

Международный союз электросвязи

**МСЭ-R**  
Сектор радиосвязи МСЭ

**Рекомендация МСЭ-R BS.1196-8**  
(10/2019)

**Кодирование звуковых сигналов  
для цифрового радиовещания**

**Серия BS**  
**Радиовещательная служба (звуковая)**



Международный  
союз  
электросвязи

## Предисловие

Роль Сектора радиосвязи заключается в обеспечении рационального, справедливого, эффективного и экономичного использования радиочастотного спектра всеми службами радиосвязи, включая спутниковые службы, и проведении в неограниченном частотном диапазоне исследований, на основании которых принимаются Рекомендации.

Всемирные и региональные конференции радиосвязи и ассамблеи радиосвязи при поддержке исследовательских комиссий выполняют регламентарную и политическую функции Сектора радиосвязи.

### Политика в области прав интеллектуальной собственности (ПИС)

Политика МСЭ-R в области ПИС излагается в общей патентной политике МСЭ-T/МСЭ-R/ИСО/МЭК, упоминаемой в Резолюции МСЭ-R 1. Формы, которые владельцам патентов следует использовать для представления патентных заявлений и деклараций о лицензировании, представлены по адресу: <http://www.itu.int/ITU-R/go/patents/en>, где также содержатся Руководящие принципы по выполнению общей патентной политики МСЭ-T/МСЭ-R/ИСО/МЭК и база данных патентной информации МСЭ-R.

### Серии Рекомендаций МСЭ-R

(Представлены также в онлайн-форме по адресу: <http://www.itu.int/publ/R-REC/en>.)

Серия	Название
BO	Спутниковое радиовещание
BR	Запись для производства, архивирования и воспроизведения; пленки для телевидения
<b>BS</b>	<b>Радиовещательная служба (звуковая)</b>
BT	Радиовещательная служба (телевизионная)
F	Фиксированная служба
M	Подвижные службы, служба радиоопределения, любительская служба и относящиеся к ним спутниковые службы
P	Распространение радиоволн
RA	Радиоастрономия
RS	Системы дистанционного зондирования
S	Фиксированная спутниковая служба
SA	Космические применения и метеорология
SF	Совместное использование частот и координация между системами фиксированной спутниковой службы и фиксированной службы
SM	Управление использованием спектра
SNG	Спутниковый сбор новостей
TF	Передача сигналов времени и эталонных частот
V	Словарь и связанные с ним вопросы

*Примечание.* – Настоящая Рекомендация МСЭ-R утверждена на английском языке в соответствии с процедурой, изложенной в Резолюции МСЭ-R 1.

Электронная публикация  
Женева, 2020 г.

© ITU 2020

Все права сохранены. Ни одна из частей данной публикации не может быть воспроизведена с помощью каких бы то ни было средств без предварительного письменного разрешения МСЭ.

## РЕКОМЕНДАЦИЯ МСЭ-R BS.1196-8\*

**Кодирование звуковых сигналов для цифрового радиовещания**

(Вопрос МСЭ-R 19-1/6)

(1995-2001-2010-2012-02/2015-10/2015-2017-01/2019-10/2019)

**Сфера применения**

В настоящей Рекомендации определены системы кодирования источников звуковых сигналов, применимые для цифрового звукового и телевизионного радиовещания. Наряду с этим определена система, применимая для совместимого в обратном направлении многоканального расширения систем цифрового звукового и телевизионного радиовещания.

**Ключевые слова**

Звуковой сигнал, кодирование звуковых сигналов, вещание, цифровой, радиовещание, звуковой, телевизионный, кодек.

Ассамблея радиосвязи МСЭ,

*учитывая,*

- a)* что в Рекомендации МСЭ-R BS.1548 указаны требования пользователей к системам кодирования звуковых сигналов для цифрового радиовещания;
- b)* что в Рекомендации МСЭ-R BS.775 рассматривается многоканальная звуковая система с сопровождающим изображением и без сопровождающего изображения и что для системы цифрового радиовещания необходима высококачественная многоканальная звуковая система, в которой используется эффективное уменьшение скорости передачи;
- c)* что усовершенствованная звуковая система, определенная в Рекомендации МСЭ-R BS.2051, содержит трехмерные конфигурации каналов и использует статические или динамические метаданные для управления сигналами на основе объекта, на основе сцены и на основе канала;
- d)* что Рекомендации МСЭ-R BS.1116 рассматривается субъективная оценка звуковых систем с небольшим ухудшением качества, в том числе многоканальных звуковых систем;
- e)* что в Рекомендации МСЭ-R BS.1534 (MUSHRA) рассматривается субъективная оценка звуковых систем с промежуточным качеством звука;
- f)* что кодирование с низкой скоростью передачи для звуковых сигналов высокого качества было протестировано Сектором радиосвязи МСЭ;
- g)* что единообразие методов кодирования источников звуковых сигналов в различных службах может обеспечить повышенную гибкость систем и снижение затрат на приемные устройства;
- h)* что многие радиовещательные службы уже используют или определили использование звуковых кодеков из семейств MPEG-1, MPEG-2, MPEG-4, AC-3 и E-AC-3;
- i)* что в Рекомендации МСЭ-R BS.1548 перечислены кодеки, которые, как было продемонстрировано, отвечают требованиям радиовещательных организаций в отношении подачи, распространения и передачи;
- j)* что те радиовещательные организации, которые еще не начали предоставлять услуги, должны иметь возможность выбрать систему, которая бы наилучшим образом подходила к их применению;
- k)* что радиовещательным организациям может потребоваться принять во внимание совместимость с традиционными радиовещательными системами и оборудованием при выборе системы;

---

\* Настоящую Рекомендацию следует довести до сведения Международной организации по стандартизации (ИСО) и Международной электротехнической комиссии (МЭК).

- l) что при введении многоканальной звуковой системы необходимо принимать во внимание существующие моно- и стереоприемники;
- m) что совместимое в обратном направлении многоканальное расширение существующей системы кодирования звуковых сигналов может обеспечить большую эффективность скорости передачи, чем одновременная передача;
- n) было бы предпочтительно, чтобы система кодирования звуковых сигналов могла кодировать сигналы и речи, и музыки с одинаково высоким качеством,

*рекомендует,*

**1** чтобы для новых применений цифровой звуковой или телевизионной радиовещательной передачи, когда не требуется совместимость с традиционными передачами и оборудованием, применялась одна из следующих систем кодирования звуковых сигналов с низкой скоростью передачи:

- расширенное HE AAC, как указано в ИСО/МЭК 23003-3:2012;
- E-AC-3, как указано в ETSI TS 102 366 (2014-08);
- AC-4, как указано в ETSI TS 103 190-1 v1.3.1 (2018-02) и ETSI TS 103 190-2 v1.2.1 (2018-02);
- MPEG-H 3D Audio LC Profile, как указано в ИСО/МЭК 23008-3:2019;
- DTS-UHD, как указано в ETSI TS 103 491 v1.1.1 (2017-04);

ПРИМЕЧАНИЕ 1. – Расширенное HE AAC является более гибким расширенным набором MPEG-4 HE AAC-v2, HE AAC и AAC LC, и включает унифицированное кодирование речи и звука (USAC) MPEG-D.

ПРИМЕЧАНИЕ 2. – E-AC-3 является более гибким расширенным набором AC-3.

ПРИМЕЧАНИЕ 3. – Спецификации AC-4, MPEG-H 3D Audio LC Profile и DTS-UHD предусматривают возможность поддерживать усовершенствованную звуковую систему, описанную в Рекомендации МСЭ-R BS.2051, и для обеспечения соответствия кодеков пользователям следует ссылаться на Рекомендацию МСЭ-R BS.1548.

**2** чтобы для применений цифровой звуковой или телевизионной радиовещательной передачи, когда требуется совместимость с традиционными передачами и оборудованием, применялась одна из следующих систем кодирования звуковых сигналов с низкой скоростью передачи:

- уровень II MPEG-1, как указано в ИСО/МЭК 11172-3:1993;
- уровень II MPEG-2 при половинной частоте дискретизации, как указано в ИСО/МЭК 13818-3:1998;
- MPEG-2 AAC-LC или MPEG-2 AAC-LC при SBR, как указано в ИСО/МЭК 13818-7:2006;
- MPEG-4 AAC-LC, как указано в ИСО/МЭК 14496-3:2009;
- MPEG-4 HE AAC v2, как указано в ИСО/МЭК 14496-3:2009;
- AC-3, как указано в ETSI TS 102 366 (2014-08);

ПРИМЕЧАНИЕ 4. – ИСО/МЭК 11172-3 иногда именуется 13818-3, поскольку эта спецификация включает 11172-3 путем ссылки.

ПРИМЕЧАНИЕ 5. – Предлагается поддерживать расширенное HE AAC, как указано в ИСО/МЭК 23003-3:2012. В него входят все упомянутые выше версии AAC, гарантируя тем самым совместимость с новыми перспективными, а также с унаследованными системами радиовещания по всему миру, используя одну единственную реализацию декодера.

**3** чтобы для совместимого в обратном направлении многоканального расширения систем цифрового телевизионного и звукового радиовещания следует использовать многоканальные звуковые расширения, описанные в ИСО/МЭК 23003-1:2007;

ПРИМЕЧАНИЕ 6. – Поскольку технология MPEG Surround, описанная в ИСО/МЭК 23003-1:2007, независима от технологии сжатия (основного кодера), используемой для передачи совместимого в обратном направлении сигнала, описанные инструменты многоканального расширения могут использоваться в сочетании с любыми из систем кодирования, рекомендованными в пунктах 1 и 2 раздела *рекомендует*.

**4** чтобы для линий распространения и подачи можно было использовать одну из следующих систем кодирования со скоростью на один звуковой сигнал (то есть на моносигнал или на компонент

независимо кодируемого стереосигнала) не менее указанной для соответствующей системы, за исключением вспомогательных данных:

- уровень II MPEG-1 (см. стандарт ИСО/МЭК 11172-3) со скоростью не менее 180 кбит/с на один монофонический звуковой сигнал;
- MPEG-4 AAC (см. стандарт ИСО/МЭК IS 14496-3) со скоростью не менее 144 кбит/с на один монофонический звуковой сигнал при наличии в системе до 5 каскадов;
- AC-4, как указано в ETSI TS 103 190-1 v1.3.1 (2018-02) и ETSI TS 103 190-2 v1.2.1 (2018-02), со скоростью не менее 128 кбит/с на один монофонический звуковой сигнал при наличии в системе до пяти каскадов;
- MPEG-H 3D Audio, как указано в ИСО/МЭК 23008-3, со скоростью не менее 144 кбит/с на один монофонический звуковой сигнал при наличии в системе до пяти каскадов;

**5** чтобы для комментаторских линий использовалось кодирование уровня III ИСО/МЭК 11172-3 на скорости не менее 60 кбит/с, за исключением вспомогательных данных для моносигналов, и не менее 120 кбит/с для стереосигналов, используя совместное стереокодирование;

**6** чтобы для применений высокого качества частота дискретизации составляла 48 кГц;

**7** чтобы входной сигнал кодера звука с низкой цифровой скоростью был свободен от предискажений и чтобы в кодере также не применялось никаких предискажений,

*далее рекомендует*

**1** обращаться к последней версии Рекомендации МСЭ-R BS.1548 за сведениями о конфигурациях систем кодирования, которые, как было доказано, отвечают требованиям подачи, распространения и передачи;

**2** что необходимо провести дальнейшие исследования требований для усовершенствованной звуковой системы, определенной в последней версии Рекомендации МСЭ-R BS.2051, и по завершении этих исследований обновить настоящую Рекомендацию.

ПРИМЕЧАНИЕ. – Информацию о кодеках, включенных в настоящую Рекомендацию, можно найти в Приложениях 1–8.

## **Приложение 1 (информационное)**

### **Звуковые MPEG-1 и MPEG-2, уровни II и III**

#### **1 Кодирование**

Кодер обрабатывает цифровой звуковой сигнал и выдает сжатый поток битов. Алгоритм кодера не стандартизирован, и поэтому могут использоваться различные средства кодирования, такие как определение порога слухового маскирования, квантование и масштабирование (Примечание 1). Вместе с тем выход кодера должен быть таким, чтобы декодер, соответствующий настоящей Рекомендации, вырабатывал звуковой сигнал, подходящий для запланированного применения.

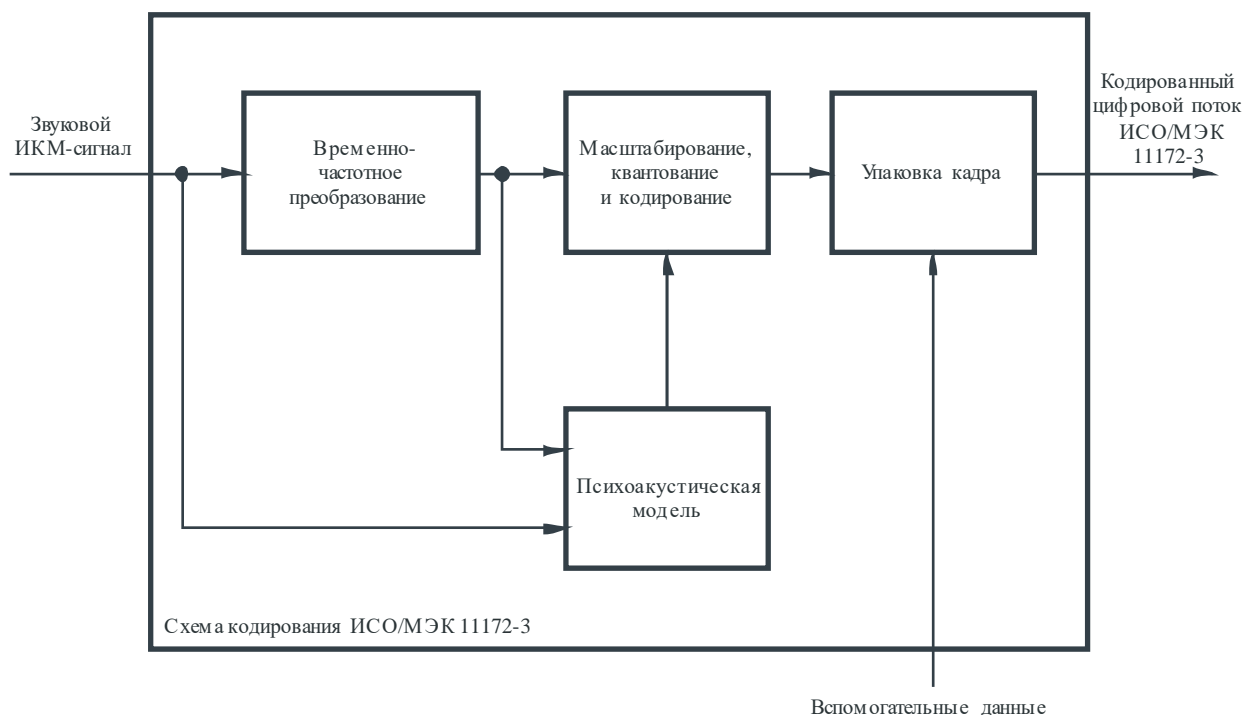
ПРИМЕЧАНИЕ 1. – Кодер, отвечающий описанию, приведенному в Приложениях С и D к Документу ИСО/МЭК 11172-3, 1993 год, будет удовлетворять минимальным требованиям к показателям работы.

Описание, представленное ниже, относится к типовому кодеру, показанному на рисунке 1. Входные отсчеты звукового сигнала подаются на кодер. При временно-частотном преобразовании создается фильтрованное и субдискретизированное представление входного звукового потока. Отображенными отсчетами могут быть либо отсчеты в подполосах (как в уровнях I или II, см. ниже), либо трансформированные отсчеты в подполосах (как в уровне III). Психоакустическая модель, в которой

используется быстрое преобразование Фурье, действующая параллельно с временно-частотным преобразованием звукового сигнала, создает набор данных для управления процессами квантования и кодирования. Эти данные различаются в зависимости от реального выполнения схемы кодера. Одной из возможностей является использование оценки порога маскирования для управления работой квантователя. Блок масштабирования, квантования и кодирования создает набор кодированных символов из преобразованных входных отсчетов. Необходимо еще раз отметить, что передаточная функция этого блока зависит от реализации системы кодирования. Блок "упаковка кадра" собирает реальный поток битов для выбранного уровня из выходных данных других блоков (например, данных по распределению битов, коэффициентов масштабирования, кодированных отсчетов в подполосах) и, при необходимости, добавляет другую информацию в поле вспомогательных данных (например, о защите от ошибок).

