

Международный союз электросвязи

МСЭ-R

Сектор радиосвязи МСЭ

Рекомендация МСЭ-R BS.1196-2
(03/2010)

**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 Приложении 1 к Резолюции 1 МСЭ-R. 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
BS	Радиовещательная служба (звуковая)
BT	Радиовещательная служба (телевизионная)
F	Фиксированная служба
M	Подвижная спутниковая служба, спутниковая служба радиоопределения, любительская спутниковая 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 1 МСЭ-R.

Электронная публикация
Женева, 2010 г.

© ITU 2010

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.1196-2*,**

Кодирование звуковых сигналов для цифрового радиовещания

(Вопрос МСЭ-R 19/6)

(1995-2001-2010)

Сфера применения

В настоящей Рекомендации определяются системы кодирования источников, применимых для цифрового звукового и телевизионного радиовещания. В ней также определяется система, применимая для совместимого в обратном направлении многоканального расширения систем цифрового звукового и телевизионного радиовещания.

Ассамблея радиосвязи МСЭ,

учитывая,

- a) что требования пользователей к системам кодирования звуковых сигналов для цифрового радиовещания указаны в Рекомендации МСЭ-R BS.1548;
- b) что многоканальная звуковая система с сопровождающим изображением и без него является предметом Рекомендации МСЭ-R BS.775 и что высококачественная многоканальная звуковая система, использующая эффективное уменьшение скорости передачи, является необходимой для системы цифрового радиовещания;
- c) что субъективная оценка звуковых систем с небольшим ухудшением качества, в том числе многоканальных звуковых систем, является предметом Рекомендации МСЭ-R BS.1116;
- d) что субъективная оценка звуковых систем с промежуточным качеством звука является предметом Рекомендации МСЭ-R BS.1534 (MUSHRA);
- e) что кодирование с низкой скоростью передачи для звуковых сигналов высокого качества было протестировано Сектором радиосвязи МСЭ;
- f) что единообразие методов кодирования источников звуковых сигналов в различных службах может обеспечить повышенную гибкость систем и снижение затрат на приемные устройства;
- g) что несколько радиовещательных служб уже используют или определили использование звуковых кодеков из семейств MPEG-1, MPEG-2, MPEG-4, AC-3 и E-AC-3;
- h) 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;
- l) что при введении многоканальной звуковой системы необходимо принимать во внимание существующие моно- и стереоприемники;

* В 2003 году 6-я Исследовательская комиссия по радиосвязи внесла редакционные поправки в настоящую Рекомендацию в соответствии с Резолюцией МСЭ-R 44.

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

м) что совместимое в обратном направлении многоканальное расширение существующей системы кодирования звуковых сигналов может обеспечить большую эффективность скорости передачи, чем одновременная передача,

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:

- MPEG-4 HE AAC v2, как указано в ИСО/МЭК 14496-3:2009;
- E-AC-3 как указано в ETSI TS 102 366 (2008-08);

ПРИМЕЧАНИЕ 1. – MPEG-4 HE AAC v2 и E-AC-3 являются более гибкими расширенными наборами MPEG-4 AAC-LC и AC-3.

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 (2008-08);

ПРИМЕЧАНИЕ 1. – ИСО/МЭК 11172-3 иногда именуется 13818-3, поскольку эта спецификация включает 11172-3 путем ссылки.

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 1. – 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 пп. 1 и 2 раздела *рекомендует*.

4 чтобы для линий распространения и подачи можно было использовать кодирование уровня II ИСО/МЭК 11172-3 на скорости не менее 180 кбит/с на звуковой сигнал (т. е. на моносигнал или на компонент независимо кодируемого стереосигнала), за исключением вспомогательных данных;

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;

8 чтобы соблюдение настоящей Рекомендации было добровольным. Вместе с тем Рекомендация может содержать определенные обязательные положения (например, для обеспечения функциональной совместимости или применимости), и соблюдение Рекомендации достигается, когда выполняются все эти обязательные положения. Для выражения требований используются слова "shall" ("должен", "обязан") или некоторые другие обязывающие термины, такие как "must" ("должен"), а также их отрицательные эквиваленты. Использование таких слов не предполагает полного или частичного соблюдения настоящей Рекомендации,

далее рекомендует

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

ПРИМЕЧАНИЕ 1. – Информацию о кодах, включенных в настоящую Рекомендацию, можно найти в Дополнениях 1–4.

