеспечивается синтаксисом устойчивости к ошибкам (ER) полезного битового потока. Этот инструмент поддерживается расширенным каналом в методах кодирования, который адаптируется к специальным

требованиям данных инструментов кодирования и заданного канала связи. Этот синтаксис ER полезного битового потока обязателен для всех типов объектов ER.

Инструмент защиты от ошибок (инструмент EP) предоставляет неравную защиту от ошибок (UEP) для ГОСТ Р 53556 вместе с ER полезного битового потока. UEP — эффективный метод повышения устойчивости к ошибкам кодирующих схем. Он используется различными речевыми и звуковыми системами кодирования, работающими по каналам, подверженным ошибкам, например мобильным телефонным сетям или цифровому радиовещанию (DAB). Биты кодированного сигнала сначала группируются в различные классы согласно их чувствительности к ошибкам. Затем защита от ошибок индивидуально применяется к различным классам, давая лучшую защиту более чувствительным битам.

Повышенная устойчивость к ошибкам для AAC обеспечивается набором инструментов. Эти инструменты уменьшают восприятие искажений декодированного аудиосигнала, причиненное поврежденными битами в полезном битовом потоке.

3.2.6 Обеспечение масштабирования в системе MPEG-4 аудио

Предыдущие стандарты MPEG аудио обеспечивали определенную скорость передачи данных с помощью отдельного набора инструментов с различными конфигурациями для заданной полосы частот, указанной для использования в различных приложениях. ГОСТ Р 53556 предоставляет несколько скоростей передачи данных и полос пропускания в пределах единого потока с функциональными возможностями масштабирования. Они позволяют масштабировать данный поток в соответствии с требованиями различных каналов и приложений или делать его чувствительным к динамическим характеристикам пропускной способности. Инструменты, указанные в ГОСТ Р 53556, являются современными инструментами, предоставляющими масштабируемое сжатие речевых и других аудиосигналов.

3.3 Набор инструментов

3.3.1 Инструменты кодирования речи

3.3.1.1 Общий обзор

Инструменты кодирования речи спроектированы для передачи и декодирования синтетической и естественной речи.

ГОСТ Р 53556 предусматривает два типа инструментов кодирования речи. Инструменты для кодирования естественной речи позволяют сжать, передать и декодировать речь человека для использования в телефонной связи, персональных системах связи и системах наблюдения. Инструменты для синтетический речи обеспечивают интерфейс преобразования текста в речь в системах синтеза. Применение синтетической речи позволяет получить низкую скорость передачи данных, а также иметь встроенное подключение речи к синтетическому лицу с анимацией для использования в приложениях видео-конференц-связи с низкой скоростью передачи данных.

3.3.1.2 Кодирование естественной речи

Набор инструментов кодирования речи MPEG-4 охватывает сжатие и декодирование естественного речевого звука со скоростью передачи данных, располагающейся от 2 до 24 кбит/с. При кодировании с переменной скоростью передачи данных допускается скорость ниже 2 кбит/с, например 1,2 кбит/с.

Используются два основных метода кодирования речи: первый — алгоритм параметрического речевого кодирования — HVXC (кодирование речевых сигналов с использованием гармонических векторов) для низких скоростей передачи данных; и второй метод кодирования — CELP (линейное предсказание с кодовым возбуждением). Речевые кодеры MPEG-4 предназначаются для следующих приложений: мобильной и спутниковой связи, телефонной связи через Интернет, для компактных носителей и баз речевых данных. Они соответствуют широкому диапазону требований, охватывающих скорость передачи данных, функциональность и качество звука.

Кодирование речевых сигналов с использованием гармонических векторов HVXC

MPEG-4 работает при фиксированных скоростях передачи данных от 2,0 кбит/с до 4,0 кбит/с и использует методику масштабирования скорости передачи данных; также работает при более низких скоростях передачи данных, обычно 1,2—1,7 кбит/с, используя методику переменной скорости передачи данных. HVXC обеспечивает качество связи с малыми потерями и качество речи, близкое к качеству (near-toll-quality speech) с полосой частот 100—3800 Гц при частоте дискретизации 8 кГц. HVXC также позволяет осуществлять независимое изменение скорости и тона в процессе декодирования, что является мощной функцией для быстрого доступа к базам с речевыми данными. Функциональность HVXC включает в себя режим с фиксированной скоростью передачи данных 2,0—4,0 кбит/с и режим с переменной скоростью передачи данных — до 2,0 кбит/с включительно.