РИСУНОК 1

Блок-схема типового кодера



BS.1196-8-01

## 2 Уровни

В зависимости от применения могут использоваться различные уровни системы кодирования повышенной сложности и с улучшенными показателями работы.

*Уровень I.* На этом уровне производится основное преобразование входных цифровых звуковых сигналов в 32 подполосы, фиксированное разбиение на сегменты для форматирования данных в блоки, определение адаптивного распределения битов и квантование с использованием компрессии и форматирования блоков с помощью психоакустической модели. В одном цикле уровня I представлены 384 отсчета на канал.

*Уровень II.* На этом уровне обеспечивается дополнительное кодирование распределения битов, содержатся коэффициенты масштабирования и отсчеты. В одном цикле уровня II представлены  $3 \times 384 = 1152$  отсчета на канал.

*Уровень III.* На этом уровне вводится повышенное разрешение по частоте, основанное на использовании гибридного блока фильтров (блок фильтров на 32 подполосы с переменной длиной, определяющей дискретные коэффициенты косинусоидального преобразования). На этом уровне

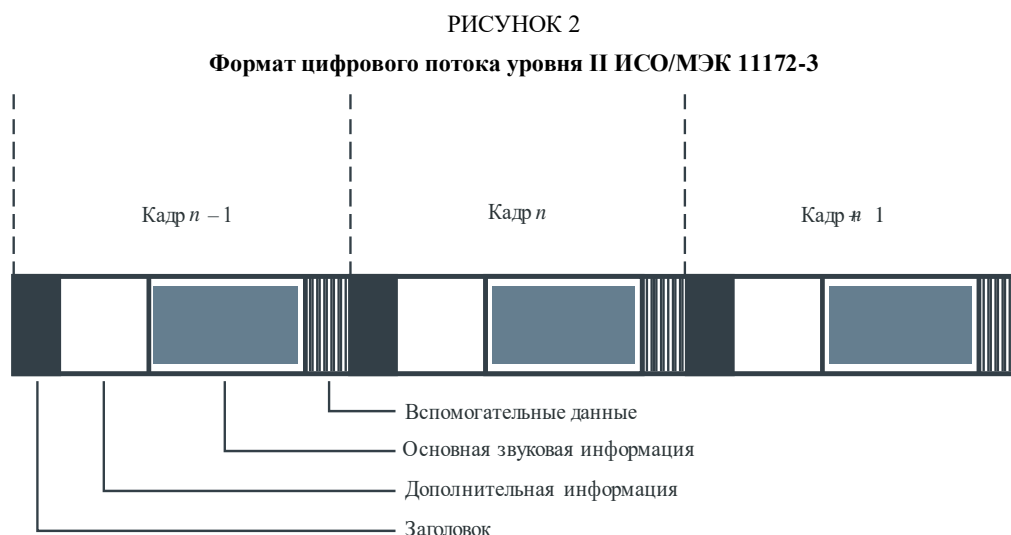
добавляются неоднородный квантователь, адаптивное разбиение на сегменты и энтропийное кодирование квантованных значений. В одном цикле уровня III представлены 1152 отсчета на канал.

Существуют четыре различных режима, возможных на любом из уровней:

- одиночный канал;
- двойной канал (два независимых звуковых сигнала, кодированных в одном цифровом потоке, например при двуязычном применении);
- стереофония (левый и правый сигналы стереофонической пары, кодированные в одном цифровом потоке);
- комбинированная стереофония (левый и правый сигналы стереофонической пары, кодированные в одном цифровом потоке с данными об используемой стереофонической рассогласованности и избыточности). Режим комбинированной стереофонии может использоваться для повышения качества звука при низких цифровых скоростях и/или для снижения скорости передачи данных для стереофонических сигналов.

### 3 Формат кодированного цифрового потока

Обзор цифрового потока ИСО/МЭК 11172-3 приведен на рисунке 2 для уровня II и на рисунке 3 для уровня III. Кодированный поток битов состоит из последовательных кадров. В зависимости от уровня кадр включает следующие поля.

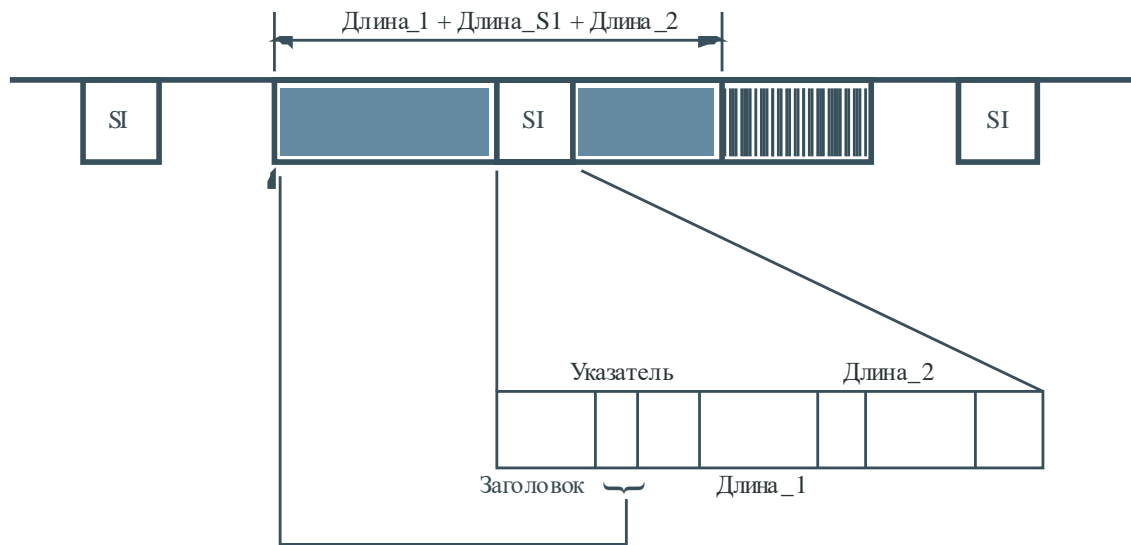


Уровень II

- Заголовок — часть цифрового потока, содержащая информацию о синхронизации и статусе
- Дополнительная информация — часть цифрового потока, содержащая информацию о распределении битов и коэффициентах масштабирования
- Основная звуковая информация — часть цифрового потока, содержащая кодированные отсчеты поддиапазонов
- Вспомогательные данные — часть цифрового потока, содержащая сведения, определяемые пользователем

РИСУНОК 3

## Формат цифрового потока уровня III ИСО/МЭК 11172-3



 Основная звуковая информация

 Вспомогательные данные

## Уровень III

Дополнительная информация	– часть цифрового потока, содержащая заголовок, указатель, длину $_1$ и длину $_2$ , сведения о коэффициенте масштабирования и т. д.
Заголовок	– часть цифрового потока, содержащая информацию о синхронизации и статусе
Указатель	– указывает на начало основной звуковой информации
Длина $_1$	– длина первой части основной звуковой информации
Длина $_2$	– длина второй части основной звуковой информации
Основная звуковая информация	– часть цифрового потока, содержащая кодированный звук
Вспомогательные данные	– часть цифрового потока, содержащая сведения, определяемые пользователем

BS.1196-8-03

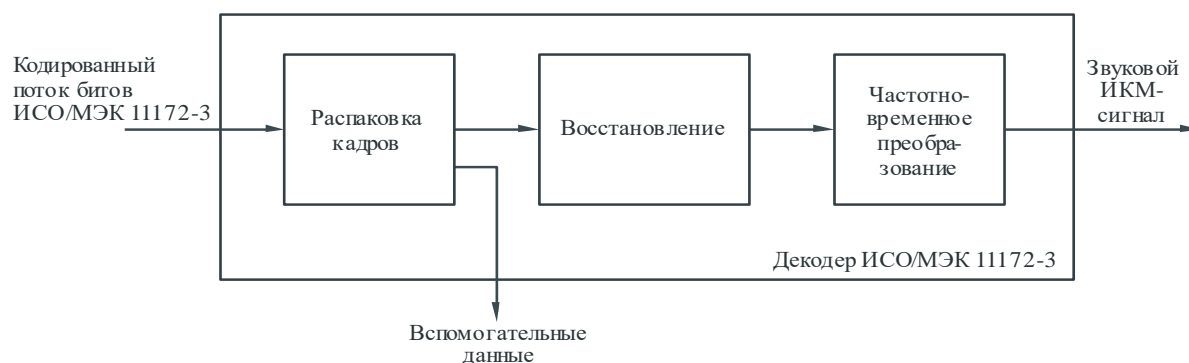
#### 4 Декодирование

Декодер принимает кодированные цифровые потоки звуковых сигналов в синтаксической структуре, определенной в Документе ИСО/МЭК 11172-3, декодирует элементы данных и использует информацию для создания выходного цифрового звукового сигнала.

Кодированный звуковой цифровой поток поступает на декодер. В процессе распаковки и декодирования дополнительно производится обнаружение ошибок, если проверка ошибок применялась в кодере. Цифровой поток распаковывается для восстановления различных частей информации, таких как заголовок звукового цикла, распределение битов, коэффициенты масштабирования, преобразованные отсчеты и, в необязательном порядке, вспомогательные данные. В процессе восстановления восстанавливается квантованная версия набора преобразованных отсчетов. Частотно-временное преобразование переводит эти преобразованные отсчеты обратно в линейные звуковые отсчеты ИКМ.



РИСУНОК 4  
Блок-схема декодера



BS.1196-8-04

## Приложение 2 (информационное)

### Звуковые MPEG-2 и MPEG-4 AAC

#### 1 Введение

В стандарте ИСО/МЭК 13818-7 описываются звуковые стандарты, не имеющие совместимости в обратном направлении, которые называются перспективным звуковым кодированием (AAC) MPEG-2. Этот стандарт является стандартом для реализации многоканальных систем более высокого качества по сравнению с качеством, достигаемым при необходимости обеспечения совместимости в обратном направлении с MPEG-1.

Система AAC имеет три профиля для предоставления компромисса между требуемой памятью, мощностью обработки и качеством звука.

– *Основной профиль*

Основной профиль обеспечивает более высокое качество звука на любой данной скорости передачи данных. Для обеспечения высокого качества звука могут быть использованы все средства, за исключением регулировки усиления. Требуемая память и мощность обработки выше, чем в случае использования профиля LC. Декодер основного профиля может декодировать цифровой поток, кодированный с использованием профиля LC.

– *Профиль пониженной сложности (LC)*

Требуемая мощность обработки и память при использовании профиля LC меньше, чем при применении основного профиля, в то время как качество работы сохраняется высоким. Профиль LC не имеет предсказателя и средства регулировки усиления, но порядок временного ограничения шума (TNS) ограничен.

– *Профиль масштабируемой частоты дискретизации (SSR)*

Профиль SSR может предоставить масштабируемый частотный сигнал с устройством регулировки усиления. При его использовании можно выбрать полосы частот для декодирования, таким образом требуется меньшее аппаратное обеспечение для декодера. Например, для декодирования на частоте дискретизации 48 кГц только самой низкой полосы частот декодер может воспроизводить полосу частот звукового сигнала шириной 6 кГц при минимальной сложности декодирования.

Как показано в таблице 1, система AAC поддерживает 12 типов частот дискретизации в диапазоне от 8 до 96 кГц и до 48 звуковых каналов. В таблице 2 показаны конфигурации каналов по умолчанию, которые включают монофонию, два канала, пять каналов (три основных/два задних канала), пять каналов с каналом низкочастотных эффектов (НЧЭ) (ширина полосы менее 200 Гц) и т. д. Кроме конфигураций по умолчанию, можно задать число громкоговорителей на каждой позиции (основной, боковой и задней), позволяя реализовать гибкую организацию многоканальной громкоговорящей системы. Поддерживается также возможность уменьшения числа каналов. Пользователь может установить коэффициент для сведения многоканальных звуковых сигналов в двухканальный сигнал. Поэтому качество звука может контролироваться с использованием устройства воспроизведения, имеющего только два канала.

ТАБЛИЦА 1

## Поддерживаемые частоты дискретизации

Частота дискретизации (Гц)
96 000
88 200
64 000
48 000
44 100
32 000
24 000
22 050
16 000
12 000
11 025
8 000

ТАБЛИЦА 2

Конфигурации каналов по умолчанию <sup>(1)</sup>

Значение <sup>(2)</sup>	Число громкоговорителей	Звуковые синтаксические элементы, перечисленные в порядке приема	Отображение по умолчанию элемента на громкоговоритель <sup>(3)</sup>	Название канала, определенное в Рекомендации МСЭ-R BS.775 или BS.2051 <sup>(4)</sup>
1	1	single_channel_element одноканальный_элемент	M + 000	Моно
2	2	channel_pair_element элемент_пары_каналов	M + 030, M – 030	Левый, правый
3	3	single_channel_element() одноканальный_элемент()	M + 000	Центральный
		channel_pair_element() элемент_пары_каналов()	M + 030, M – 030	Левый, правый
4	4	single_channel_element() одноканальный_элемент()	M + 000	Центральный
		channel_pair_element() элемент_пары_каналов()	M + 030, M – 030	Левый, правый

		single_channel_element() одноканальный_элемент()	M + 180	Моно с эффектом объемного звука
--	--	---	---------	---------------------------------

ТАБЛИЦА 2 (продолжение)

Значение <sup>(2)</sup>	Число громкоговорителей	Звуковые синтаксические элементы, перечисленные в порядке приема	Отображение по умолчанию элемента на громкоговоритель <sup>(3)</sup>	Название канала, определенное в Рекомендации МСЭ-R BS.775 или BS.2051 <sup>(4)</sup>
5	5	single_channel_element() одноканальный_элемент()	M + 000	Центральный
		channel_pair_element() элемент_пары_каналов()	M + 030, M – 030	Левый, правый
		channel_pair_element() элемент_пары_каналов()	M + 110, M – 110	Левый с объемным звуком, правый с объемным звуком
6	5 + 1	single_channel_element() одноканальный_элемент()	M + 000	Центральный
		channel_pair_element() элемент_пары_каналов()	M + 030, M – 030	Левый, правый
		channel_pair_element() элемент_пары_каналов()	M + 110, M – 110	Левый с объемным звуком, правый с объемным звуком
		lfe_element() элемент_НЧЭ()	LFE1	Низкочастотных эффектов
7	7 + 1 фронтальный	single_channel_element() одноканальный_элемент()	M + 000	Н/П <sup>(5)</sup>
		channel_pair_element() элемент_пары_каналов()	M + 030, M – 030	
		channel_pair_element() элемент_пары_каналов()	M + 045, M – 045	
		channel_pair_element() элемент_пары_каналов()	M + 110, M – 110	
		lfe_element() элемент_НЧЭ()	LFE1	
8–10	–	–	Зарезервирован	–
11	6 + 1	single_channel_element() одноканальный_элемент()	M + 000	Н/П
		channel_pair_element() элемент_пары_каналов()	M + 030, M – 030	
		channel_pair_element() элемент_пары_каналов()	M + 110, M – 110	
		single_channel_element() одноканальный_элемент()	M + 180	
		lfe_element() элемент_НЧЭ()	LFE1	
12	7 + 1 тыловой	single_channel_element() одноканальный_элемент()	M + 000	Центральный
		channel_pair_element() элемент_пары_каналов()	M + 030, M – 030	Левый, правый
		channel_pair_element() элемент_пары_каналов()	M + 090, M – 090	Левый боковой с объемным звуком, правый боковой с объемным звуком
		channel_pair_element() элемент_пары_каналов()	M + 135, M – 135	Левый задний с объемным звуком, правый задний с объемным звуком
		lfe_element() элемент_НЧЭ()	LFE1	Низкочастотных эффектов