Дополнение 1

Звуковые MPEG-1 и MPEG-2, уровень II и III

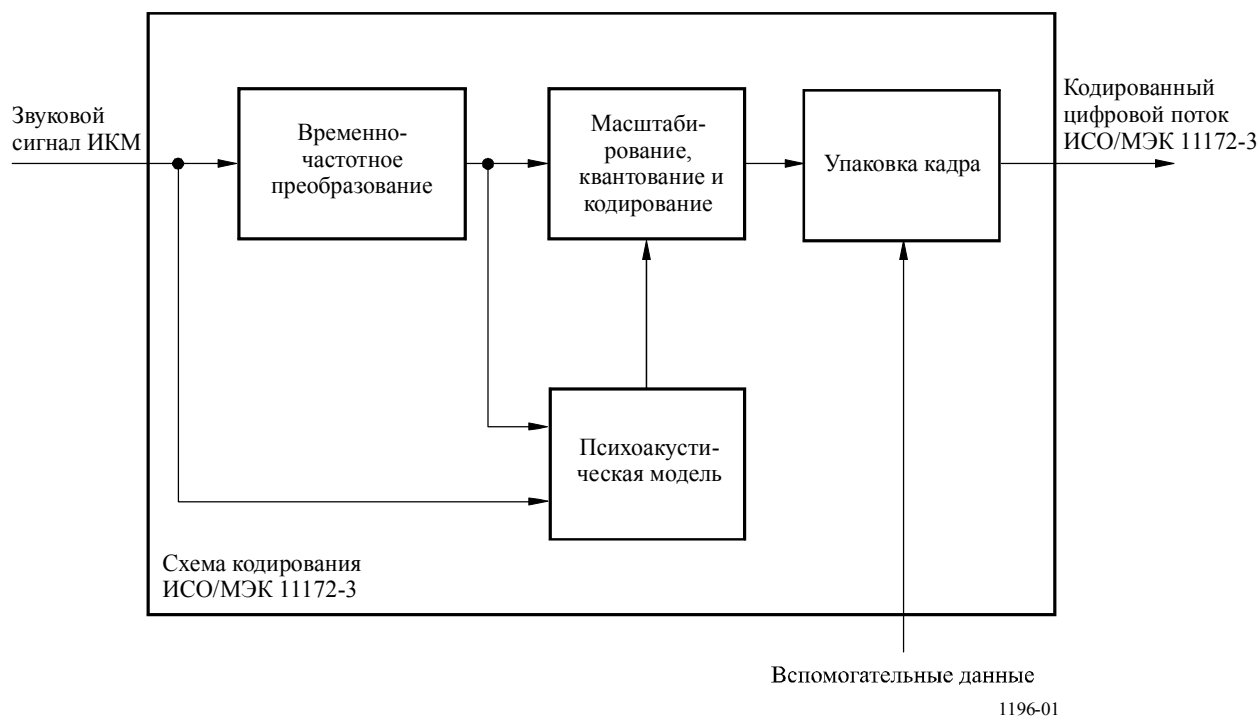
1 Кодирование

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, выработывал 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 рис. 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 "упаковки 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



2 Уровни

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

Уровень I: на этом уровне производятся 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 компрессирования 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:

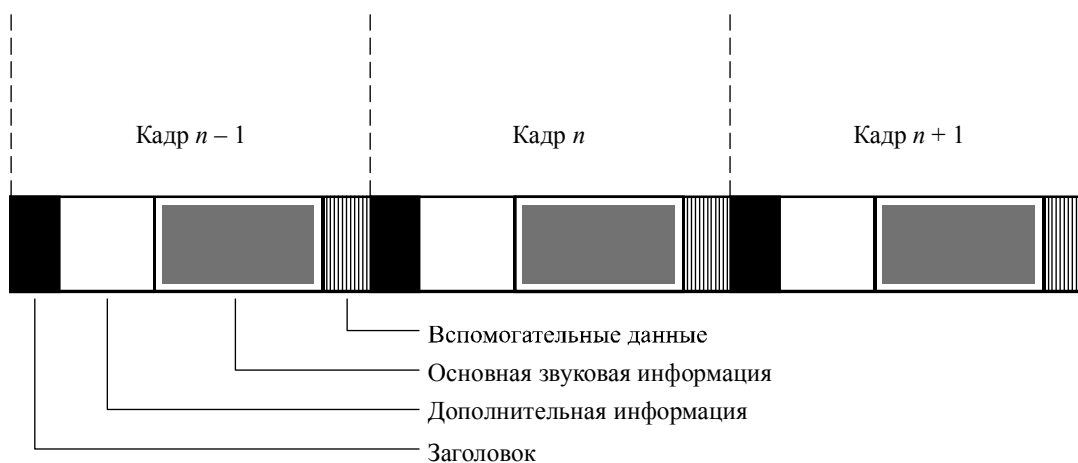
- одиночный канал;
- двойной канал (два независимых звуковых сигнала, закодированных в одном цифровом потоке, например при двуязычном приложении);
- стереофония (левый и правый сигналы стереофонической пары, закодированные в одном цифровом потоке);
- комбинированная стереофония (левый и правый сигналы стереофонической пары, закодированные 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 рис. 2 для уровня II и на рис. 3 для уровня III. Кодированный поток битов состоит из последовательных кадров. В зависимости от уровня кадр включает следующие поля:

РИСУНОК 2

Формат цифрового потока уровня II ИСО/МЭК 11172-3

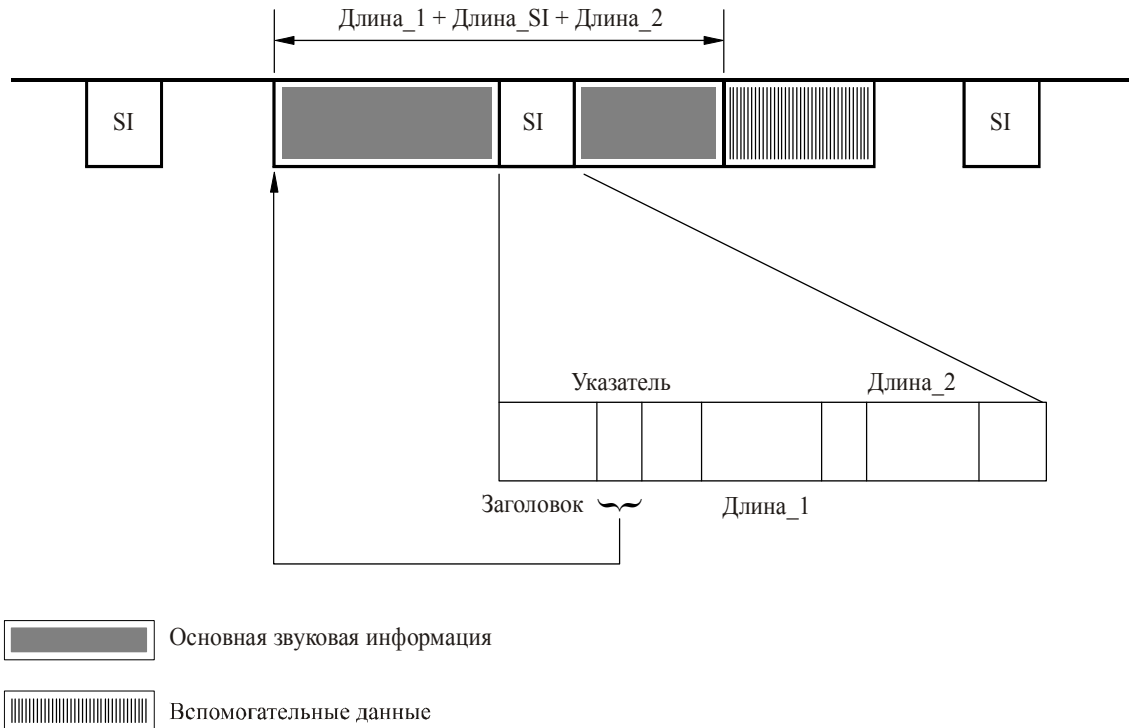


Уровень II:

Заголовок:	часть цифрового потока, содержащая информацию о синхронизации и статусе;
Дополнительная информация:	часть цифрового потока, содержащая информацию о распределении битов и о 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



Уровень III:

Дополнительная информация (SI):	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;
Служебная информация:	часть цифрового потока, содержащая сведения, определяемые пользователем.

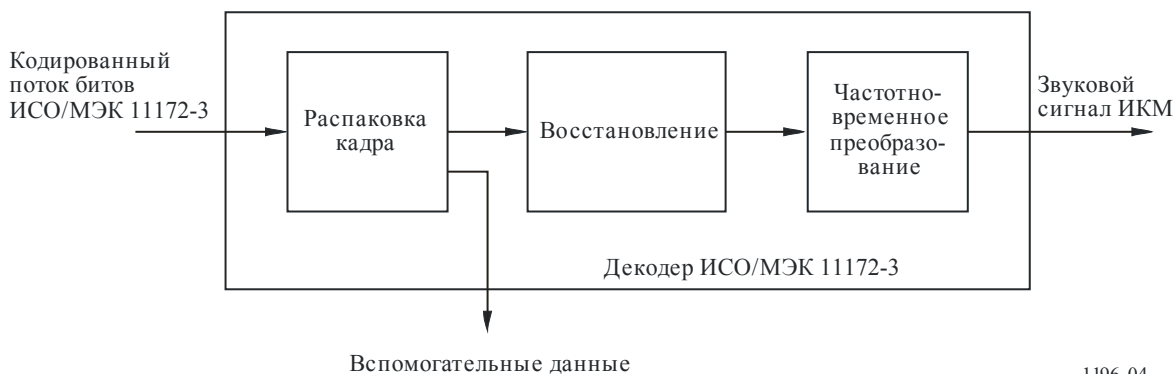
1196-03

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



Дополнение 2

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

Число громкоговорителей	Звуковые синтаксические элементы, перечисленные в порядке приема	Отображение по умолчанию элемента на громкоговоритель
1	single_channel_element одноканальный_элемент	Основной центральный громкоговоритель
2	channel_pair_element элемент_пары_каналов	Левый и правый основные громкоговорители
3	single_channel_element() одноканальный_элемент()	Основной центральный громкоговоритель
	channel_pair_element() элемент_пары_каналов	Левый и правый основные громкоговорители
4	single_channel_element() одноканальный_элемент()	Основной центральный громкоговоритель
	channel_pair_element() элемент_пары_каналов()	Левый и правый основные громкоговорители
	single_channel_element() одноканальный_элемент()	Тыловой громкоговоритель объемного звучания
5	single_channel_element() одноканальный_элемент()	Основной центральный громкоговоритель
	channel_pair_element() элемент_пары_каналов()	Левый и правый основные громкоговорители
	channel_pair_element() элемент_пары_каналов()	Левый и правый тыловые громкоговорители объемного звучания
5 + 1	single_channel_element() одноканальный_элемент()	Основной центральный громкоговоритель
	channel_pair_element() элемент_пары_каналов()	Левый и правый основные громкоговорители
	channel_pair_element() элемент_пары_каналов()	Левый и правый тыловые громкоговорители объемного звучания
	Lfe_element() элемент_НЧЭ()	Громкоговоритель низкочастотных эффектов