Метод ER HVXC позволяет работать в режиме с переменной скоростью передачи данных до 4,0 кбит/с при более высоком качестве кодирования. Поэтому ER HVXC обеспечивает режимы фиксированной скорости передачи данных 2,0—4,0 кбит/с, а при переменной скорости передачи данных — меньше чем 2,0 кбит/с, а также меньше чем 4,0 кбит/с в масштабируемых и немасштабируемых режимах. В режиме с переменной скоростью передачи данных в неголосовых сигналах обнаруживаются неречевые части, для которых используется меньшее число бит для кодирования, чем уменьшается средняя скорость передачи данных. Метод ER HVXC обеспечивает качество связи, близкое по восприятию звуковых сигналов (near-toll-quality) в полосе частот 100—3800 Гц при частоте дискретизации 8 кГц. Когда режим с переменной скоростью передачи данных разрешен, возможна работа с более низкой средней скоростью передачи данных. Закодированная речь, использующая режим переменной скорости со средней скоростью передачи данных 1,5 кбит/с и средней скоростью передачи данных 3,0 кбит/с, по существу, обладает тем же качеством, что и при фиксированной скорости 2,0 кбит/с и 4,0 кбит/с соответственно. Функция изменения тона и скорости во время декодирования поддерживается для всех режимов. Метод ER HVXC имеет синтаксис полезного битового потока с классами чувствительности к ошибкам, который используется с инструментом EP; при этом поддерживается функциональность маскировки ошибок для использования в каналах, подверженных ошибкам, например мобильных каналах связи. ER HVXC речевой кодер предназначается для следующего диапазона приложений: мобильной и спутниковой связи, телефонной связи через Интернет, для компактных носителей и баз речевых данных.

CELP MPEG-4 — известный алгоритм кодирования с новой функциональностью. Обычные кодеры CELP предлагают сжатие при единственной скорости передачи данных и оптимизированы для конкретных приложений. Сжатие — одна из функций, выполняемых CELP MPEG-4. Дополнительно MPEG-4 позволяет использовать один основной кодер во множестве приложений, выполнять масштабируемость скорости передачи данных и ширины полосы пропускания, а также генерировать полезные битовые потоки в произвольных скоростях передачи данных. Кодер CELP MPEG-4 поддерживает две частоты дискретизации, а именно 8 и 16 кГц. Ширина полосы пропускания составляет 100—3800 Гц для частоты дискретизации 8 кГц и 50—7000 Гц — для частоты дискретизации 16 кГц. Инструмент сжатия тишины включает в себя детектор голосовой активности (VAD), модуль прерывистой передачи (DTX) и модуль генератора комфортного шума (CNG). Инструмент кодирует/декодирует входной сигнал с более низкой скоростью передачи данных во время паузы в разговоре (тихие фреймы). Во время разговора (в речевых фреймах) используются кодирование и декодирование CELP MPEG-4.

Инструмент сжатия тишины уменьшает среднюю скорость передачи данных благодаря использованию сжатия с более низкой скоростью передачи данных для тишины. В кодере используется детектор голосовой активности для того, чтобы различить области с нормальной речевой активностью и области с тишиной или фоном. Во время нормальной речевой активности используется кодирование CELP. В других временных интервалах передается дескриптор включения тишины (SID) для более низкой скорости передачи данных. Дескриптор SID запускает генератор комфортного шума (CNG) в декодере. Амплитуда и спектральная форма комфортного шума зависит от энергии и LPC-параметров, которые определяются методами, аналогичными используемым в нормальном фрейме CELP. Эти параметры являются опционной частью SID и могут быть обновлены по требованию.

MPEG провел обширные проверочные испытания в реальных условиях прослушивания, чтобы доказать эффективность набора инструментов кодирования речи.

3.3.1.3 Интерфейс преобразования текста в речь

Возможность преобразования текста в речь (TTS) стала распространенным типом информации и играет важную роль в различных областях мультимедийного приложения. Например, используя функциональность TTS, можно создать мультимедийный контент с дикторским текстом, не записывая естественную речь. До MPEG-4 не было способа для поставщика мультимедийного контента дать команды неизвестной системе TTS. С помощью MPEG-4 стандартизирован интерфейс TTS как отдельный общий интерфейс для систем TTS. Этот интерфейс позволяет передавать речевую информацию в международном фонетическом алфавите (IPA) или в текстовой (письменной) форме любого языка.

Гибридный/многоуровневый масштабируемый интерфейс TTS MPEG-4 является развитием обычной оболочки TTS. Этот расширенный интерфейс TTS, используя просодическую информацию, взятую из естественной речи, и суммируя ее с входным текстом, генерирует наиболее высокое качество синтетической речи. Интерфейс и его формат полезного битового потока являются масштабируемыми в терминах этой добавленной информации; например, если некоторые параметры просодической информации не доступны, декодер может сгенерировать отсутствующие параметры по определенным правилам. Нормативные алгоритмы для синтеза речи и преобразование текста в фонемы в MPEG-4 не указаны, однако,

отвечая цели, которая заложена в основе интерфейса TTS MPEG-4, декодер должен полностью использовать всю предоставленную информацию согласно уровню потребностей пользователя.