ТАБЛИЦА 2 (окончание)

Значение <sup>(2)</sup>	Число громкоговорителей	Звуковые синтаксические элементы, перечисленные в порядке приема	Отображение по умолчанию элемента на громкоговоритель <sup>(3)</sup>	Название канала, определенное в Рекомендации МСЭ-R BS.775 или BS.2051 <sup>(4)</sup>
13	22 + 2	single_channel_element() одноканальный_элемент()	M + 000	Фронтальный центральный
		channel_pair_element() элемент_пары_каналов()	M + 030, M – 030	Фронтальный левый центральный, фронтальный правый центральный
		channel_pair_element() элемент_пары_каналов()	M + 060, M – 060	Фронтальный левый, фронтальный правый
		channel_pair_element() элемент_пары_каналов()	M + 090, M – 090	Левый боковой, правый боковой
		channel_pair_element() элемент_пары_каналов()	M + 135, M – 135	Тыловой левый, тыловой правый
		single_channel_element() одноканальный_элемент()	M + 180	Тыловой центральный
		lfe_element() элемент_НЧЭ()	LFE1	Низкочастотных эффектов-1
		lfe_element() элемент_НЧЭ()	LFE2	Низкочастотных эффектов-2
		single_channel_element() одноканальный_элемент()	U + 000	Верхний фронтальный центральный
		channel_pair_element() элемент_пары_каналов()	U + 045, U – 045	Верхний фронтальный левый, верхний фронтальный правый
		channel_pair_element() элемент_пары_каналов()	U + 090, U – 090	Верхний левый боковой, верхний правый боковой
		single_channel_element() одноканальный_элемент()	T + 000	Верхний центральный
		channel_pair_element() элемент_пары_каналов()	U + 135, U – 135	Верхний тыловой левый, верхний тыловой правый
		single_channel_element() одноканальный_элемент()	U + 180	Верхний тыловой центральный
		single_channel_element() одноканальный_элемент()	B + 000	Нижний фронтальный центральный
channel_pair_element() элемент_пары_каналов()	B + 045, U – 045	Нижний фронтальный левый, нижний фронтальный правый		
14	7 + 1 верхний	single_channel_element() одноканальный_элемент()	M + 000	Центральный
		channel_pair_element() элемент_пары_каналов()	M + 030, M – 030	Левый, правый
		channel_pair_element() элемент_пары_каналов()	M + 110, M – 110	Левый с объемным звуком, правый с объемным звуком
		lfe_element() элемент_НЧЭ()	LFE1	Низкочастотных эффектов
		channel_pair_element() элемент_пары_каналов()	U + 030, U – 030	Левый верхний фронтальный, правый верхний фронтальный
15	–	–	Зарезервирован	–

(1) Список приведен из таблицы 1.19 ISO/IEC 14496-3:2009/Amd.4:2013.

(2) Конфигурация звуковых выходных каналов указывается четырехбитовым полем, в котором содержится значение конфигурации каналов, определенное в ISO/IEC 23001-8:2013 "Не зависящие от метода кодирования кодовые точки". MPEG-2 применяется к значениям конфигурации каналов, не превышающим 7. MPEG-4 AAC применяется к значениям конфигурации каналов, не превышающим 15.

(3) Идентификация громкоговорителей с помощью обозначений, соответствующих Рекомендации МСЭ-R BS.2051.

(4) Следует обратить внимание на то, что обозначения и названия каналов зависят от фактической конфигурации каналов.

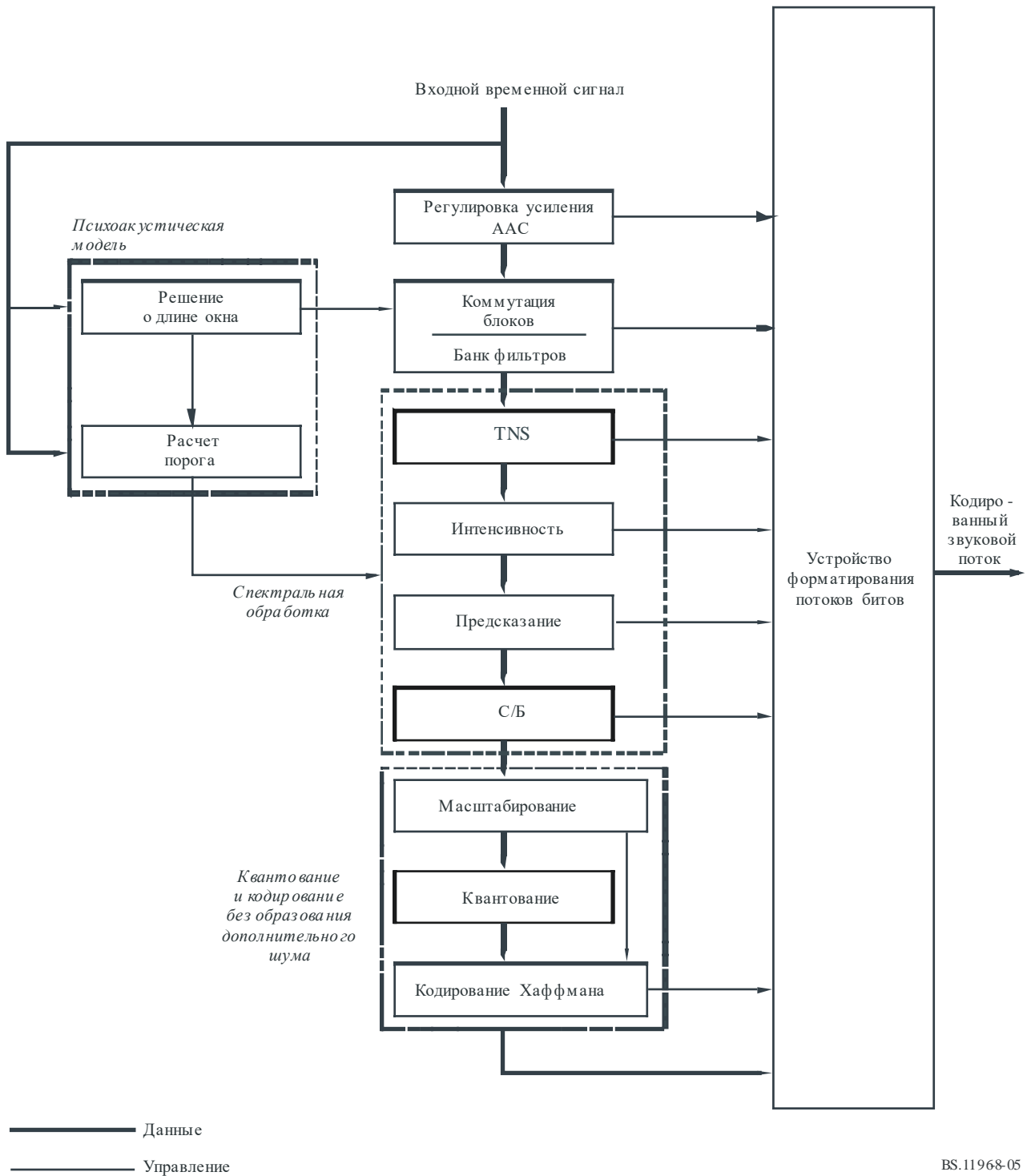
(5) Н/П – неприменимо; конфигурация каналов отсутствует в Рекомендации МСЭ-R BS.2051 или Рекомендации МСЭ-R BS.775.

## 2 Кодирование

Основная структура кодера AAC MPEG-2 показана на рисунке 5. В систему AAC входят следующие средства кодирования.

- Регулировка усиления – регулировка усиления разбивает входной сигнал на четыре равноудаленные полосы частот. Регулировка усиления используется в профиле SSR.
- Банк фильтров – модифицированное дискретное косинусное преобразование (MDCT), реализуемое банком фильтров, разлагает входной сигнал на спектральные составляющие подвыборки с частотным разрешением, равным 23 Гц, и временным разрешением, равным 21,3 мс (128 спектральных составляющих), или с частотным разрешением в 187 Гц и временным разрешением в 2,6 мс (1024 спектральных составляющих) при дискретизации, равной 48 кГц. Форма окна выбирается между двумя альтернативными формами окон.
- Временное ограничение шума (TNS) – после анализа, реализуемого банком фильтров, выполняется операция TNS. Метод TNS позволяет кодеру осуществлять контроль над тонкой временной структурой шума квантования.
- Кодирование средних/боковых (С/Б) стереосигналов и кодирование интенсивности стереосигналов – кодирование интенсивности стереосигналов и кодирование С/Б стереосигналов может быть применено для многоканальных звуковых сигналов. Для сокращения передаваемой информации о направлении при кодировании интенсивности стереосигналов передается только энергетическая огибающая. Вместо передачи изначального левого и правого сигналов при кодировании С/Б стереосигналов могут передаваться нормированная сумма (С как посередине) и разница сигналов (Б как сбоку).
- Предсказание – для снижения избыточности стационарных сигналов осуществляется предсказание временной области между спектральными составляющими последующих кадров в подвыборках.
- Квантование и помехоустойчивое кодирование – в устройстве квантования используется нелинейный квантователь с шагом размером в 1,5 дБ. Кодирование Хаффмана применяется для квантованного спектра, различных коэффициентов масштабирования и информации о направлении.
- Устройство форматирования потока битов – наконец, устройство форматирования потока битов используется для уплотнения потока битов, который состоит из квантованных и кодированных спектральных коэффициентов и некоторой дополнительной информации, поступающей от каждого инструмента.
- Психоакустическая модель – текущий порог маскирования рассчитывается с использованием психоакустической модели входного сигнала. Применяется психоакустическая модель, аналогичная модели 2 ИСО/МЭК 11172-3. Отношение сигнал/шум, получаемое на основании порога маскирования, и входной уровень сигнала используются в процессе квантования для минимизации слышимого шума квантования и, кроме того, для выбора соответствующего средства кодирования.

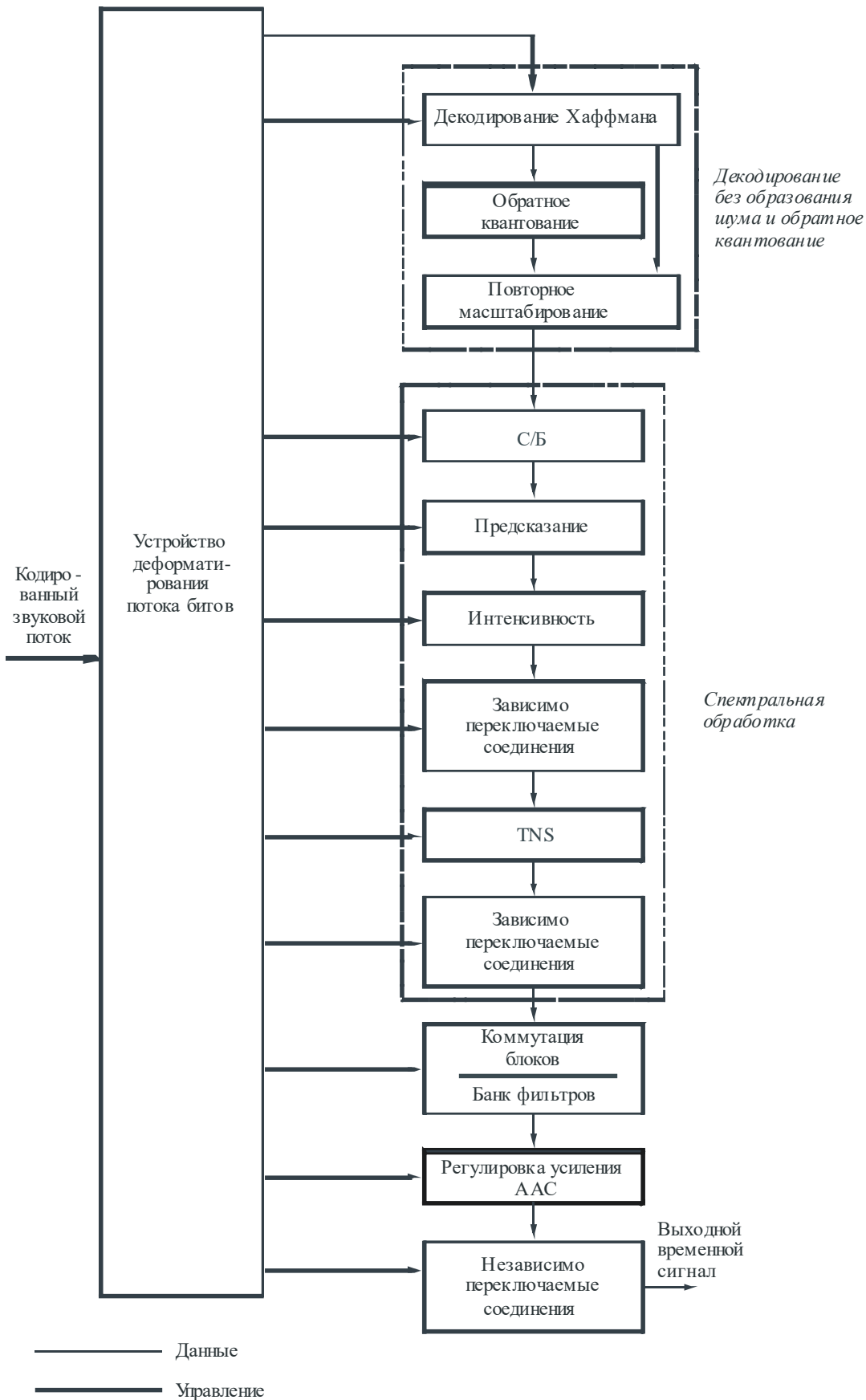
РИСУНОК 5  
Функциональная схема кодера AAC MPEG-2



### 3 Декодирование

Основная структура декодера AAC MPEG-2 показана на рисунке 6. В основном процесс декодирования является процессом, обратным кодированию.

РИСУНОК 6  
Функциональная схема декодера AAC MPEG-2

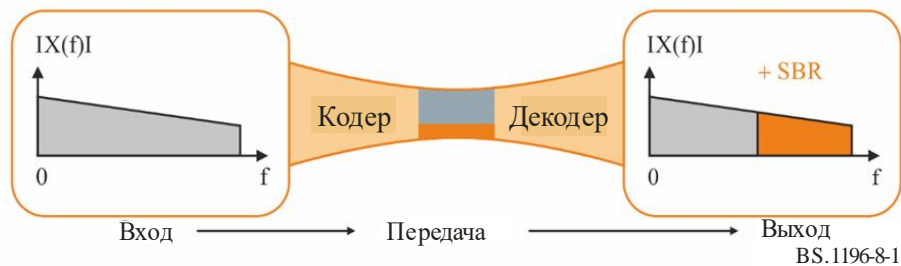


Функции декодера состоят в обнаружении описания квантованного звукового спектра в потоке битов, декодировании квантованных значений и другой информации о восстановлении квантованных спектров, обработке восстановленных спектров с помощью любого устройства, действующего в потоке битов, в целях получения реального спектра сигнала, описываемого входным потоком битов, и, наконец, преобразовании спектров частотной области во временную область с или без использования дополнительных средств регулировки усиления. Следуя цели первоначального восстановления и масштабирования восстановления спектра, имеется много дополнительных средств, которые изменяют один или более из спектров для обеспечения более эффективного кодирования. Возможность "сквозного прохода" сохраняется для каждого из дополнительных средств, работающих в спектральной области, и для всех случаев, когда спектральная операция пропускается, спектры на его входе проходят непосредственно через устройство без изменения.

#### 4 Высокоэффективное AAC и дублирование спектральной полосы

При высокоэффективном AAC (HE AAC) осуществляется дублирование спектральной полосы (SBR). SBR представляет собой метод высокоэффективного кодирования высоких частот в алгоритмах сжатия звука. Он позволяет повысить показатели работы низкоскоростных кодеков звука и речи, либо увеличивая ширину полосы звука на данной скорости, либо повышая эффективность кодирования на заданном уровне качества.