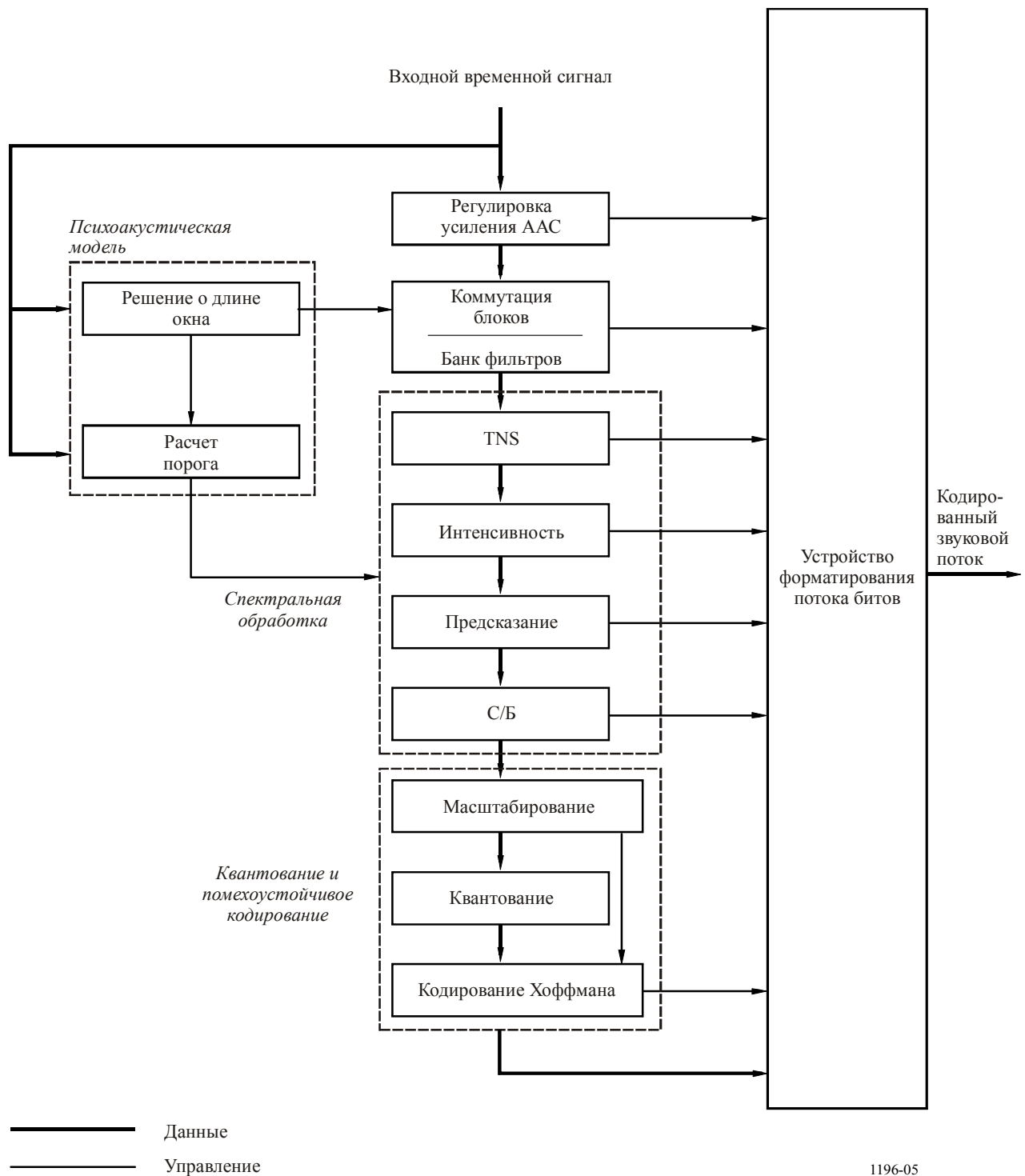
Число громкоговорителей	Звуковые синтаксические элементы, перечисленные в порядке приема	Отображение по умолчанию элемента на громкоговоритель
7 + 1	single_channel_element() одноканальный_элемент()	Основной центральный громкоговоритель
	channel_pair_element() элемент_пары_каналов()	Левый и правый центральные основные громкоговорители
	channel_pair_element() элемент_пары_каналов()	Левый и правый наружные основные громкоговорители
	channel_pair_element() элемент_пары_каналов()	Левый и правый тыловые громкоговорители объемного звучания
	lfe_element() элемент_НЧЭ	Громкоговоритель низкочастотных эффектов

2 Кодирование

Основная структура кодера AAC MPEG-2 показана на рис. 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 Хоффмана 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 используется в процессе квантования для минимизации слышимого шума квантования и, кроме того, для выбора соответствующего средства кодирования.