Аналогично интерфейсу систем синтеза текста в речь MPEG-4 специфицирует объединенный метод кодирования фонематической информации, параметров анимации лица (FA) и других параметров анимации (AP). Используя данную методику, отдельный полезный битовый поток может использоваться для управления интерфейсом преобразования текста в речь и декодером визуального объекта анимации лица. Функциональность TTS простирается от обычного TTS до кодирования естественной речи, а в прикладных областях — от простого TTS до звуковой презентации с TTS и дублирования кинофильмов с использованием TTS.

3.3.2 Инструменты кодирования звуковых сигналов

3.3.2.1 Назначение

Инструменты кодирования звуковых сигналов разработаны для передачи и декодирования записанных музыкальных и других аудиосаундтреков.

3.3.2.2 Инструменты основного аудиокодирования

ГОСТ Р 53556 стандартизирует кодирование сигналов естественного звука при скоростях передачи данных в пределах от 6 кбит/с до нескольких сотен кбит/с на аудиоканал для моно-, двух каналов и мультиканальных стереоканалов. Общее высококачественное сжатие обеспечивается стандартом MPEG-4 AAC, представляющим собой стандарт MPEG-2 AAC с определенными усовершенствованиями. Для скорости передачи цифровых сигналов 64 кбит/с на канал и выше этот кодер прошел проверочное тестирование при строгих условиях выполнения критерия «неразличимого качества», как это определено Европейским союзом радиовещания.

Инструменты основного аудиокодирования (GA — general audio) включают набор инструментов AAC, расширенный альтернативным квантованием и схемами кодирования Twin-VQ и BSAC. Общий аудиокодер использует перцептуальный банк фильтров, сложную модель маскирования, технику формирования шума, канальное сопряжение, бесшумное кодирование и распределение бит, чтобы обеспечить максимальное сжатие при максимально возможном качестве. Стандарты психоакустического кодирования, разрабатываемые группой MPEG, представляют современное состояние этой технологии, начиная со стандарта MPEG-1 аудио. Стандарт MPEG-4, основное аудиокодирование, продолжает эту традицию.

Для скоростей передачи данных в пределах от 6 кбит/с до 64 кбит/с на канал ГОСТ Р 53556 обеспечивает расширения инструментов GA кодирования, которые позволяют владельцу контента достичь кодирования высшего качества для желательной скорости передачи данных. Более того, с GA-кодером доступны различные варианты масштабирования скорости передачи данных. Методы малой скорости передачи данных и способы масштабирования, предоставленные в пределах этого комплекта инструментов, были также проверены с помощью формальных текстов MPEG.

Функциональные возможности кодирования с малой задержкой в соответствии с ГОСТ Р 53556 позволяют распространить использование GA с низкой скоростью передачи данных на приложения, требующие малой задержки в цепочке кодирования/декодирования (например, полнодуплексная связь в реальном времени). В отличие от традиционных кодеров с малой задержкой, основанных на технологии кодирования речи, концепция этого кодера с малой задержкой основана на общем перцептуальном звуковом кодировании и таким образом является подходящей для широкого диапазона сигналов звуковой частоты. В частности — кодер получен из испытанной архитектуры MPEG-2/4 AAC, и ему доступны все возможности кодирования 2 (стерео) или больше (многоканальных) звуковых каналов кодером с малой задержкой. Он работает до частот дискретизации 48 кГц и использует меньшую длину фрейма — 512 или 480 отсчетов по сравнению с 1024 или 960 отсчетами, используемыми в стандарте MPEG-2/4 AAC для обеспечения кодирования общих сигналов звуковой частоты с алгоритмической задержкой, не превышающей 20 мс. Размер окна, используемого в анализе и синтезе банка фильтров, уменьшен в два раза. Никаких переключений блоков не используется с тем, чтобы избежать задержки «предвидения» из-за блока, переключающего решение. Для того чтобы уменьшить артефакты предэха в случае сигналов с переходными процессами, использовано переключение формы окна. Для частей сигнала без переходных процессов используется синусное окно, в то время как «окно с малым перекрытием» используется для частей с переходным процессом. Использование буфера битов в кодирующем устройстве минимизировано для получения желательной задержки. В крайнем случае никакие битовые буфера не используются вообще.