Кодируется и передается только нижняя часть спектра. Человеческое ухо наиболее чувствительно именно к этой части спектра. Вместо передачи верхней части спектра SBR используется как процесс посткодирования для восстановления более высоких частот на основании анализа переданных более низких частот. Точное восстановление обеспечивается передачей связанных с SBR параметров в кодированном потоке битов на очень низкой скорости передачи данных.



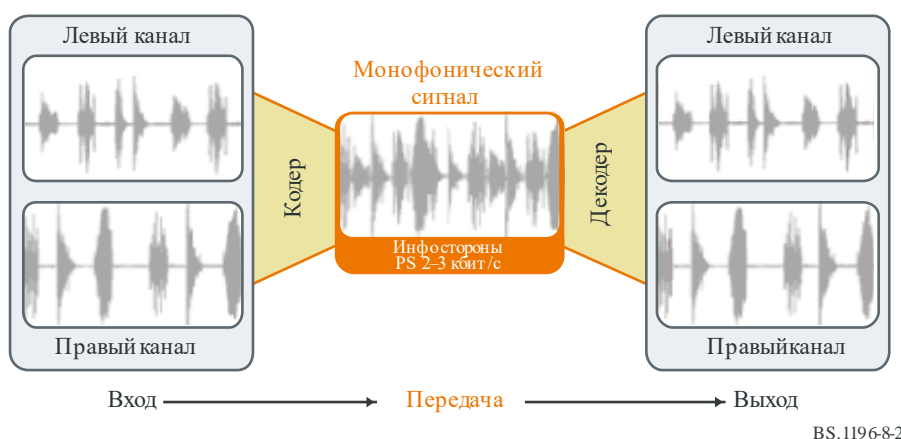
Поток битов HE AAC представляет собой расширение потока битов звука AAC. Дополнительные данные SBR встраиваются в элемент заполнения AAC, тем самым гарантируя совместимость со стандартом AAC. Технология HE AAC является двухскоростной системой. Совместимый в обратном направлении обычный поток битов звука AAC направляется на половинной частоте дискретизации расширения SBR, таким образом декодер AAC, который не способен декодировать данные расширения SBR, выдаст выходной сигнал времени на частоте дискретизации, составляющей половину от производимой декодером HE AAC.

#### 5 Высокоэффективное AAC версии 2 и параметрическое стерео

HE AAC v2 является расширением HE AAC и представляет параметрическое стерео (PS) для повышения эффективности сжатия звука для стереосигналов низкой скорости.

Кодер анализирует звуковой стереосигнал и конструирует параметрическое представление стереоизображения. Теперь нет необходимости передавать оба канала, и кодируется только монофоническое воспроизведение первоначального стереосигнала. Этот сигнал передается вместе с параметрами, необходимыми для восстановления стереоизображения.





В результате воспринимаемое качество звука низкоскоростного потока битов звука (например, 24 кбит/с), включающего параметрическое стерео, значительно выше, чем качество аналогичного потока битов без параметрического стерео.

Поток битов HE AAC v2 строится на потоке битов HE AAC. Дополнительные данные параметрического стерео встраиваются в элемент расширения монопотока HE AAC, тем самым гарантируя совместимость с HE AAC, равно как и с AAC.

Декодер HE AAC, неспособный декодировать расширение параметрического стерео, выдает монофонический выходной сигнал на полную ширину полосы. Обычный декодер AAC, неспособный декодировать данные расширения SBR, выдает монофонический выходной сигнал времени на половинной частоте дискретизации.

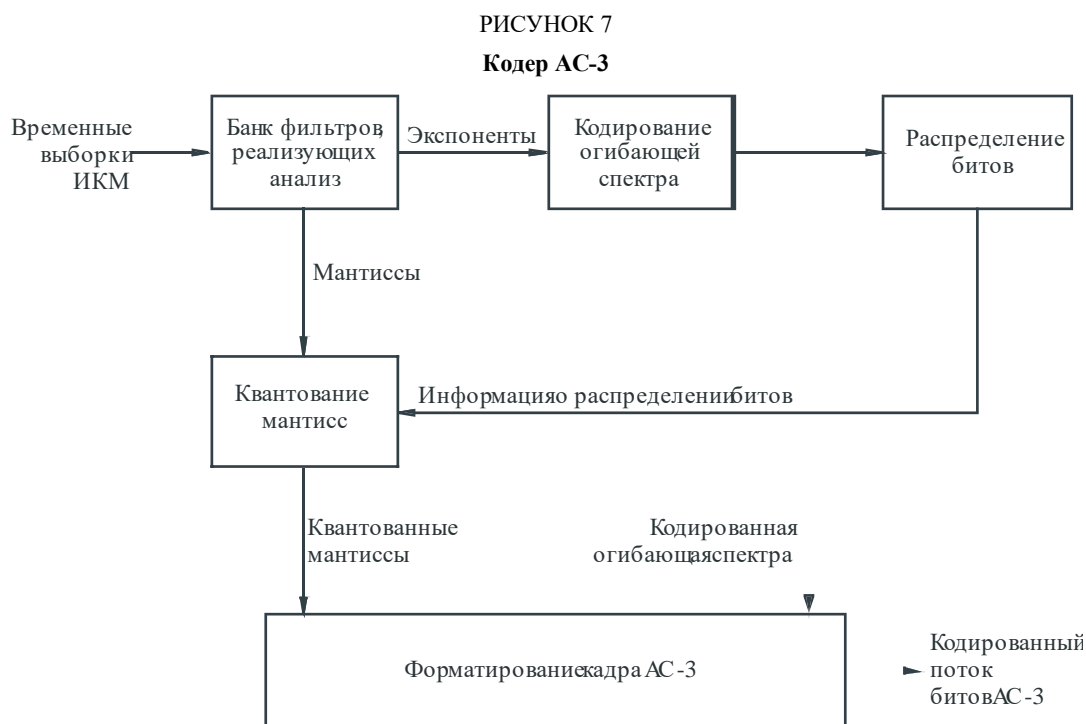
### Приложение 3 (информационное)

#### Звуковые AC-3 и E-AC-3

##### 1 Кодирование

Цифровой алгоритм сжатия AC-3 может кодировать от 1 до 5,1 каналов звукового источника из представления ИКМ в последовательный поток битов на скоростях передачи данных от 32 кбит/с до 640 кбит/с. Алгоритм AC-3 достигает высокой эффективности кодирования (отношение скорости передачи данных на входе к скорости передачи данных на выходе) путем грубого квантования представления звукового сигнала в частотной области. Структурная схема этого процесса показана на рисунке 7. Первым шагом в процессе кодирования является преобразование звука из последовательности временных выборок ИКМ в последовательность блоков частотных коэффициентов. Это осуществляется в банке фильтров, реализующих анализ. Перекрывающиеся блоки 512 временных выборок умножаются на временное окно и преобразуются в частотную область. Из-за перекрывающихся блоков каждая входная выборка ИКМ представляется двумя последовательными преобразованными блоками. Представление в частотной области может быть в таком случае уменьшено вдвое так, чтобы в каждом блоке содержалось 256 частотных коэффициентов. Отдельные частотные коэффициенты представлены в двоичной экспоненциальной нотации как двоичная экспонента и мантисса. Набор экспонент кодируется в грубое представление спектра сигнала, который называется спектральной огибающей. Эта спектральная огибающая используется внутренней стандартной программой распределения битов, которая определяет, сколько битов используется для кодирования каждой отдельной мантиссы. Спектральную огибающую и грубо квантованные мантиссы

для 6 звуковых блоков (1536 звуковых выборок) форматируют в кадр АС-3. Поток битов АС-3 является последовательностью кадров АС-3.



BS.1196-8-07

Реальный кодер АС-3 гораздо сложнее, чем это показано на рисунке 7. Включаются также следующие функции, не изображенные выше:

- присоединяется заголовок кадра, содержащий информацию (скорость передачи данных, частота дискретизации, число кодированных каналов и т. д.), необходимую для синхронизации и декодирования кодированного потока данных;
- вводятся коды обнаружения ошибок для обеспечения возможности проверки декодером того, что принятый кадр данных не содержит ошибок;
- можно динамически изменять спектральное разрешение банка фильтров, реализующих анализ, с тем чтобы лучше соответствовать временно-частотной характеристике каждого звукового блока;
- можно кодировать спектральную огибающую с изменяемым временно-частотным разрешением;
- может выполняться более сложное распределение битов и изменение параметров стандартной внутренней программы распределения битов, с тем чтобы создать более оптимальное распределение битов;
- каналы могут быть объединены на высоких частотах в целях достижения большей эффективности кодирования для работы на более низких скоростях передачи данных;
- при двухканальном режиме может с успехом выполняться процесс повторного кодирования с помощью матричной схемы для обеспечения дополнительной эффективности кодирования и предоставления возможности получения улучшенных результатов в случае декодирования двухканального сигнала матричным декодером объемного звучания.

## 2 Декодирование

В основном процесс декодирования является процессом, обратным кодированию. Декодер, показанный на рисунке 8, должен синхронизировать кодированный поток битов, проверять ошибки и деформатировать различные типы данных, как, например, кодированную спектральную огибающую и квантованные мантиссы. Результаты работы стандартной программы распределения битов используются для распаковки и деквантования мантисс. Спектральная огибающая декодируется для создания экспонент. Экспоненты и мантиссы обратно преобразуются во временную область для создания декодированных временных выборок ИКМ.

РИСУНОК 8  
Декодер АС-3



BS.1196-8-08

Реальный декодер АС-3 сложнее, чем это показано на рисунке 8. Включаются следующие функции, не изображенные выше:

- в случае обнаружения ошибок в данных может применяться скрывание ошибок или выключение;
- каналы, высокочастотный контент которых был объединен, должны быть разъединены;
- должно применяться декодирование с помощью матричной схемы (в двухканальном режиме) каждый раз, когда каналы были повторно кодированы с помощью матричной схемы;
- разрешение банка фильтров, реализующих синтез, должно динамически меняться так же, как это имело место в ходе процесса кодирования в банке фильтров кодера, реализующих анализ.

## 3 Е-АС-3

Расширенное АС-3 (Е-АС-3) добавляет несколько дополнительных инструментов кодирования к вышеописанному базовому кодеку АС-3. Дополнительные инструменты кодирования обеспечивают повышенную эффективность кодирования, которая дает возможность работы на более низких скоростях передачи, тогда как дополнительные характеристики обеспечивают дополнительную гибкость применения.

Дополнительные инструменты кодирования:

- адаптивный гибридный преобразователь – дополнительный уровень, применяемый в банке фильтров анализа/синтеза для обеспечения более тонкого спектрального разрешения (1/6 от AC-3);
- переходная предшумовая обработка – дополнительный инструмент для сокращения переходных предшумов;
- спектральное расширение – синтез декодером компонентов наиболее высоких частот на основании побочной информации, создаваемой кодером;
- расширенное объединение – при объединении каналов охватывается фаза наряду с амплитудой.

Дополнительные характеристики:

- более точная детализация скорости передачи данных;
- более высокая максимальная скорость передачи данных (3 Мбит/с);
- подпотоки могут нести дополнительные звуковые каналы, например 7.1 chs или комментаторские дорожки.

## Приложение 4 (информационное)

### MPEG Surround

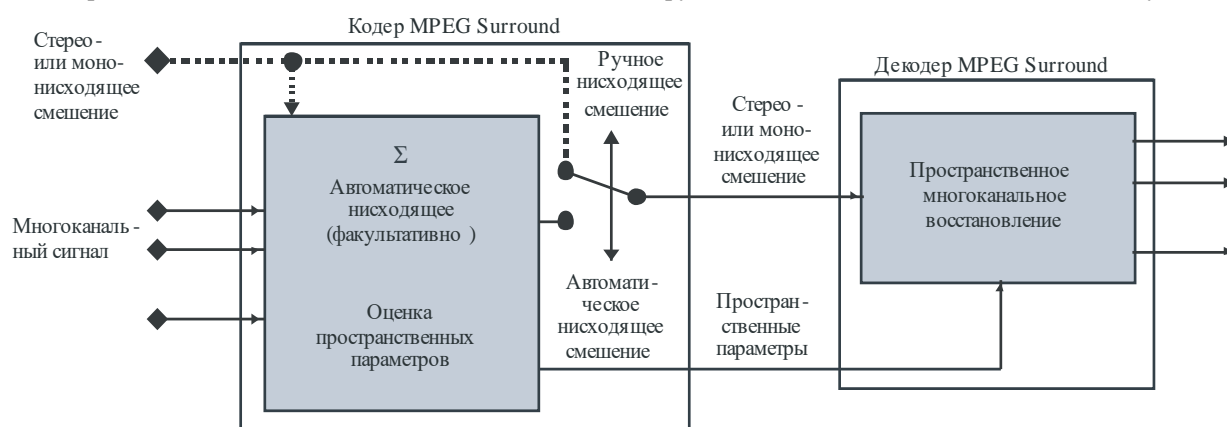
#### 1 Введение

Технология ИСО/МЭК 23003-1 или MPEG Surround обеспечивает чрезвычайно эффективный метод кодирования многоканального звука и позволяет передавать окружающий звук на скоростях, которые обычно используются для кодирования моно- или стереозвуча. С ее помощью можно представить многоканальный звуковой сигнал  $N$  каналов на основании нисходящего смешения каналов  $M < N$  и дополнительных контрольных данных. В предпочитаемых режимах эксплуатации кодер MPEG Surround производит моно- или стереонисходящее смешение из многоканального входного звукового сигнала. Это нисходящее смешение кодируется с использованием стандартного базового кодека звука, например одной из систем кодирования, рекомендуемых в пунктах 1 и 2 раздела *рекомендует*. В дополнение к нисходящему смешению MPEG Surround генерирует описание пространственного изображения параметров многоканального звука, которое добавляется как поток вспомогательных данных к базовому кодеку звука способом, совместимым в обратном направлении. Традиционные моно- или стереодекодеры игнорируют вспомогательные данные и воспроизводят звуковой стерео- или моносигнал нисходящего смешения. Оснащенные MPEG декодеры сначала декодируют моно- или стереонисходящее смешение, а затем используют параметры пространственного изображения, извлеченные из потока вспомогательных данных, для генерирования высококачественного многоканального звукового сигнала.

На рисунке 9 показан принцип MPEG Surround.

РИСУНОК 9

**Принцип MPEG Surround, нисходящее смешение кодируется с использованием базового кодека звука**



BS.1196-8-09

Используя MPEG Surround, существующие услуги легко усовершенствовать для обеспечения окружающего звука способом, совместимым в обратном направлении. Стереодекoder в существующем традиционном пользовательском устройстве игнорирует данные MPEG Surround data и воспроизводит стереосигнал без ухудшения качества, тогда как снабженный MPEG Surround декодер поставляет многоканальный звук высокого качества.

## 2 Кодирование

Цель кодера MPEG Surround заключается в представлении многоканального входного сигнала как совместимого в обратном направлении моно- или стереосигнала в сочетании с пространственными параметрами, которые позволяют восстанавливать многоканальный выходной сигнал, схожий в отношении восприятия с первоначальными многоканальными входными сигналами. Наряду с автоматически генерируемым нисходящим смешением может использоваться созданное ввне нисходящее смешение ("художественное смешение"). Нисходящее смешение сохранит пространственные характеристики входного звука.

MPEG Surround строится на технологии параметрического стерео в сочетании с HE AAC, в результате чего возникает стандартная спецификация HE AAC v2. Путем сочетания множественных модулей параметрического стерео и других разработанных в последнее время модулей были определены различные структуры, поддерживающие разные сочетания ряда выходных каналов и каналов нисходящего смешения. В качестве примера: для многоканального входного сигнала а 5.1 имеются три различные конфигурации – одна конфигурация для базирующихся на нисходящем смешении стереосистем (конфигурация 525) и две различные конфигурации для базирующихся на нисходящем смешении моносистем (конфигурация 515<sub>1</sub> и 515<sub>2</sub>, использующая другое сочетание узлов).