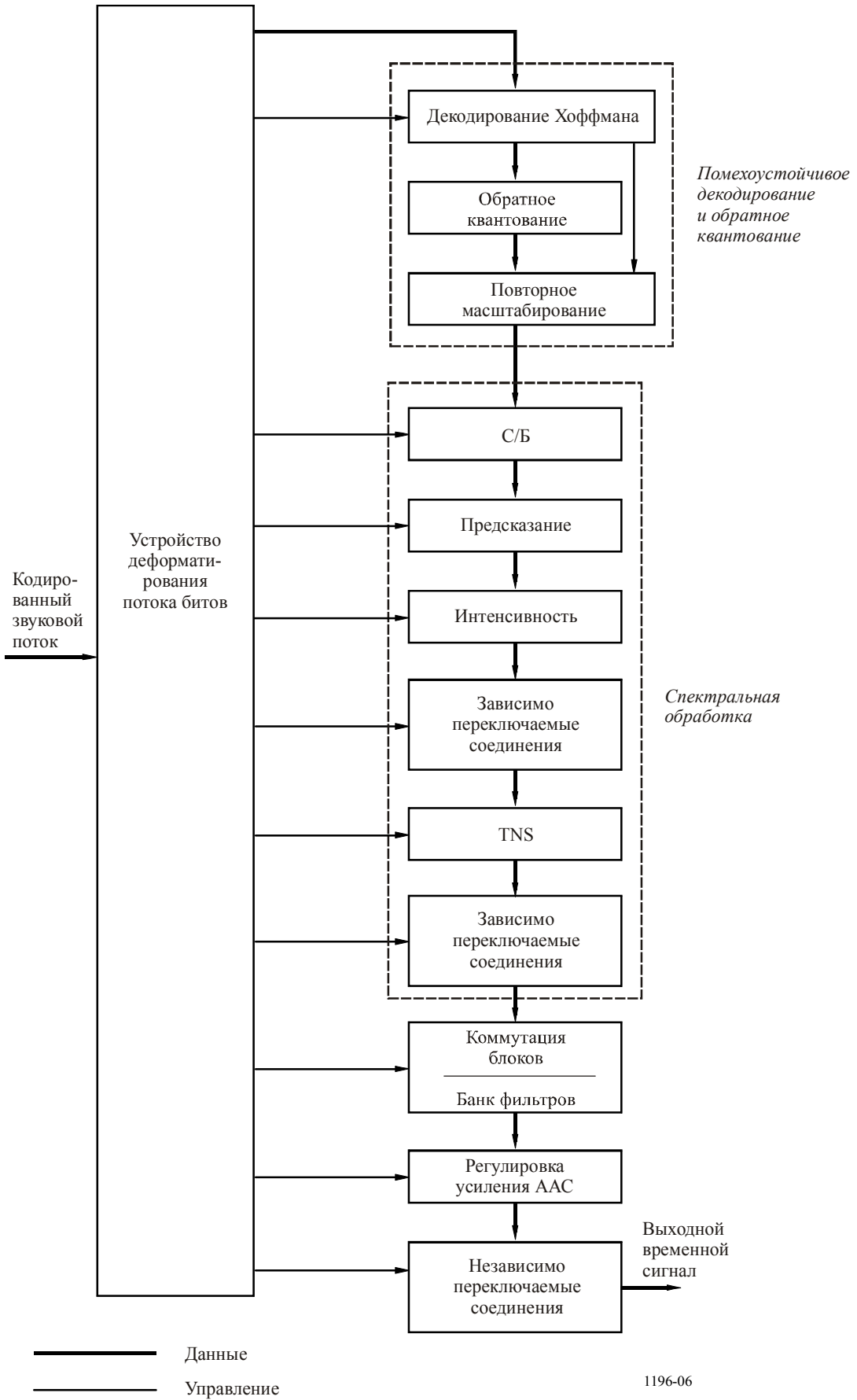
РИСУНОК 5
Функциональная схема кодера AAC MPEG-2