Побитовое арифметическое кодирование MPEG-4 BSAC используется в комбинации с инструментами AAC кодирования и заменяет бесшумовое кодирование квантованных спектральных данных

и масштабных коэффициентов. MPEG-4 BSAC обеспечивает тонкую градацию масштабирования с шагом 1 кбит/с на аудиоканал и с шагом 2 кбит/с — для сигнала стерео. Используются один поток базового уровня и много малых потоков уровня расширения. Для того чтобы получить тонкую масштабируемость шага, к квантованным спектральным данным применяется схема нарезания бит. Сначала квантованные спектральные значения группируют в диапазонах частот. Каждая из этих групп содержит квантованные спектральные значения в их двоичном представлении. После этого биты группы обрабатывают в секторах согласно их значениям. То есть сначала обрабатывают все старшие значащие биты (MSB) квантованных значений в группе. Эта нарезка битов при кодировании использует схему арифметического кодирования для получения энтропийного кодирования с минимальной избыточностью. Для того чтобы получить тонкую градацию масштабирования, эффективно используются инструменты MPEG-4 system; аудиоданные с тонкой градацией могут быть сгруппированы в слои большого шага, и эти слои большого шага могут быть далее сгруппированы при связывании слоев большого шага от нескольких субкадров. Кроме того, конфигурация полезной нагрузки, передаваемая по элементарному потоку (ES), может быть изменена динамически (посредством использования возможностей обратного канала MPEG-4) в зависимости от окружающей среды, такой как телефонная сеть или при интерактивном взаимодействии с пользователем. Это означает, что BSAC может обеспечить в реальном времени корректировку качества обслуживания. В дополнение к тонкой градации масштабирования это может улучшить качество аудиосигнала, который декодируется из потока, передаваемого по подверженному ошибкам каналу, такому как сеть мобильной связи или канал цифрового аудиовещания (DAB — digital audio broadcasting).

Копирование спектральной полосы MPEG-4 SBR (spectral band replication) является инструментом расширения ширины полосы, используемым в комбинации с общим аудиокодеком AAC. Объединение с кодеком AAC MPEG позволяет существенно усовершенствовать эксплуатационные показатели — понизить скорость передачи данных или улучшить аудиокачество. Это достигается копированием верхней полосы, то есть высокочастотной части спектра. Только малое число данных, представляющих параметрическое описание верхней полосы, кодируется и используется в процессе декодирования. Скорость передачи данных намного ниже скорости передачи данных, требуемой при использовании обычного AAC кодирования верхней полосы.

3.3.2.3 Инструменты параметрического кодирования звуковых сигналов

Инструмент параметрического аудиокодирования звуковых сигналов HILN (harmonic and individual lines plus noise) кодирует неречевые сигналы как музыку при скоростях передачи данных 4 кбит/с и выше, используя параметрическое представление аудиосигнала. Базовой идеей этой методики является разложение входного сигнала на аудиообъекты, которые описываются соответствующими моделями источника и представляются модельными параметрами. Модели объектов для синусоид, гармонических тонов и шума используются в кодере HILN. Кодер HILN позволяет независимо изменять скорость и тон во время декодирования.

Инструменты параметрического кодирования звуковых сигналов объединяют кодирование с очень низкой скоростью передачи данных общих сигналов аудио с возможностью изменения скорости воспроизведения или тона во время декодирования без необходимости применения блока обработки эффектов. Комбинирование инструментов кодирования речи и музыки улучшает эффективность объектно-ориентированного кодирования, которое разрешает выбор и (или) переключение между различными кодирующими методами.

Этот подход позволяет представить более расширенную модель источника, чем только предположение о стационарности сигнала в течении фрейма, которое обосновывает спектральное разложение, используемое, например, в MPEG-4 основном аудиокодере. Как известно из кодирования речи, где применяется специализированная модель источника, основанная на генерации голоса в человеческом вокальном тракте, расширенные модели источника могут быть выгодными специально для схем кодирования с очень низкой скоростью передачи данных.

Из-за очень низких скоростей передачи данных могут быть переданы параметры только для небольшого числа объектов. Поэтому используется модель восприятия, для того чтобы выбрать только те объекты, которые являются самыми важными для качества восприятия сигнала.

В кодере HILN параметры частоты и амплитуды квантованы согласно «только значимым различиям», известным из психоакустики. Спектральная огибающая шума и гармонических тонов описывается с использованием моделирования LPC, известного из кодирования речи. Корреляция между параметрами одного фрейма и между последовательными фреймами используется для предсказания параметра. Наконец, квантованные параметры подвергаются энтропийному кодированию и мультиплексированию, чтобы сформировать битовый поток.