В MPEG Surround входят ряд инструментов, позволяющих применять характеристики, которые дают возможность использовать этот стандарт в широких масштабах. Одной из ключевых характеристик MPEG Surround является способность постепенно масштабировать качество пространственного изображения от очень низкого пространственного заголовка до прозрачности. Еще одной из ключевых характеристик является то, что вход декодера можно сделать совместимым с существующими матричными технологиями окружающего звука.

Эти и другие характеристики реализуются следующими основными инструментами кодирования.

- Остаточное кодирование – наряду с пространственными параметрами также могут передаваться остаточные сигналы при использовании метода гибридного кодирования. Эти сигналы заменяют часть некоррелированных сигналов (являющихся частью узлов параметрического стерео). Остаточные сигналы кодируются путем преобразования сигналов области КЗФ в область MDCT, после чего коэффициенты MDCT кодируются с использованием AAC.

- Матричная совместимость – факультативно стереонисходящее смешение может заранее обрабатываться, чтобы быть совместимым с традиционными матричными технологиями окружающего звука для обеспечения совместимости в обратном направлении с декодерами, которые способны декодировать только поток стереобитов, но снабжены матричным декодером окружающего звука.
- Произвольные сигналы нисходящего смешения – система MPEG Surround способна обрабатывать не только генерированное кодером нисходящее смешение, но и художественное смешение, поставляемое кодеру в дополнение к первоначальному многоканальному сигналу.
- MPEG Surround по ИКМ – обычно пространственные параметры MPEG Surround несутся в части вспомогательных данных базовой схемы сжатия звука. Для применений, в которых нисходящее смешение передается как ИКМ, MPEG Surround также поддерживает метод, позволяющий переносить пространственные параметры по несжатым аудиоканалам. В основе этого лежит технология, известная как скрытые данные.

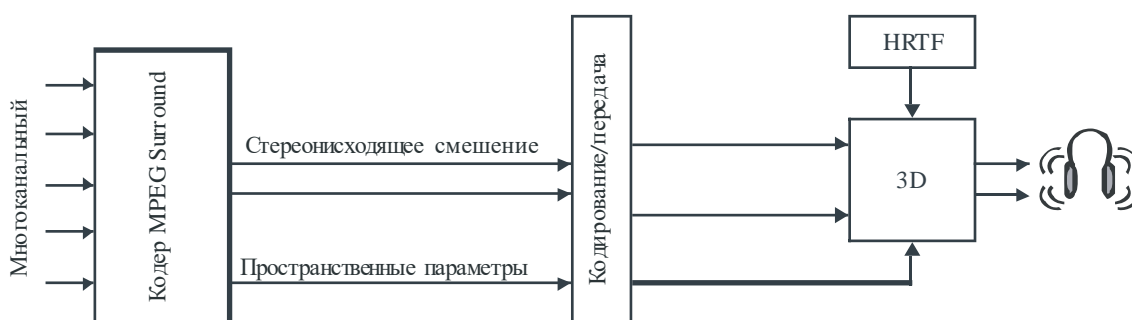
### 3 Декодирование

Наряду с обработкой многоканального выходного сигнала декодер MPEG Surround также поддерживает обработку альтернативных конфигураций выходного сигнала.

- Виртуальный окружающий звук – система MPEG Surround может эксплуатировать пространственные параметры для сведения нисходящего смешения до выходного стереосигнала виртуального окружающего звука для воспроизведения в традиционных наушниках. В стандарте указывается не функция передачи, учитывающая строение головы (HRTF), но только интерфейс к этим HRTF, что предоставляет свободу при реализации в зависимости от случая использования. Обработка виртуального окружающего звука может применяться как в декодере, так и в кодере, причем последний обеспечивает возможность ощущения виртуального окружающего звука при нисходящем смешении, без необходимости применения декодера MPEG Surround. Вместе с тем декодер MPEG Surround может отменить обработку виртуального окружающего звука при нисходящем смешении и повторно применить альтернативный виртуальный окружающий звук. Основной принцип отображен на рисунке 10.

РИСУНОК 10

#### Декодирование виртуального окружающего звука MPEG Surround



BS.1196-8-10

- Расширенный матричный режим – в случае традиционного стереоконтента, когда побочная пространственная информация отсутствует, MPEG Surround способен оценить побочную пространственную информацию из нисходящего смешения и тем самым создать многоканальный выходной звук при уровне качества выше, чем то, которое обеспечивают обычные матричные системы окружающего звука.
- Отсечение – вследствие лежащей в его основе структуры декодер MPEG Surround может поставить свой выходной сигнал для конфигураций каналов, где число каналов меньше, чем число каналов в многоканальном входном сигнале кодера.

#### 4 Профили и уровни

Декодер MPEG Surround может применяться в виде версии высокого качества и версии низкой мощности. Обе версии работают на одном и том же потоке данных, но с разными выходными сигналами.

Базовый профиль MPEG Surround определяет шесть различных иерархических уровней, которыми предусматривается различное число входных и выходных каналов, различные диапазоны частот дискретизации и различная ширина полосы декодирования остаточного сигнала. Уровень декодера должен быть равен уровню битового потока или выше его, чтобы обеспечить надлежащее декодирование. Наряду с этим декодеры уровней 1, 2 и 3 способны декодировать все битовые потоки уровней 2, 3 и 4, хотя, возможно, при несколько более низком качестве ввиду ограничений декодера. Вместе с тем качество и формат выходного сигнала декодера MPEG Surround зависит от конкретной конфигурации декодера. Тем не менее аспекты конфигурации декодера полностью ортогональны по отношению к различным уровням этого профиля.

#### 5 Присоединение к звуковым кодекам

MPEG Surround работает как расширение на периоды до и после обработки поверх традиционных схем кодирования звука. Ввиду этого он способен обеспечить возможность работы практически любого кодера звука. Формирование кадров в MPEG Surround отличается высокой гибкостью, что позволяет обеспечить синхронность с широким диапазоном кодеров, а также имеются средства для оптимизации соединения с кодерами, которые уже используют параметрические инструменты (например, дублирование спектральной полосы).

## Приложение 5 (информационное)

### Расширенное высокоэффективное AAC (расширенное HE AAC)

#### 1 Введение

Профиль расширенного HE AAC, определенный в рамках унифицированного кодирования речи и звука (USAC) MPEG-D ИСО/МЭК 23003-3. USAC является стандартом кодирования звуковых сигналов, которое позволяет кодировать речевые, звуковые сигналы, а также любое сочетание речевых и звуковых сигналов при соответствующем качестве звучания для всех звуковых материалов в широком диапазоне скоростей передачи. Он поддерживает одноканальное и многоканальное кодирование на больших скоростях передачи, где обеспечивает практически прозрачное качество. В то же время он обеспечивает очень эффективное кодирование на очень низких скоростях передачи, полностью сохраняя при этом ширину полосы звукового сигнала.

Если у предыдущих кодеков звуковых сигналов были специфические сильные и слабые стороны при кодировании или речевого или звукового контента, то USAC способен кодировать любой контент с одинаково высокой точностью, независимо от типа такого контента.

Для того чтобы достичь одинаково высокого качества кодирования звуковых и речевых сигналов, в USAC применяются хорошо зарекомендовавшие себя методы кодирования, основанные на модифицированном дискретном косинусном преобразовании (MDCT), известные по MPEG-4 для звука (MPEG-4 AAC, HE AAC, HE AAC v2), объединяя их с такими специализированными элементами речевого кодера, как линейное предсказание с возбуждением по алгебраической кодовой книге (ACELP). Средства параметрического кодирования, такие как дублирование спектральной полосы (SBR) MPEG-4 и MPEG Surround в MPEG-D, улучшены и хорошо интегрированы в этот кодек.

В результате получилось высокоэффективное кодирование, работающее вплоть до самых низких скоростей передачи.

В настоящее время в стандарте USAC определяются два профиля.

– *Базовый профиль USAC*

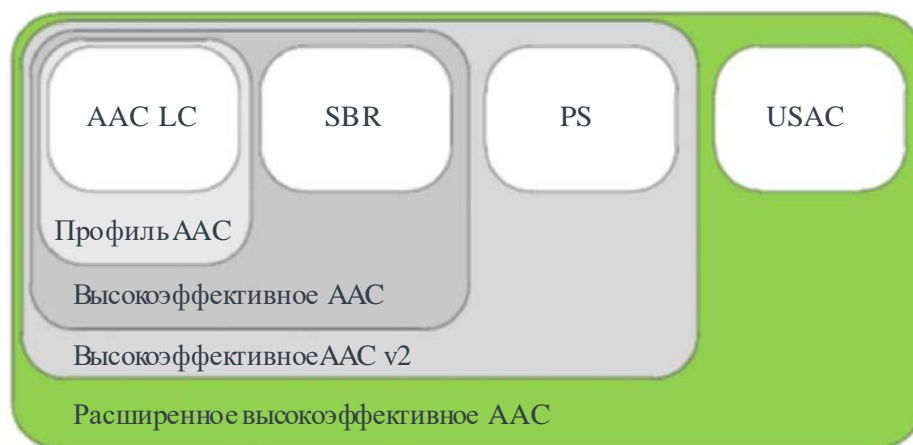
В базовом профиле USAC представлены функции стандарта USAC в полном объеме при сохранении общей вычислительной сложности на низком уровне. Исключены те средства, для которых требуется значительный объем памяти или вычислительной мощности.

– *Расширенный профиль HE AAC*

Этим профилем, путем добавления возможностей USAC, расширяется существующий профиль HE AAC v2, чтобы специально ориентироваться на применения, в которых должна сохраниться совместимость с существующим семейством профилей AAC (AAC, HE AAC и HE AAC v2). В этот профиль включен уровень 2 базового профиля USAC. Вследствие этого декодеры с расширенным профилем HE AAC могут декодировать битовые потоки HE AAC v2, а также битовые потоки USAC (до двух каналов).

РИСУНОК 11

Структура расширенного высокоэффективного AAC



BS.11968-11

В USAC поддерживаются частоты дискретизации от 7,35 кГц до 96 кГц, при этом он обеспечивает высокое качество звука для битовых скоростей в диапазоне начиная с 8 кбит/с и до битовых скоростей, на которых достигается воспринимаемая прозрачность. Это было подтверждено результатами верификационного испытания (документ MPEG2011/N12232), полученными из PГ11/ПК29/ОТК1 ИСО/МЭК, которые приложены к Документу 6В/286 (Rev.2).

Конфигурация каналов могла выбираться произвольно. 13 различных конфигураций каналов, устанавливаемых по умолчанию, могут обеспечивать эффективную передачу сигналов в большинстве обычных сценариев работы применений. В эти конфигурации, устанавливаемые по умолчанию, включены все конфигурации каналов MPEG-4, такие как моно, стерео, 5.0 и 5.1 объемный, или даже схемы громкоговорителей 7.1 или 22.2.

## 2 Кодирование

Как это обычно принято при стандартизации MPEG, стандарт ИСО/МЭК 23003-3 определяет только процесс декодирования файлов и потоков данных USAC MPEG-D. Он нормативно не определяет процесс кодирования.

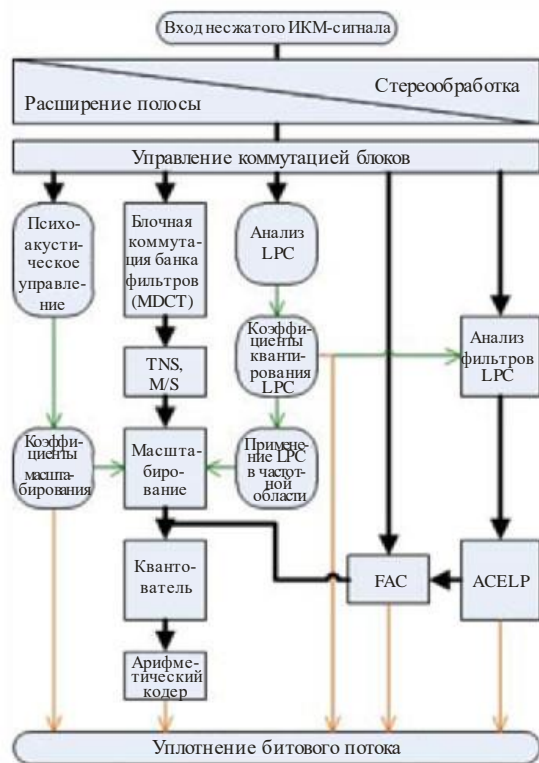
Типовая возможная структура кодера показана на рисунке 12.

Кодер состоит из следующих средств кодирования.



- Обработка стереосигнала. На низких/промежуточных битовых скоростях в USAC применяются технологии параметрического стереокодирования. Принцип их работы подобен работе средства параметрического стерео, описанного в Дополнении 2.5 к ИСО/МЭК 23003-3, но отличается тем, что базируется на MPEG Surround, как описано в Приложении 4, и поэтому называется MPEG Surround 2-1-2 (MPS 2-1-2). Кодер извлекает из входного звукового сигнала высокоэффективное параметрическое представление стереофонического образа. Эти параметры передаются в битовом потоке совместно с монозвуковым сигналом, полученным в результате нисходящего смещения. По выбору кодер может передавать остаточный сигнал, который подкорректирует процесс реконструкции стереосигнала в декодере. Механизм кодирования остаточного сигнала позволяет плавно переходить от полностью параметрического до полностью дискретного стереокодирования канала. Средство MPS 2-1-2 является неотъемлемой частью кодека USAC. При более высоких битовых скоростях, когда параметрическое кодирование и ACELP обычно не действуют, стереокодирование может осуществляться исключительно в области MDCT посредством комплексного стереопредсказания. Поэтому данный метод называют стереокодированием с комплексным предсказанием. Его можно рассматривать как обобщение традиционного С/Б стереокодирования.
- Расширение полосы. Параметрическое расширение полосы – это многократно улучшенная версия дублирования спектральной полосы (SBR) MPEG-4, которая описана в Дополнении 2.4 к ИСО/МЭК 23003-3. Кодер оценивает огибающую спектра, тональность более высоких полос звуковых частот и передает соответствующие параметры на декодер. В кодере может выбираться один из двух различных типов средств преобразования (гармонического или с копированием) и один из трех коэффициентов преобразования (1:2, 3:8, 1:4). Это улучшенное средство SBR является неотъемлемой частью кодека USAC.
- Банк фильтров, коммутация блоков. Банк фильтров на основе MDCT образует основу ядра кодера. В зависимости от примененного механизма формирования шумов квантования разрешение преобразования может быть выбрано 1024, 512, 256 или 128 спектральных линий. В комбинации с коэффициентом преобразования SBR 3:8 разрешение может быть изменено до  $\frac{3}{4}$  от перечисленных выше вариантов, обеспечивая лучшую временную дробность даже при более низких частотах дискретизации.
- Временное ограничение шума (TNS), С/Б стереокодирование, квантование. Эти средства были взяты из AAC и применены так, как описано в Дополнении 2.2 к ИСО/МЭК 23003-3.
- Арифметический кодер с адаптацией к контексту. Кодирование спектральных коэффициентов MDCT без образования шума (то есть энтропийное) осуществляется арифметическим кодером, который выбирает таблицы вероятностей, основываясь на ранее кодированных спектральных линиях.
- Психоакустическое управление, изменение коэффициента масштабирования – коэффициент масштабирования основывается на психоакустической модели, подобной той, которая используется в AAC, см. Дополнение 2.2 к ИСО/МЭК 23003-3.
- Масштабирование, основанное на параметрах кодирования с линейным предсказанием (LPC) – это средство формирования спектрального шума может использоваться в качестве альтернативы упомянутому выше изменению коэффициентов масштабирования. Взвешенная версия частотного представления набора коэффициентов фильтра LPC применяется к спектральным коэффициентам MDCT до квантования и кодирования.
- ACELP. В кодере с линейным предсказанием, возбуждаемым алгебраическим кодом (ACELP), применяется хорошо зарекомендовавшее себя представление возбуждения адаптивной/инновационной кодовой таблицей, известное по современным речевым кодекам.
- Уплотнение битового потока. Окончательный битовый поток формируется из различных элементов, которые производятся средствами кодера.
- FAS. Средство прямой коррекции ступенчатости (FAS) является механизмом, обеспечивающим плавный переход от кодирования на основе MDCT, подверженного ступенчатости, к кодированию ACELP во временной области.