3 Декодирование

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

РИСУНОК 6
Функциональная схема декодера AAC MPEG-2

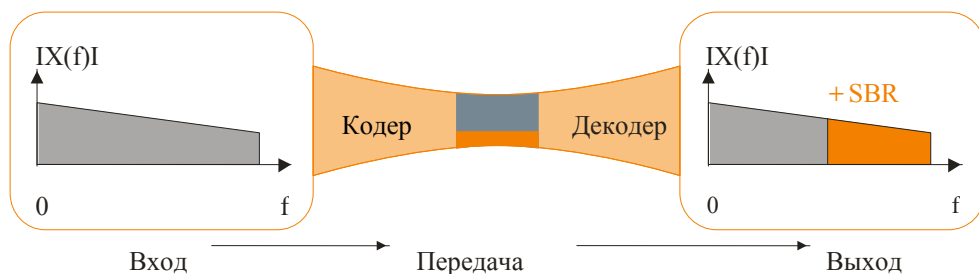


Функции декодера состоят в обнаружении описания квантованного звукового спектра в потоке битов, декодировании квантованных значений и другой информации о восстановлении квантованных спектров, обработке восстановленных спектров с помощью любого устройства, действующего в потоке битов, с целью получения 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.



1196

Поток битов HE AAC представляет собой расширение потока битов звука AAC. Дополнительные данные SBR встраиваются в элемент наполнения 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.



1196

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, который неспособен 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.

Дополнение 3

Звуковые AC-3 E-AC-3

1 Кодирование

Цифровой алгоритм сжатия AC-3 может кодировать от одного 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 рис. 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 AC-3. Поток битов AC-3 является последовательностью кадров AC-3.

РИСУНОК 7

Кодер АС-3



1196-07

Реальный кодер АС-3 гораздо сложнее, чем это показано на рис. 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 с целью 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 рис. 8, должен синхронизировать кодированный поток битов, проверять ошибки и деформатировать различные типы данных, как, например, кодированную спектральную огибающую и квантованные мантисы. Результаты работы стандартной программы распределения битов используются для распаковки и деквантования мантисс. Спектральная огибающая декодируется для создания экспонент. Экспоненты и мантисы обратно преобразуются во временную область для создания декодированных временных выборок ИКМ.

РИСУНОК 8
Декодер AC-3



Реальный декодер AC-3 сложнее, чем это показано на рис. 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 E-AC-3

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

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).
- Подпотoki могут нести дополнительные звуковые каналы, например 7.1 chs или комментаторские дорожки.