Свойство этой схемы параметрического кодирования является результатом того, что сигнал описан в терминах параметров амплитуды и частоты. Такое представление сигнала дает функциональную возможность изменения скорости и тона простой модификацией параметра в декодере. Параметрический аудиокодер HILN может быть скомбинирован с параметрическим речевым кодером MPEG-4 (HVXC), чтобы сформировать интегрированный параметрический кодер, перекрывающий более широкий диапазон сигналов и скоростей передачи данных. Этот интегрированный кодер поддерживает изменение скорости и тона. Используя инструмент классификации речи/музыки в кодирующем устройстве, возможно автоматически выбирать HVXC для речевых сигналов и HILN — для сигналов музыки. Такое автоматическое переключение HVXC/HILN было успешно продемонстрировано, и инструмент классификации описан в информативном приложении стандарта MPEG-4.

MPEG-4 SSC (sinuSoidal coding — синусоидальное кодирование) — это инструмент параметрического кодирования, который способен к полному широкополосному высококачественному кодированию аудио. Инструмент кодирования «разбивает» монофонический или стереоаудиосигнал на ряд различных объектов так, что каждый может быть эффективно параметризован и закодирован с низкой скоростью передачи данных. Эти объекты являются переходными процессами: представленными синусоидами с динамическими изменениями во временной области; синусоиды: представленными детерминированными компонентами; шумом: представленным компонентами, у которых нет ясной временной или спектральной локализации. Поскольку сигнал представлен в параметрической области, независимое высококачественное масштабирование тона и темпа возможны при низких вычислительных затратах.

3.3.3 Инструменты аудиокодирования без потерь

Прямая передача потока MPEG-4 DST (direct stream transfer) обеспечивает кодирование без потерь оцифрованных аудиосигналов.

3.3.4 Инструменты синтеза

Инструменты синтеза разработаны для передачи с очень низкой скоростью и для синтеза на стороне терминала синтетической музыки и других звуков.

Комплект инструментов, обеспечивающий общую способность синтеза звука, называют «структурированным кодированием». SA-кодер обеспечивает общие способы описания синтетического звука и нормативного создания синтетического звука в терминале декодирования. Высококачественный стереозвук при использовании этих инструментов может быть передан на скоростях передачи данных от 0 кбит/с (без затрат) и до 2—3 кбит/с для чрезвычайно выразительного звука.

Вместо того чтобы определять специфический метод синтеза, SA определяет гибкий язык для описания метода синтеза. Эта методика предоставляет авторам контента два преимущества. Во-первых, набор доступных методов синтеза не ограничен предложениями создателей стандарта; может использоваться любой имеющийся или будущий метод синтеза в MPEG-4 структурированном аудио. Во-вторых, создание синтетического звука по структурированным описаниям является нормативным в ГОСТ Р 53556. Таким образом, звук, созданный SA-кодером, будет одинаковым на любом терминале.

Синтетический звук передается через ряд инструментальных модулей, которые могут создавать звуковые сигналы под управлением партитуры. Инструмент — маленькая сеть примитивов обработки сигналов, которые управляют параметрической генерацией звука согласно некоторому алгоритму. В единственном битовом потоке структурированного аудио могут передаваться и использоваться несколько различных инструментов. Партитура — это упорядоченный во времени набор команд, который вызывает различные инструменты в определенное время, чтобы их выводом дополнить полный поток музыки. Формат для описания инструментов — SAOL (structured audio orchestra language), структурированный звуковой язык оркестра. Формат для описания партитуры — SASL (structured audio score language), структурированный звуковой язык партитур.

Эффективная передача звуковых отсчетов с помощью так называемых волновых таблиц, используемых при выборке отсчетов при синтезе, выполняется с обеспечением функциональной совместимости со стандартом ассоциации изготовителей MIDI [2] — DLS-2 (downloaded sounds level 2), на который нормативно ссылается стандарт структурированного аудио. При использовании формата DSL-2 в MPEG-4 для саундтрека может использоваться простая и популярная техника синтеза с помощью волновых таблиц отдельно или в соединении с другими видами синтеза, используя инструменты более общего назначения. Для того чтобы в будущем обеспечить функциональную совместимость с существующим контентом и авторским инструментарием, может использоваться популярный формат управления MIDI [1] (musical instrument digital interface) вместо или в дополнение к партитуре SASL с тем, чтобы управлять синтезом.

С помощью обеспечения совместимости со стандартами MIDI [1], [2] MPEG-4 структурированное аудио таким образом представляет собой объединение существующей методики описания синтетического аудио (синтез с помощью MIDI-базированных волновых таблиц) с будущей методикой универсального алгоритмического синтеза.