РИСУНОК 12  
Блок-схема кодера USAC MPEG-D



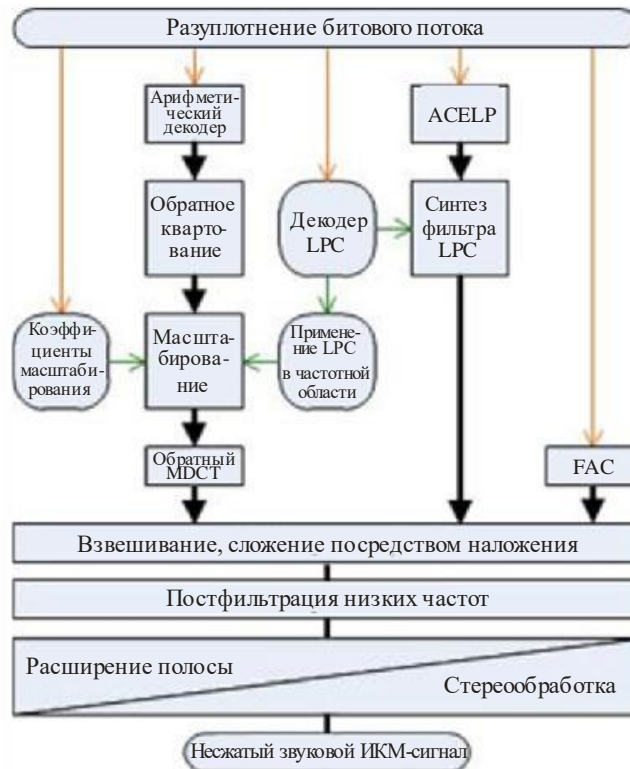
BS.1196-8-12

### 3 Декодирование

Базовая структура декодера USAC MPEG-D показана на рисунке 13. Процесс декодирования обычно осуществляется в последовательности, обратной процессу кодирования.

РИСУНОК 13

## Блок-схема декодера USAC MPEG-D



BS.11968-B

Процесс декодирования может быть укрупненно описан следующим образом.

- Разуплотнение битового потока – декодером осуществляется поиск в битовом потоке всей необходимой ему информации и направление ее на соответствующие модули декодера.
- Декодирование в ядре – в зависимости от содержания битового потока, декодер:
  - декодирует и осуществляет процедуру, обратную квантованию спектральных коэффициентов MDCT, применяет масштабирование, основанное или на информации о коэффициентах масштабирования, или на информации о коэффициентах LPC, и применяет далее (опционально) средства, основанные на MDCT, если они существуют и применимы. Наконец, применяется обратный процесс MDCT для получения соответствующего сигнала во временной области;
  - или декодирует информацию, касающуюся ACELP, производит сигнал возбуждения и синтезирует выходной сигнал с помощью фильтра LPC.
- Взвешивание, сложение посредством наложения – последовательность кадров ядра кодера соединяют или сливают в ходе обычного процесса сложения посредством наложения, известного из AAC. Переходы между кодированием, основанном на ACELP или MDCT, осуществляются путем слияния данных, декодированных FAC.
- Постфильтрация низких частот – для улучшения качества речевого сигнала может применяться опциональный фильтр, улучшающий высоту звука.
- Расширение полосы, стереопереработка – наконец, применяются средства параметрического кодирования для расширения полосы и средства стереофонического кодирования в целях воссоздания полной полосы, дискретного стереосигнала.

Для каждого из опциональных средств сохраняется возможность "сквозного прохода", и во всех случаях, когда некоторая операция исключается, данные с ее входа проходят через средство без изменения.

#### 4 Профили и уровни

В настоящее время в MPEG определяются два профиля, в которых применяется кодек USAC.

– *Базовый профиль USAC*

В базовом профиле USAC содержится полный кодек USAC, за исключением нескольких средств, которые характеризуются значительной вычислительной сложностью в наихудшем случае. Эти средства выше не были описаны. Данный профиль представляет собой четкий автономный профиль для применений и случаев использования, при которых способность поддержки семейства профилей AAC (профиль AAC, профиль HE AAC, профиль HE AAC v2) не является важной.

– *Расширенный профиль HE AAC*

В профиле расширенного высокоэффективного AAC содержатся все средства высокоэффективного профиля AAC v2, и в таком виде он способен декодировать все потоки профиля семейства AAC. Кроме того, в этот профиль включена возможность декодирования моно-/стереосигнала базового профиля USAC. Следовательно, этот профиль является естественным развитием профиля HE AAC v2, так как моно-/стереоподсистема USAC (при работе на низких скоростях) приобретает дополнительную ценность совместимых характеристик для различных типов контента при низких скоростях.

## Приложение 6 (информационное)

### Не зависящие от метода кодирования кодовые точки (CICP) для кодирования MPEG

#### 1 Введение

В ISO/IEC23001-8:2013 описаны аспекты кодирования звуковых программ, не зависящие от кодированного представления, включая позиции и схему расположения систем громкоговорителей. Конфигурации каналов по умолчанию включают конфигурации каналов, описанные в Рекомендациях МСЭ-R BS.775 или МСЭ-R BS.2051. Все конфигурации каналов показаны в таблице 3.

ТАБЛИЦА 3

#### Конфигурации каналов и схемы расположения громкоговорителей (Примечание 1)

Значение <sup>*1</sup> конфигурации каналов (Примечание 1)	Число громкогов- ворителей (Примечание 2)	Отображение по умолчанию элемента на громкоговоритель (Примечание 3)	Название канала, определенное в Рекомендации МСЭ-R BS.775 или BS.2051 (Примечание 4)
0	Любая установка		
1	1/0.0 (0 + 1 + 0)	M + 000	Моно
2	2/0.0 (0 + 2 + 0)	M + 030	Левый
		M – 030	Правый
3	3/0.0 (0 + 3 + 0)	M + 000	Центральный
		M + 030	Левый
		M – 030	Правый

ТАБЛИЦА 3 (продолжение)

Значение* <sup>1</sup> конфигурации каналов (Примечание 1)	Число громкого- ворителей (Примечание 2)	Отображение по умолчанию элемента на громкоговоритель (Примечание 3)	Название канала, определенное в Рекомендации МСЭ-R BS.775 или BS.2051 (Примечание 4)
4	3/1.0 (0 + 4 + 0)	M + 000	Центральный
		M + 030	Левый
		M – 030	Правый
		M + 180	Моно с эффектом объемного звука
5	3/2.0 (0 + 5 + 0)	M + 000	Центральный
		M + 030	Левый
		M – 030	Правый
		M + 110	Левый с объемным звуком
		M – 110	Правый с объемным звуком
6	3/2.1 (0 + 5 + 0)	M + 000	Центральный
		M + 030	Левый
		M – 030	Правый
		M + 110	Левый с объемным звуком
		M – 110	Правый с объемным звуком
		LFE1	Низкочастотных эффектов
7	5/2.1 (0 + 7 + 0)	M + 000	Н/П* <sup>2</sup>
		M + 030	
		M – 030	
		M + 045	
		M – 045	
		M + 110	
		M – 110	
		LFE1	
8	1 + 1	Канал 1	Н/П
		Канал 2	
9	2/1.0 (0 + 3 + 0)	M + 030	Левый
		M – 030	Правый
		M + 180	Моно с эффектом объемного звука
10	2/2.0 (0 + 4 + 0)	M + 030	Левый
		M – 030	Правый
		M + 110	Левый с объемным звуком
		M – 110	Правый с объемным звуком
11	3/3.1 (0 + 6 + 0)	M + 000	Н/П
		M + 030	
		M – 030	
		M + 110	
		M – 110	
		M + 180	
		LFE1	

ТАБЛИЦА 3 (продолжение)

Значение* <sup>1</sup> конфигурации каналов (Примечание 1)	Число громкоговори- телей (Примечание 2)	Отображение по умолчанию элемента на громкоговоритель (Примечание 3)	Название канала, определенное в Рекомендации МСЭ-R BS.775 или BS.2051 (Примечание 4)
12	3/4.1 (0 + 7 + 0)	M + 000	Центральный
		M + 030	Левый
		M – 030	Правый
		M + 090	Левый боковой с объемным звуком
		M – 090	Правый боковой с объемным звуком
		M + 135	Левый задний с объемным звуком
		M – 135	Правый задний с объемным звуком
13	11/11.2 (9 + 10 + 3)	LFE1	Низкочастотных эффектов
		M + 000	Фронтальный центральный
		M + 030	Фронтальный левый центральный
		M – 030	Фронтальный правый центральный
		M + 060	Фронтальный левый
		M – 060	Фронтальный правый
		M + 090	Левый боковой
		M – 090	Правый боковой
		M + 135	Тыловой левый
		M – 135	Тыловой правый
		M + 180	Тыловой центральный
		LFE1	Низкочастотных эффектов-1
		LFE2	Низкочастотных эффектов-2
		U + 000	Верхний фронтальный центральный
		U + 045	Верхний фронтальный левый
		U – 045	Верхний фронтальный правый
		U + 090	Верхний левый боковой
		U – 090	Верхний правый боковой
		T + 000	Верхний центральный
		U + 135	Верхний тыловой левый
		U – 135	Верхний тыловой правый
U + 180	Верхний тыловой центральный		
V + 000	Нижний фронтальный центральный		
V + 045	Нижний фронтальный левый		
U – 045	Нижний фронтальный правый		
14	5/2.1 (2 + 5 + 0)	M + 000	Центральный
		M + 030	Левый
		M – 030	Правый
		M + 110	Левый с объемным звуком
		M – 110	Правый с объемным звуком
		LFE1	Низкочастотных эффектов
		U + 030	Левый верхний фронтальный
U – 030	Правый верхний фронтальный		

ТАБЛИЦА 3 (продолжение)

Значение* <sup>1</sup> конфигурации каналов (Примечание 1)	Число громкого- ворителей (Примечание 2)	Отображение по умолчанию элемента на громкоговоритель (Примечание 3)	Название канала, определенное в Рекомендации МСЭ-R BS.775 или BS.2051 (Примечание 4)
15	5/5.2 (3 + 7 + 0)	M + 000	Центральный
		M + 030	Левый
		M – 030	Правый
		M + 090	Левый боковой
		M – 090	Правый боковой
		M + 135	Тыловой левый
		M – 135	Тыловой правый
		U + 045	Левый верхний
		U – 045	Правый верхний
		UH + 180	Центральный верхний
		LFE1	Левый низкочастотных эффектов
		LFE2	Правый низкочастотных эффектов
16	5/4.1 (4 + 5 + 0)	M + 000	Центральный
		M + 030	Левый
		M – 030	Правый
		M + 110	Левый с объемным звуком
		M – 110	Правый с объемным звуком
		LFE1	Низкочастотных эффектов
		U + 030	Левый верхний фронтальный
		U – 030	Правый верхний фронтальный
		U + 110	Левый верхний задний
		U – 110	Правый верхний задний
17	6/5.1 (6 + 5 + 0)	M + 000	Н/П
		M + 030	
		M – 030	
		M + 110	
		M – 110	
		LFE1	
		U + 000	
		U + 030	
		U – 030	
		U + 110	
		U – 110	
		T + 000	
18	6/7.1 (6 + 7 + 0)	M + 000	Н/П
		M + 030	
		M – 030	
		M + 110	
		M – 110	
		M + 150	
		M – 150	
		LFE1	
		U + 000	
		U + 030	
		U – 030	
		U + 110	
		U – 110	
		T + 000	

ТАБЛИЦА 3 (окончание)

Значение <sup>*1</sup> конфигурации каналов (Примечание 1)	Число громкоговори- телей (Примечание 2)	Отображение по умолчанию элемента на громкоговоритель (Примечание 3)	Название канала, определенное в Рекомендации МСЭ-R BS.775 или BS.2051 (Примечание 4)
19	5/6.1 (4 + 7 + 0)	M + 000	Центральный
		M + 030	Левый
		M – 030	Правый
		M + 090	Левый боковой с объемным звуком
		M – 090	Правый боковой с объемным звуком
		M + 135	Левый задний с объемным звуком
		M – 135	Правый задний с объемным звуком
		LFE	Низкочастотных эффектов
		U + 045	Левый верхний фронтальный
		U – 045	Правый верхний фронтальный
		U + 135	Левый верхний тыловой
		U – 135	Правый верхний тыловой
20	7/6.1 (4 + 9 + 0)	M + 000	Центральный
		M + SC	Левый экран
		M – SC	Правый экран
		M + 030	Левый
		M – 030	Правый
		M + 090	Левый боковой с объемным звуком
		M – 090	Правый боковой с объемным звуком
		M + 135	Левый задний с объемным звуком
		M – 135	Правый задний с объемным звуком
		LFE	Низкочастотных эффектов
		U + 045	Левый верхний фронтальный
		U – 045	Правый верхний фронтальный
U + 135	Левый верхний тыловой		
U – 135	Правый верхний тыловой		
21–63	Зарезервированы		

<sup>\*1</sup> Конфигурация звуковых выходных каналов указывается шестибитовым полем, в котором содержится значение конфигурации каналов, определенное в ISO/IEC 23001-8:2013 "Не зависящие от метода кодирования кодовые точки".

<sup>\*2</sup> Н/П – неприменимо; конфигурация каналов отсутствует в Рекомендации МСЭ-R BS.2051 или Рекомендации МСЭ-R BS.775.

ПРИМЕЧАНИЕ 1. – Список приведен из таблицы 8 ISO/IEC 23001-8:2013/Amd.1:2015.

ПРИМЕЧАНИЕ 2. – Представление числа громкоговорителей соответствует обозначениям "Фронтальные громкоговорители/громкоговорители с объемным звуком, громкоговорители НЧЭ" и в скобках – "Громкоговорители верхнего уровня + громкоговорители среднего уровня + громкоговорители нижнего уровня", громкоговорители НЧЭ исключены.

ПРИМЕЧАНИЕ 3. – Идентификация громкоговорителей с помощью обозначений, соответствующих Рекомендации МСЭ-R BS.2051.

ПРИМЕЧАНИЕ 4. – Обозначения и названия каналов зависят от фактической конфигурации каналов.



## Приложение 7 (информационное)

### АС-4

#### 1 Введение

АС-4 – это усовершенствованная система кодирования для цифрового радиовещания, в которой используется алгоритм цифрового сжатия и различные средства параметрического кодирования для повышения эффективности и расширения функциональных возможностей. АС-4 по умолчанию поддерживает входные и выходные форматы как на основе канала, так и на основе объекта.

В таблице 4 представлен перечень форматов каналов, по умолчанию поддерживаемых системой АС-4, в котором указаны конфигурации каналов, необходимые для передачи сигналов, определенных в Рекомендации МСЭ-R BS.1548. АС-4 также поддерживает форматы многоканального кодирования, которые можно использовать для поддержки передовых аудиосистем, описанных в Рекомендации МСЭ-R BS.2051.

ТАБЛИЦА 4

**Поддерживаемые по умолчанию форматы каналов**

Формат	Число каналов	Примечание
Моно (формат 1/0) <sup>(1)</sup>	1	
Сtereo (формат 2/0) <sup>1)</sup>	2	
3.0 (формат 3/0) <sup>1)</sup>	3	
5.0/5.1 (формат 3/2) <sup>1)</sup>	5/6	
7.0/7.1 (система I) <sup>2)</sup>	7/8	Сигнализация трех разных конфигураций громкоговорителей
7.0.4/7.1 + 4 (система J) <sup>2)</sup>	11/12	Сигнализация подмножеств с меньшим числом каналов
9.0.4/9.1.4	13/14	Сигнализация подмножеств с меньшим числом каналов
22.2 (система H) <sup>2)</sup>	24	

<sup>1)</sup> Описан в Рекомендации МСЭ-R BS.775.