Дополнение 4

MPEG Surround

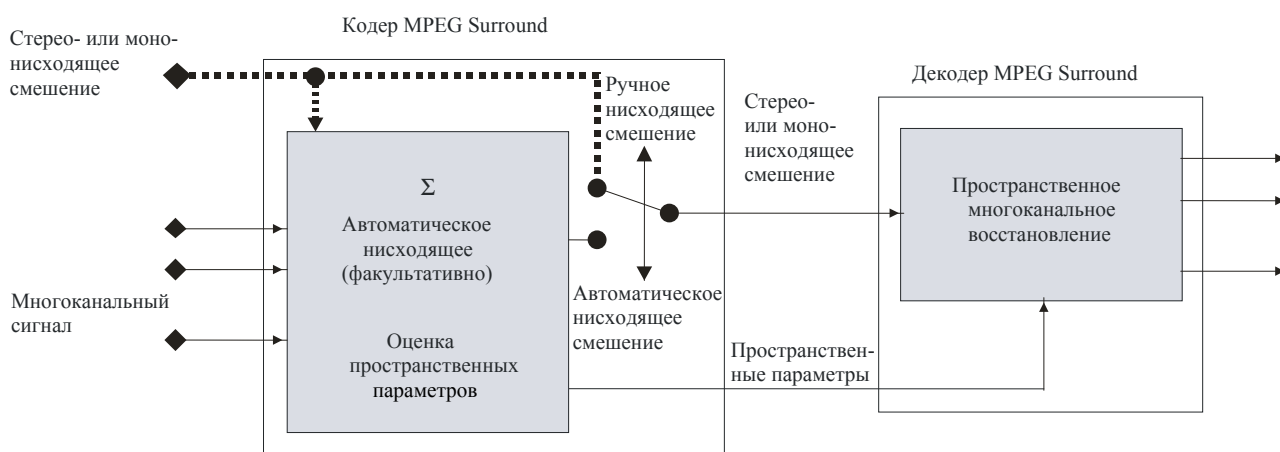
1 Введение

Технология ISO/IEC 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 стерео- нисходящее 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 моно- сигнал нисходящего смешения. Оснащенные MPEG декодеры сначала декодируют моно- или стерео- нисходящее смешение, а затем используют параметры пространственного изображения, извлеченные из потока вспомогательных данных, для генерирования высококачественного многоканального звукового сигнала.

На рис. 9 показан принцип MPEG Surround.

РИСУНОК 9

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



1196-09

Используя MPEG Surround, существующие услуги легко усовершенствовать для обеспечения "окружающего звука" способом, совместимым в обратном направлении. Стереодекодер 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 смещением может использоваться созданное ввне нисходящее смещение ("художественное смещение"). Нисходящее смещение 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 смещения. В качестве примера: для многоканального входного сигнала а 5.1 имеются три различные конфигурации; одна конфигурация для базирующихся на нисходящем смещении стереосистем (конфигурация 525) и две различные конфигурации для базирующихся на нисходящем смещении моносистем (конфигурация 515₁ и 515₂, 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 стерео нисходящее смещение 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 смещения: система MPEG Surround способна обрабатывать не только генерированное кодером нисходящее смещение, но и художественное смещение, 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 приложений, в которых нисходящее смещение 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

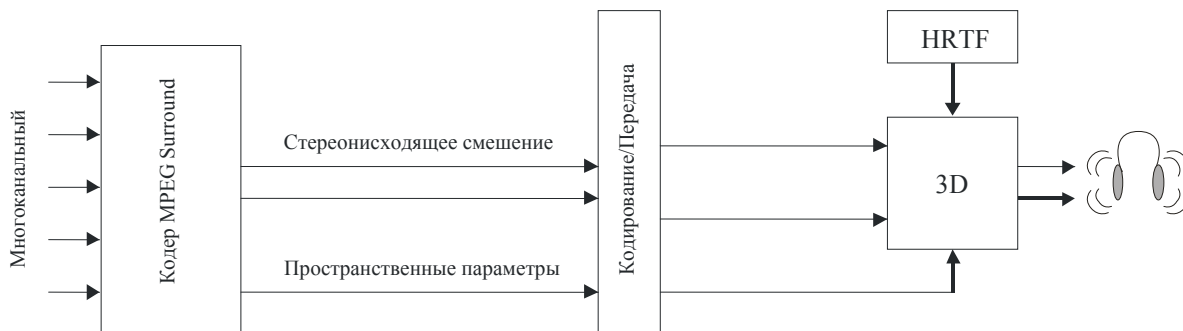
Наряду с обработкой многоканального выходного сигнала декодер MPEG Surround также поддерживает обработку альтернативных конфигураций выходного сигнала:

- Виртуальный окружающий звук: система MPEG Surround может эксплуатировать пространственные параметры для сведения нисходящего смещения 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 смещении, без необходимости применения декодера MPEG Surround. Вместе с тем декодер MPEG Surround может отменить обработку виртуального окружающего звука при нисходящем смещении и повторно применить альтернативный виртуальный окружающий звук. Основной принцип отображен на рис. 10.

РИСУНОК 10

Декодирование виртуального окружающего звука MPEG Surround



1196-10

- Расширенный матричный режим: в случае традиционного стереоконтента, когда побочная пространственная информация отсутствует, MPEG Surround способен оценить побочную пространственную информацию из нисходящего смещения 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 и 2 способны декодировать все битовые потоки уровней 2 и 3, 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).