3.3.5 Инструменты композиции

Инструменты композиции разработаны для объектно-ориентированного кодирования, интерактивных функциональных возможностей и аудиовизуальной синхронизации. Инструменты как для звуковой композиции, так и для визуальной композиции специфицированы в части стандарта MPEG-4 системы. Ниже представлен краткий обзор звуковых функциональных возможностей.

Звуковая композиция — это использование множества индивидуальных «звуковых объектов» и методов их объединения для создания единственного саундтрека. Это похоже на процесс записи саундтрека в мультиканальном соединении каждого музыкального инструмента, голоса актера и звуковых эффектов в их собственных каналах и затем «смешиванием» множества каналов на единственный канал или единственную стереопару. В MPEG-4 может передаваться само мультиканальное соединение с отдельными звуковыми источниками, использующими различные кодирующий инструменты, и ряд команд для смешивания также передается в битовом потоке. Поскольку получено множество звуковых объектов, они декодируются отдельно, но не воспроизводятся для слушателя, а используются команды для смешивания, чтобы подготовить единственный саундтрек из «сырого материала», данного в объектах. Этот финальный саундтрек затем и проигрывается слушателю.

Проиллюстрировать эффективность этого подхода можно таким примером: предположим, для определенного приложения мы хотим передать голос человека, говорящего в звукоотражающей окружающей среде с музыкальным фоном, в режиме стерео, с очень высоким качеством. Традиционный подход к кодированию потребовал бы использования общего кодирования звука с 32 кбит/с на канал или выше; но звуковой источник слишком сложен, чтобы быть хорошо смоделированным простым, основанным на модели кодером. Однако в MPEG-4 мы можем представить саундтрек как соединение нескольких объектов: голос человека с реверберацией добавляется к треку синтетической музыки. Мы передаем голос диктора, используя инструмент CELP при скорости 16 кбит/с, синтетическую музыку, используя инструмент SA при скорости 2 кбит/с, и позволяем малому числу ресурсов (только несколько сотен байтов в качестве фиксированной издержки) описывать стереосмешивание и реверберацию. Используя MPEG-4 и объектно-базированный подход, можно описывать потоком со скоростью менее 20 кбит/с поток, для которого могла бы потребоваться скорость 64 кбит/с, при традиционном кодировании с эквивалентным качеством.

Дополнительно наличие структурированной информации саундтрека позволяет включить в терминале декодирования более сложное взаимодействие с клиентской стороной. Например, слушателю можно предоставить (по желанию автора контента) возможность приглушить музыкальный фон. Эти функциональные возможности были бы невозможны, если бы музыка и речь были закодированы в один и тот же звуковой трек.

С двоичным форматом для сцен (BIFS — binary format for scenes) MPEG-4, определенном в части MPEG-4 системы, инструмент, названный AudioBIFS, позволяет авторам контента описывать звуковые сцены, используя эту объектно-основанную структуру, где множество источников могут смешиваться и объединяться и где предусмотрено интерактивное управление для их комбинации. Данный метод предоставляет возможность управления типовой разрешающей способностью по смешиванию. Динамическая загрузка заказных обрабатывающих сигнал программ позволяет автору контента точно запрашивать специфический или нормативный цифровой фильтр, ревербератор или другую эффект-обрабатывающую программу. Наконец, здесь предоставлен интерфейс к терминал-зависимым методам 3D-пространственного звучания для описания виртуального мира и другого 3D-звукового материала.

Поскольку AudioBIFS является частью общей спецификации BIFS, то структура используется, чтобы синхронизировать аудио и видео, аудио- и компьютерную графику или аудио с другим материалом.

3.3.6 Инструменты масштабируемости

Инструменты масштабируемости разработаны для создания битовых потоков, которые могут быть переданы без перекодирования на различных скоростях передачи данных.

Многие из типов потока в MPEG-4 масштабируются тем или иным способом. Далее рассмотрены несколько типов масштабируемости.

Масштабируемость скорости передачи данных позволяет преобразовать битовый поток в поток с более низкой скоростью передачи данных, чтобы преобразованный поток мог быть все еще декодирован

в значащий сигнал. Преобразование битового потока может осуществляться или во время передачи, или в декодере. Масштабируемость доступна для каждой схемы кодирования естественного аудио или для комбинации различных схем кодирования естественного аудио.

Масштабируемость ширины полосы — специфический случай масштабируемости скорости передачи данных, посредством которого от части битового потока, представляющей часть частотного спектра, можно отказаться во время передачи или декодирования. Это доступно для речевого кодера с алгоритмом компрессии речи CELP, уровень расширения которого преобразовывает узкую полосу речевого кодера в широкую полосу речевого кодера. Общие инструменты звукового кодирования, управляющие частотной областью, также предлагают очень гибкое управление пропускной способностью для различных уровней кодирования.