<sup>2)</sup> Описана в Рекомендации МСЭ-R BS.2051.

Система АС-4 позволяет кодировать от 1 до 22.2 каналов исходного аудиосигнала с ИКМ в последовательный битовый поток со скоростью передачи данных от 24 кбит/с до 1536 кбит/с. Кроме канальных представлений система АС-4 поддерживает кодирование динамических звуковых объектов. Полное описание синтаксиса битового потока АС-4 содержится в документе ETSI TS 103 190-2.

В таблице 5 приведены предлагаемые скорости передачи битов при различных конфигурациях каналов для удовлетворения требований, изложенных в Рекомендации МСЭ-R BS.1548.

ТАБЛИЦА 5

**Скорость передачи битов для конкретных конфигураций каналов**

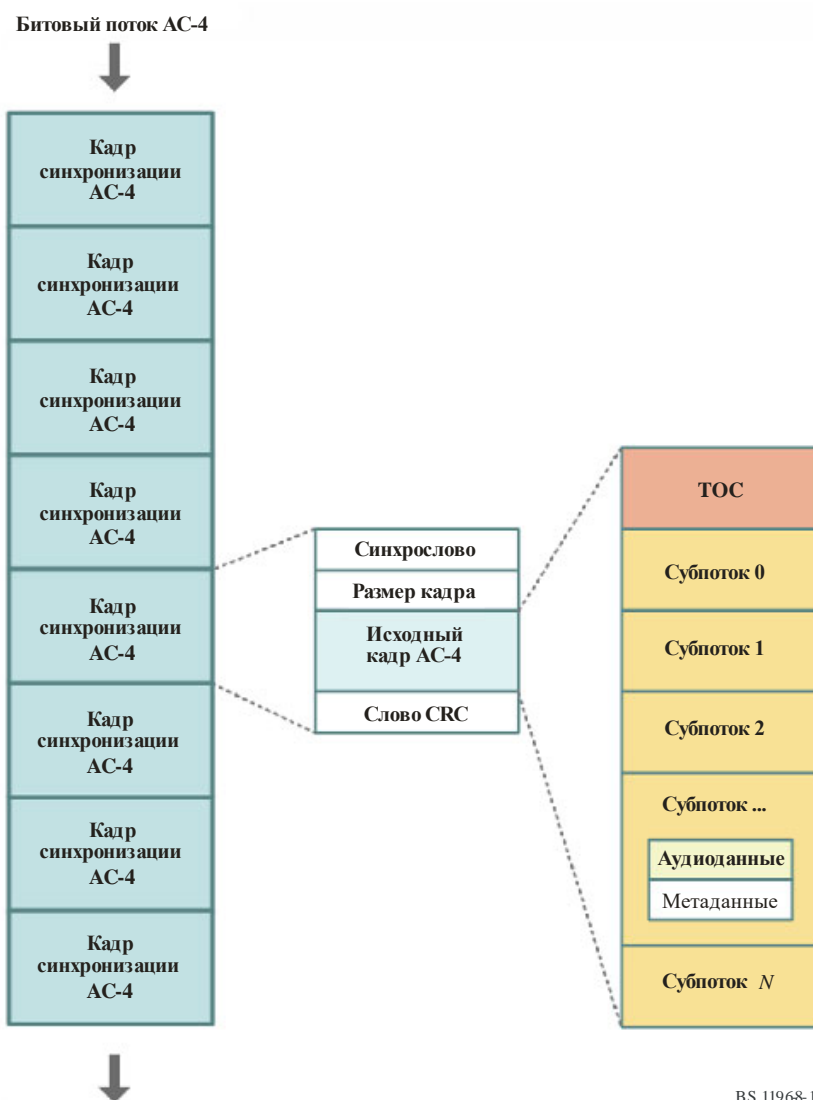
Режим	Скорость передачи битов
Сtereo 2.0	96 кбит/с
Объемный 5.1	192 кбит/с
Объемный 22.2	1536 кбит/с

АС-4 также по умолчанию поддерживает следующие системные функции:

- интеллектуальное управление громкостью в соответствии с Рекомендациями МСЭ-R BS.1770 и МСЭ-R BS.1771, включая сигнализацию для индикации соответствия международным и некоторым региональным стандартам громкости, действующим в настоящее время;
- поддержка кодирования и декодирования представлений звука на основе канала и на основе объекта;
- поддержка условий прослушивания согласно требованиям Рекомендации МСЭ-R BS.1909, в частности "домашних условий" и "мобильных условий", включая управление расширенным динамическим диапазоном, применимое к широкому спектру типов устройств как для домашних так и для мобильных условий;
- улучшенная передача разговора;
- синхронное кодирование видеок кадров, позволяющее выравнивать аудиокадры с видеок кадрами;
- встроенная поддержка передачи и сигнализации расширенных вспомогательных данных или метаданных.

Битовый поток АС-4, показанный на рисунке 14, состоит из кадров синхронизации АС-4, которые начинаются с синхрослова и заканчиваются словом циклической проверки избыточности (CRC). Синхрослово позволяет декодеру легко определить начало декодируемого кадра АС-4, а слово CRC – обнаруживать ошибки в битовом потоке и выполнять любое требуемое маскирование ошибок. Фактический кадр кодека, или исходный кадр АС-4, состоит из оглавления (TOC) и по меньшей мере одного субпотока.

РИСУНОК 14  
Общий синтаксис битового потока



Каждый субпоток включает кодированные аудиоданные, а также связанные с ними метаданные (вспомогательные данные). ТОС содержит необходимую информацию о том, как декодировать субпоток или несколько субпотоков.

## 2 Кодирование

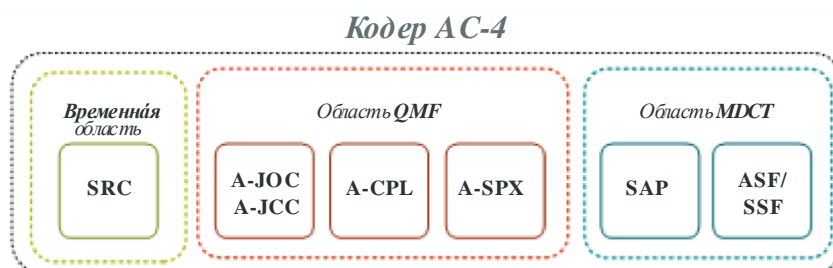
Кодер АС-4 нормативно не задан; поддерживаются различные средства кодирования, позволяющие создать совместимый кодированный битовый поток.

АС-4 использует квантование и кодирование области преобразования с использованием модифицированного дискретного косинусного преобразования (MDCT) с гибким переключением окон и параметрическим кодированием в банке псевдоквадратурных зеркальных фильтров. Система АС-4 поддерживает кодирование от 1 до 22.2 каналов исходного аудиосигнала с ИКМ в последовательный битовый поток со скоростью передачи данных от 24 кбит/с до 1536 кбит/с. При кодировании 2.0- и 5.1-канального стереозвуча требованиям Рекомендации МСЭ-R BS.1548 соответствуют скорости передачи данных 96 кбит/с и 192 кбит/с. АС-4 также поддерживает три режима распределения битов – режим с постоянной скоростью передачи битов, режим со средней битовой скоростью и режим с переменной битовой скоростью.

Кодер AC-4 может быть реализован с использованием разных средств кодирования, показанных на рисунке 15, для повышения эффективности и гибкости и расширения функциональных возможностей в зависимости от режима работы и/или назначения. Эти средства применяются в порядке слева направо, то есть аудиосигнал с ИКМ подается на вход устройств, показанных слева, а кодированный битовый поток AC-4 выводится справа. Ниже приведено описание этих средств кодирования.

- SRC (Sample Rate Converter) – преобразователь частоты дискретизации, обеспечивает в кодере AC-4 длительность кадра, соответствующую видеокадру, для режимов синхронного кодирования кадров. В зависимости от частоты кадров входной сигнал преобразуется в одну из внутренних частот дискретизации 46 034 Гц, 46 080 Гц, 48 000 Гц или 51 200 Гц, используемых последующими инструментами кодера на основе QMF и MDCT.
- A-JCC (Advanced Joint Channel Coding) – средство расширенного совместного кодирования каналов, выполняет понижающее микширование входного иммерсивного канала (более 5 каналов) путем сокращения числа каналов с кодированием соответствующих параметров. Эти параметры позволяют декодеру полностью восстановить входной канал.
- A-JOC (Advanced Joint Object Coding) – средство расширенного совместного кодирования объектов, принимает аудиообъекты в качестве входных объектов и пространственно кодирует их, создавая меньшее число выходных объектов, чтобы уменьшить количество сигналов в коде MDCT. Параметры, закодированные с меньшим числом выходных объектов, позволяют декодеру восстановить объекты.
- A-CPL (Advanced Coupling) – усовершенствованное средство соединения, выполняет понижающее микширование двухканального аудиосигнала в одноканальный и кодирует соответствующие параметры, которые позволяют восстановить исходный двухканальный входной сигнал.
- A-SPX (Advanced Spectral Extension) – средство расширенного развертывания спектра, кодирует параметры, связанные с высокочастотным содержанием входного сигнала, а затем развертывает спектр в декодере. В число параметров входят характеристики огибающей, тональности и шума. Спектральное и временное разрешение параметров можно адаптировать к характеристикам входного сигнала.
- SAP (Stereo Audio Processing) – средство обработки стереозвука, выполняет совместное кодирование каналов в области MDCT между двумя или более входными каналами.
- ASF (Audio Spectral Front-end) – средство входного спектрального анализа звука, метод квантования и кодирования на основе MDCT с использованием переключения окон преобразования. Модуль управления переключением окон выбирает оптимальную длину преобразования кадра в зависимости от типа входного сигнала. После нелинейного квантования и помехоустойчивого кодирования коэффициенты MDCT и дополнительная управляющая информация сохраняются в битовом потоке. Распределение битов по времени и частота регулируются битовым буфером и перцепционной моделью. Модель битового буфера также учитывает биты, используемые другими средствами кодирования и общими метаданными.
- SSF (Speech Spectral Front-end) – средство входного спектрального анализа речи, альтернативный метод квантования и кодирования на основе MDCT, специально предназначенный для кодирования речи и работающий с короткими преобразованиями. Выполняет квантование и кодирование в области MDCT с использованием предсказателя поддиапазона. Инструменты SSF и ASF являются взаимоисключающими, в связи с чем для кодирования MDCT используется либо ASF, либо SSF.

РИСУНОК 15  
Средства кодирования, имеющиеся в системе AC-4



BS.1196-8-15

### 3 Декодирование

Процесс декодирования – в основном обратный процессу кодирования – показан на рисунке 16.

РИСУНОК 16  
Средства декодирования



BS.1196-8-16

Битовый поток AC-4 и конструкция декодера поддерживают реализацию декодеров пониженной сложности для устройств с ограниченными возможностями (например, мобильных телефонов/планшетов). ↴

Для этого используются два метода.

- 1) Декодирование в ядре. AC-4 поддерживает два режима декодирования: полное декодирование и декодирование в ядре. Декодер AC-4 способен работать в режиме декодирования в ядре, когда декодируется основное подмножество каналов (представление содержимого всех входных каналов) кодированной программы, что обеспечивает согласованное воспроизведение программы при уменьшенной вычислительной сложности.
- 2) Масштабируемое декодирование. AC-4 также поддерживает масштабирование частоты дискретизации, так что повышенные частоты дискретизации, в частности 96 кГц и 192 кГц, поддерживаются масштабируемым образом. Устройства, поддерживающие только выходные сигналы с частотой 48 кГц, могут декодировать только базовый уровень.

В дополнение к декодированию поддерживаемых по умолчанию форматов каналов, указанных в таблице 4, AC-4 поддерживает использование вспомогательных данных или метаданных. Это позволяет предсказуемым образом производить понижающее микширование выходного сигнала декодированного канала для уменьшения числа выходных каналов в соответствии с параметрами устройства.

## Приложение 8 (информационное)

### Профиль пониженной сложности формата MPEG-H 3D Audio

#### 1 Введение

Формат MPEG-H 3D Audio – это стандарт кодирования звука, предназначенный для поддержки кодирования звука в форме звуковых каналов, звуковых объектов или по технологии Ambisonics высшего порядка (НОА) и обеспечивающий решения для нормирования громкости и управления динамическим диапазоном. Контент каждого типа (каналы, объекты или сигналы НОА) может использоваться отдельно или в сочетании с другими. Использование групп звуковых каналов, объектов или НОА позволяет обеспечить интерактивность или персонализацию программ, например, путем выбора разных языковых дорожек, регулировки усиления или положения объектов во время рендеринга в декодере MPEG-H.

Спецификация формата MPEG-H 3D Audio публикуется как ИСО/МЭК 23008-3:2015. В начале 2017 года была опубликована Поправка 3 с описанием профиля пониженной сложности (LC-профиль) формата MPEG-H 3D Audio и дополнительной технологии.

MPEG-H 3D Audio LC Profile поддерживает до 24 выходных громкоговорителей и 56 базовых каналов кодека (28 из которых могут декодироваться одновременно).

Примеры возможных схем расположения громкоговорителей:

- 2 стереосистемы 2.0 (формат 2/0, описанный в Рекомендации МСЭ-R BS.775);
- многоканальное воспроизведение звука в системах 5.1 (формат 3/2, описанный в Рекомендации МСЭ-R BS.775);
- иммерсивный звук 10.2 (система F, описанная в Рекомендации МСЭ-R BS.2051);
- иммерсивный звук 22.2 (система H, описанная в Рекомендации МСЭ-R BS.2051).

Этот стандарт может использоваться в широком спектре применений, включая хранение и передачу стерео- и объемного звука. Он обеспечивает поддержку интерактивности и иммерсивного звука, что важно для удовлетворения требований систем доставки медиаданных последующих поколений, в частности новых систем телевизионного вещания и развлекательных потоковых услуг, а также поддержку контента и услуг виртуальной реальности.

Например, в телевизионном вещании комментарии или разговор могут передаваться как аудиообъекты и объединяться в декодере MPEG-H 3D Audio с иммерсивным каналом. Это позволит эффективно передавать разговор на нескольких языках, а также даст возможность слушателю по своему усмотрению настраивать баланс между речью и другими звуковыми элементами. Эту концепцию можно распространить на другие элементы, обычно не присутствующие в трансляции, такие как звуковое описание для слабовидящих, комментариев режиссера или разговоры участников спортивных состязаний.

MPEG-H 3D Audio LC Profile поддерживает управление громкостью в соответствии с рекомендациями МСЭ-R BS.1770 и МСЭ-R BS.1771, включая сигнализацию соответствия для указания соответствия международным и некоторым региональным стандартам громкости. Он также поддерживает управление расширенным динамическим диапазоном (DRC), применимое к широкому спектру типов устройств как для домашних, так и для мобильных условий.

#### 2 Кодирование

Архитектура кодека MPEG-H 3D Audio построена на принципе перцепционного кодека для сжатия входных сигналов разного класса на основе стандарта унифицированного кодирования речи и звука (USAC) MPEG. USAC позволяет выполнять сжатие монофонических и многоканальных аудиосигналов со скоростью 8 кбит/с на канал и выше.

В целях удовлетворения новых требований, возникающих в контексте 3D-звука, эта технология дополнена инструментами, которые используют перцептивные эффекты 3D-воспроизведения и тем самым еще больше повышают эффективность кодирования.

- Функция интеллектуальное заполнение интервалов (IGF) обеспечивает улучшенные параметры заполнения шумом. IGF – это инструмент, который параметрическим способом восстанавливает части передаваемого спектра, используя подходящую информацию по участкам спектра, смежным по частоте и времени. Кодер управляет выбором и обработкой этих участков спектра на основе анализа входного сигнала. Это позволяет заполнять спектральные интервалы с учетом спектральных коэффициентов, что воспринимается лучше, чем заполнение обычно применяемыми псевдослучайными шумовыми последовательностями.
- Помимо повышения эффективности кодирования, базовый декодер USAC-3D оснащен новыми механизмами сигнализации для разных схем размещения 3D-контента/громкоговорителей и разных типов сигналов в сжатом потоке (звуковой канал – звуковой объект – сигнал НОА).