Масштабируемость сложности кодирующего устройства позволяет кодерам различной сложности генерировать правильный и значащий битовый поток. Например, для этого существует высококачественный малой сложности модуль для широкополосного кодера CELP, дающий возможность выбора между существенно меньшей сложностью кодера и оптимизированным качеством кодирования.

Масштабируемость сложности декодера позволяет данный битовый поток декодировать декодерами различных уровней сложности. Подтип масштабируемости сложности декодера — «изящное ухудшение», при котором декодер динамически контролирует доступные ресурсы и уменьшает масштаб сложности декодирования (и, таким образом, качество звука), если ресурсы ограничены. Структурированный звуковой декодер допускает этот тип масштабируемости, и автор контента может обеспечить несколько различных алгоритмов синтеза звуков фортепьяно, и сам контент, в зависимости от доступных ресурсов, решает, какой из них использовать.

3.3.7 Восходящий поток

Инструменты восходящего потока разработаны для динамического управления серверным потоком, то есть для управления скоростью передачи данных и качеством управления с обратной связью.

Восходящий поток или обратный канал позволяет пользователю на удаленной стороне управлять передачей контента с сервера. Потоки обратного канала несут пользовательскую информацию о взаимодействии.

3.3.8 Средства обеспечения устойчивости к ошибкам

3.3.8.1 Краткий обзор

Средства обеспечения устойчивости к ошибкам включают в себя инструменты, обеспечивающие как устойчивость к ошибкам, так и защиту от ошибок.

Средства обеспечения устойчивости к ошибкам обеспечивают улучшение работы на подверженных ошибкам каналах передачи. Эти средства состоят из устойчивого к ошибкам переупорядочивания потока битов общего инструмента защиты от ошибок и инструментов специфического кодирования с устойчивостью к ошибкам.

3.3.8.2 Устойчивое к ошибкам форматирование потока битов.

Устойчивое к ошибкам форматирование потока битов позволяет эффективно использовать улучшенную технологию канального кодирования — неравную защиту от ошибок (UEP — unequal error protection), которая может быть приспособлена к потребностям различных инструментов кодирования. Основная идея состоит в том, чтобы перестроить звуковое информационное наполнение фрейма в зависимости от его чувствительности к ошибкам в одной или более частях, принадлежащих к различным категориям чувствительности к ошибкам (error sensitivity categories — ESC). Эта перестановка может быть осуществлена в любых данных поэлементно или даже поразрядно. Устойчивый к ошибкам фрейм потока битов создается связыванием этих частей.

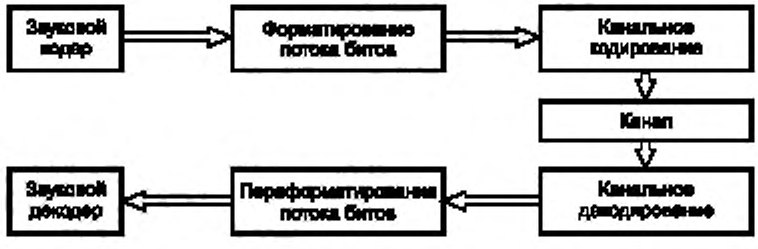

Рисунок 1— Основной принцип передачи устойчивого к ошибкам потока

Основной принцип устойчивого к ошибкам форматирования потока битов представлен на рисунке 1. Битовый поток форматируется согласно чувствительности к ошибкам отдельных элементов потока битов или отдельных битов. Этот поток кодируется, передается и декодируется в канале. Перед декодированием звука поток бит переформатируется в его первоначальное состояние.

3.3.8.3 Защита от ошибок

Инструмент защиты от ошибок обеспечивает неравную защиту от ошибок (UEP). Он получает несколько классов битов от инструментов аудиокодирования и затем применяет кодирование с упреждающей коррекцией ошибок (forward error correction — FEC) и (или) с циклическими избыточными кодами (cyclic redundancy codes — CRC) для каждого класса, согласно его чувствительности к ошибкам.

Инструмент защиты от ошибок (инструмент EP) обеспечивает неравную защиту от ошибок (UEP) и является характеристикой набора кодеков. Главными особенностями этого инструмента являются:

- обеспечение набора исправляюших/обнаруживающих ошибки кодов с широким и малым шагом масштабирования; оба типа кодов как производительны, так и избыточны;
- обеспечение универсальной структуры с эффективной широкополосной защитой от ошибок, которая охватывает и потоки фрейма фиксированной длины, и потоки фрейма переменной длины;
- обеспечение управления конфигурацией UEP с малыми издержками.

3.3.8.4 Инструменты защиты от ошибок для усовершенствованного звукового кодирования

Существуют несколько инструментов, чтобы для усовершенствованного звукового кодирования увеличить защиту от ошибок. Эти инструменты улучшают воспринятое звуковое качество декодированного звукового сигнала в случае повреждения битовых потоков, которые могут произойти, например в шумных каналах передачи.

Инструмент виртуальных кодовых книг (VCB11) расширяет информацию секционирования битового потока AAC. Это позволяет обнаружить серьезные ошибки в спектральных данных потока битов MPEG-4, AAC. Виртуальные кодовые книги используются, чтобы ограничить наибольшее абсолютное значение, возможное для любого диапазона масштабных коэффициентов; используется значение «escape» (выход за пределы). Пока все виртуальные кодовые книги используют кодовую книгу 11; шестнадцать виртуальных кодовых книг, представленных VCB11, обеспечивают шестнадцать различных ограничений спектральных величин, принадлежащих соответствующему подпункту. Поэтому ошибки, которые приводят к спектральным величинам, превышающим обозначенный предел, могут быть определены и соответственно скрыты.

Инструмент обратимого кодирования переменной длины (RVLC — reversible variable length coding) заменяет Хаффман-кодирование и дифференциально импульсно-кодово-модуляционное (ДИКМ) кодирование масштабных коэффициентов в потоке битов AAC. RVLC использует симметричные кодовые слова, чтобы дать возможность прямого и обратного декодирования данных масштабных коэффициентов. Для того чтобы иметь отправную точку для обратного декодирования, передается общее число битов RVLC части битового потока. Так как применяется ДИКМ-кодирование масштабных коэффициентов, то передается также значение последнего масштабного коэффициента для того, чтобы дать возможность обратного ДИКМ-декодирования. Так как не все узлы кодового дерева RVLC используются как кодовые слова, появляется возможность обнаружения некоторых ошибок.

Алгоритм переупорядочивания кодовых слов Хаффмана (huffman codeword reordering — HCR) для спектральных данных AAC основан на факте, что некоторые из кодовых слов могут быть помещены в известные позиции так, чтобы эти кодовые слова могли быть декодированы независимо от любой ошибки в других кодовых словах. Поэтому этот алгоритм исключает распространение ошибок в этих кодовых словах, так называемых приоритетных кодовых словах (priority codewords — PCW). Для достижения этого определены сегменты известной длины, и эти кодовые слова помещают в начало этих сегментов. Остающиеся кодовые слова (неприоритетные кодовые слова, non-PCW) заполняют промежутки, оставленные PCW, при использовании специального алгоритма, который минимизирует распространение ошибок на non-PCWs кодовые слова. Этот алгоритм переупорядочивания не увеличивает размер спектральных данных. Перед применением алгоритма переупорядочивания определяются PCW, сортируя кодовые слова согласно их важности.

Библиография

[1] MIDI 1.0. Детальная спецификация, версия 96.2 (принята Ассоциацией изготовителей MIDI, 1996 г.) (The Complete MIDI 1.0 Detailed Specification v. 96.2, MIDI Manufacturers Association, 1996)

[2] Спецификация MIDI загружаемых звуков, версия 98.2 (принята Ассоциацией изготовителей MIDI, 1998 г.) (The MIDI Downloadable Sounds Specification v. 98.2, MIDI Manufacturers Association, 1998)

УДК 621.396:006.354 ОКС 33.170

Ключевые слова: звуковое вещание, электрические параметры, каналы и тракты, технологии MPEG-кодирования, синтетический звук, масштабирование, защита от ошибок

Редактор переиздания *Е.В. Яковлева*
Технические редакторы *В.Н. Прусакова, И.Е. Черепкова*
Корректор *Е.И. Рычкова*
Компьютерная верстка *Г.В. Струковой*

Сдано в набор 13.07.2020. Подписано в печать 30.09.2020. Формат 60 × 84 1/8. Гарнитура Ариал.
Усл. печ. л. 2,33. Уч.-изд. л. 1,80.

Подготовлено на основе электронной версии, предоставленной разработчиком стандарта

ИД «Юриспруденция», 115419, Москва, ул. Орджоникидзе, 11.
gost.e-ecolog.ru gost.e-ecolog.ru

Создано в единичном исполнении во ФГУП «СТАНДАРТИНФОРМ»
для комплектования Федерального информационного фонда стандартов,
117418 Москва, Нахимовский пр-т, д. 31, к. 2.
gost.e-ecolog.ru stinfo.ru

gost.e-ecolog.ru