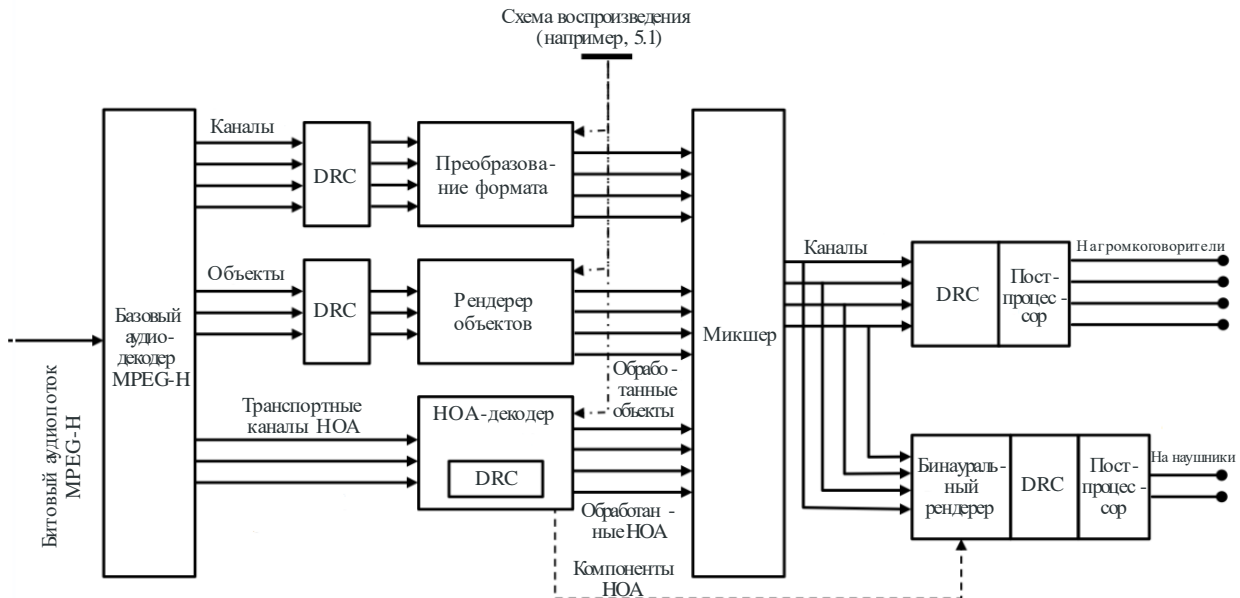
Еще один новый аспект в механизме полезной нагрузки сжатого звука – улучшенное поведение при мгновенном переключении скорости или быстрой настройке в контексте динамической адаптивной потоковой передачи (DASH) MPEG. Для этой цели в синтаксис добавлены так называемые кадры немедленного воспроизведения, которые обеспечивают бесщелевой переход из одного потока в другой. Это особенно целесообразно для адаптивной потоковой передачи по IP-сетям.

### 3 Декодирование

Структурная схема верхнего уровня общей архитектуры MPEG-H 3D Audio LC Profile представлена на рисунке 17.

РИСУНОК 17

Структурная схема верхнего уровня архитектуры аудиодекодера MPEG-H 3D LC Profile



Основными компонентами являются так называемый базовый декодер USAC-3D, набор рендереров для сигналов разного класса и микшер. На первом этапе различные базовые сигналы преобразуются из своих представлений со сжатыми данными с помощью декодера USAC-3D.

Затем сигналы разного класса (формы сигналов канала и объекта или сигналы коэффициентов НОА) подаются в соответствующие рендереры, которые преобразуют эти сигналы для подачи на громкоговорители применительно к конкретной схеме воспроизведения со стороны приемника. Когда

все обработанные сигналы доступны в формате воспроизведения, они объединяются в каскаде микширования для подачи на громкоговорители. Если требуется бинауральное представление, сигнал преобразуется в виртуальную 3D-сцену для воспроизведения в наушниках. В одном потоке MPEG-H можно передавать любую комбинацию сигналов разного типа, например комбинацию сигналов канала вместе с сигналами объекта или НОА-сцену вместе с объектами.

Используются следующие рендереры:

- преобразователь формата для преобразования сигналов канала из их технологического формата в формат воспроизведения;
- рендерер объектов для размещения дорожек статических или динамических объектов в схеме воспроизведения;
- рендерер НОА для преобразования из представления НОА на основе сцены в фактическую схему воспроизведения;
- бинауральный рендерер для преобразования виртуальной компоновки громкоговорителей на выходной сигнал для наушников.

Кроме того, можно управлять воспроизведением и рендерингом сигналов разного класса с помощью пользовательского интерфейса, если эти сигналы маркируются соответствующими метаданными как сигналы, предполагающие возможность интерактивного режима.

## **Приложение 9 (информационное)**

### **DTS-UHD**

#### **1 Введение**

В составе системы кодирования DTS-UHD имеется подсистема кодирования звука (АСЕ), которая служит эффективным средством сжатия звуковых волн. DTS-UHD поддерживает кодирование звука на основе канала (СВА), звуковых полей Ambisonic и звука на основе объекта (ОВА). Система кодирования DTS-UHD стандартизирована документом ETSI TS 104 491 и образует часть спецификации транспортного потока DVB MPEG 2 – ETSI TS 101 154. Она обеспечивает полностью иммерсивное восприятие, а дополнительно к этому путем использования объектов, каналов, звуковых полей или сочетания всех этих средств обеспечивает интерактивность и персонализацию. Кроме того, она поддерживает дополнительные средства обеспечения доступности для пользователей с нарушениями зрения и слуха.

DTS-UHD поддерживает все схемы расположения громкоговорителей, указанные в Рекомендации МСЭ-R BS.2051.

Помимо поддержки звука на основе канала, в DTS-UHD предусмотрена возможность кодирования до 224 дискретных звуковых объектов, которые могут быть далее распределены по 32 группам объектов и 32 презентациям в рамках одного потока.

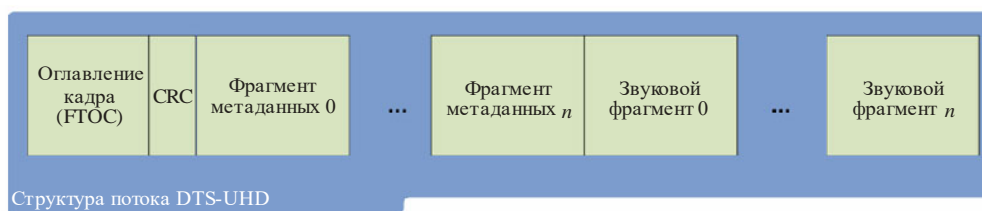
Поток DTS-UHD представляет собой последовательность кадров DTS-UHD, содержащих элементы трех основных типов:

- оглавление кадра (FTOC), которое позволяет переходить непосредственно к нужным элементам кадра;
- фрагменты метаданных;
- звуковые фрагменты.

Структура потока показана на рисунке ниже.



РИСУНОК 18  
Структура потока DTS-UHD



BS.1196-8-18

FTOC присутствует в каждом кадре DTS-UHD и сообщает, к какому типу относится данный кадр – кадр синхронизации или обычный кадр. Кадры синхронизации содержат все параметры, необходимые для того, чтобы инициировать сеанс декодирования.

Звуковые фрагменты содержат сжатые оптимальным образом звуковые волны; эти фрагменты могут быть кадрами синхронизации или обычными кадрами. Звуковой фрагмент разделен на ряд подпотоков, которые могут представлять волны монофонического, стереофонического или LFE-формата.

Фрагмент метаданных полностью описывает звуковой компонент; в частности, он содержит тип компонента, перечень звуковых фрагментов, необходимых для рендеринга компонента, и сведения о громкости и сжатии динамического диапазона (компрессии). Фрагмент метаданных точки доступа (кадр синхронизации DTS-UHD) ссылается на звуковой фрагмент, также являющийся кадром синхронизации, и содержит полный набор метаданных, необходимых для того, чтобы начать кодирование.

## 2 Кодер

ETSI не издал нормативной спецификации кодера ACE DTS-UHD. В кодере реализован полный набор инструментов, позволяющих кодировать множество битовых потоков для поддержки различных конфигураций громкоговорителей.

Технология кодирования DTS-UHD обеспечивает сохранение качества звука и позволяет формировать битовые потоки, отвечающие требованиям к высококачественной передаче, изложенным в Рекомендации МСЭ-R BS.1548 для следующих скоростей передачи:

Конфигурация громкоговорителей	Скорость передачи (кбит/с)
0+2+0 (стерео)	128
0+5+0 (5.1)	192
4+7+0 (7.1.4)	288

Кодер DTS-UHD обеспечивает полный контроль над размещением кадров синхронизации, чтобы звуковая служба могла управлять своими точками входа.

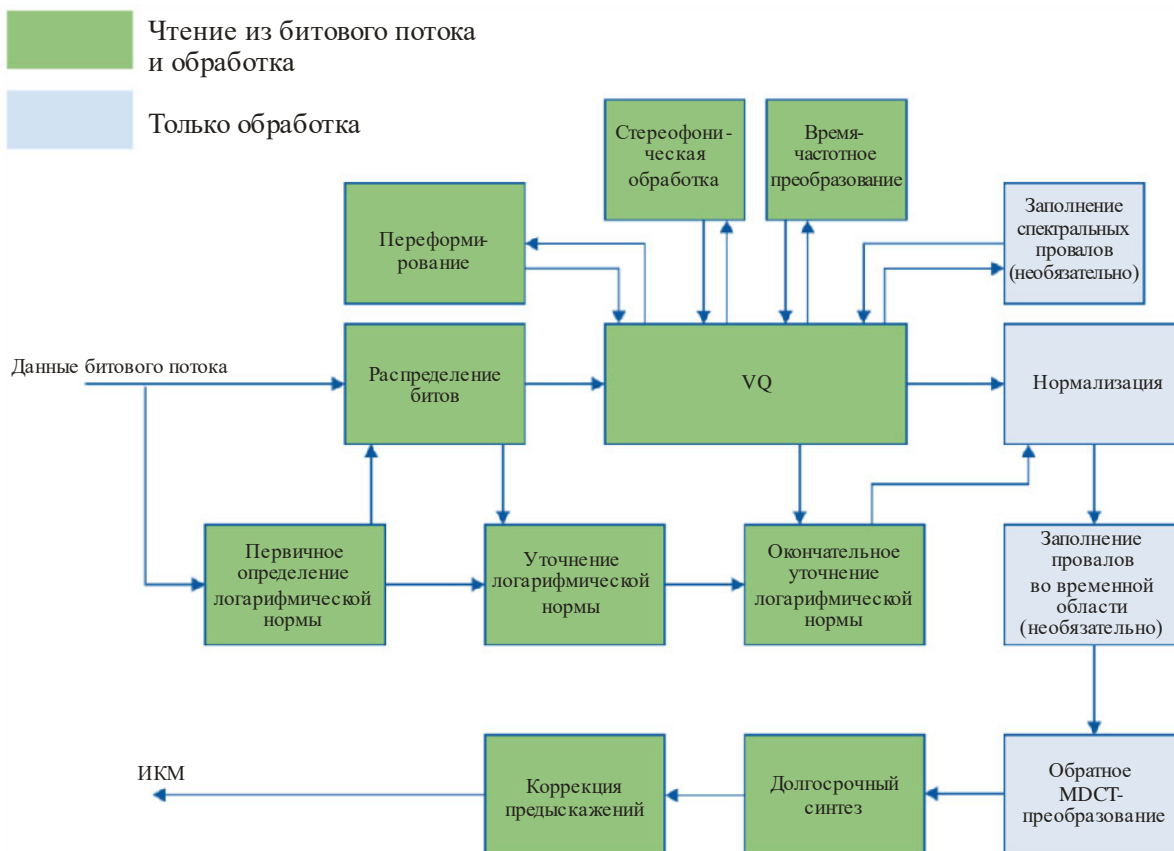
Функциональность кодирования с переменной скоростью передачи (VBR) в DTS-UHD гарантирует постоянство скорости передачи (CBR) в каждой группе кадров (GoF), то есть в интервале между двумя последовательно идущими кадрами синхронизации. Кадры синхронизации задаются кодером, и интервал синхронизации может по необходимости меняться динамически в процессе кодирования. Кодер DTS-UHD может также задавать нужную максимальную мгновенную пиковую скорость передачи в режиме VBR. Модель управления скоростью передачи (BRC) в DTS-UHD устанавливает логику динамического распределения битов каждого звукового кадра по различным объектам, каналам и парам каналов в битовом потоке. В модели BRC для кодирования наиболее значимого с точки зрения восприятия контента в кадре отводится большее число бит при одновременном стремлении обеспечить одинаковое качество звука на множестве всех объектов и каналов кадра. Это позволяет повысить эффективность кодирования при повышенном числе звуковых элементов, пользуясь неоднородностью

распределения значимого с точки зрения восприятия контента по объектам и каналам. Кодер также динамически распределяет биты звуковым кадрам для обеспечения постоянства воспринимаемого качества звука во времени (с ограничениями, накладываемыми CBR по группам кадров и заданной максимальной пиковой скоростью передачи). Следует отметить, что функциональность BRC не требует передачи дополнительной информации и что работа декодера не зависит от этой функциональности. Каждый объект, канал или пара каналов остаются независимо декодируемыми и полноценно воспроизводимыми в режиме BRC.

### 3 Декодер

Декодер ACE работает со стереофоническими, монофоническими и LFE-потоками ACE. В кадре ACE потоки разделены по этим трем типам. Стереофонические потоки дополнительно оптимизированы за счет информации, общей для обоих каналов.

РИСУНОК 19  
Декодер ACE



BS.1196-8-19

Декодирование монофонического или стереофонического потока ACE состоит из следующих этапов:

**Первичное определение логарифмической нормы:** по битовому потоку в первом приближении определяются значения мощности в спектральной полосе, которые представляются в логарифмическом масштабе и называются значениями логарифмической нормы.

**Распределение битов:** для каждой спектральной полосы рассчитывается число битов, доступных для декодирования спектральной полосы с нормализацией. Этот расчет зависит от информации, содержащейся в битовом потоке, и (потенциально) значений первичной логарифмической нормы.

**Уточнение логарифмической нормы:** небольшая часть распределенных битов используется для уточнения значений логарифмической нормы.

VQ (векторное квантование): большинство распределенных битов используется для воссоздания нормализованных значений спектральных частот (по одной спектральной полосе за прием).

Переформирование: предварительно воссозданные прореженные спектральные полосы, соответствующие исходным "полным" полосам, переформируются, чтобы ближе аппроксимировать полный спектр.

Стерефоническая обработка: в случае стерефонических потоков два воссозданных канала преобразуются к исходному (немикшированному) виду.

Время-частотное преобразование: спектральные полосы подвергаются простому преобразованию Хаара, в результате чего все они приводятся к одному и тому же время-частотному представлению.

Заполнение спектральных провалов: спектральные полосы, которые исходно содержали некоторую энергию, но в воссозданном виде содержат только нулевые значения, заполняются "шумом", который генерируется по более ранним воссозданным значениям спектральных полос.

Окончательное уточнение логарифмической нормы: оставшиеся после векторного квантования биты используются для окончательного уточнения значений логарифмической нормы.

Нормализация: спектральные полосы восстанавливаются с приведением к значениям мощности, соответствующим окончательным значениям логарифмической нормы.

Заполнение провалов во временной области: спектральные полосы, содержащие только нулевые значения, могут заполняться синтетическим шумом, зависящим от мощности в спектральной полосе в предыдущих кадрах.

Обратное MDCT-преобразование: из спектрального представления сигнал преобразуется к представлению во временной области (ИКМ).

Долгосрочный синтез: для воссоздания компонентов сигнала, удаленных на этапе кодирования, применяется фильтр долгосрочного синтеза.

---