

## РЕКОМЕНДАЦИЯ МСЭ-R BS.1115-1\*,\*\*

**Кодирование звука с низкой цифровой скоростью**

(Вопрос МСЭ-R 19/6)

(1994-2004)

Ассамблея радиосвязи МСЭ,

*учитывая,*

- a) что высококачественные цифровые звуковые службы доступны потребителю через различные средства передачи информации, такие как компактные диски (CD), магнитная лента и радиовещание;
- b) что основа для цифровых звуковых систем звукового радиовещания содержится в Рекомендациях МСЭ-R BS.774 и МСЭ-R ВО.789;
- c) что требования пользователей к системам кодирования звука для цифрового радиовещания заданы в Рекомендации МСЭ-R BS.1548;
- d) что основное качество звука и стереоотображения, требуемое для цифрового звукового радиовещания, должно быть фактически неотличимо от основного качества CD;
- e) что характеристики систем кодирования звука с низкой цифровой скоростью, используемых в соединительных линиях программы (линии доставки, распределения и комментариев), подводимых к цифровым звуковым системам радиовещания, должны быть такими, чтобы можно было доставлять звуковые сигналы самого высокого качества к цифровому передатчику звукового радиовещания;
- f) что для случаев линий доставки и распределения это предполагает уровень характеристик, превышающий такой уровень для CD, из-за требований к запасам на обработку и каскадирование и (для случая линий доставки) к запасам на перегрузку или "просвет";
- g) что для случая комментаторских линий это предполагает уровень характеристик, способный доставлять слушателю речь превосходного качества, но в случае, если по таким линиям передается материал музыкальных программ, ожидается снижение уровня качества;
- h) что кодирование звука с низкой цифровой скоростью обычно вызывает задержку, составляющую несколько десятков миллисекунд, которая может сказаться на эксплуатационной практике, как, например, при подаче сигналов в программах, принимаемых непосредственно с эфира;
- j) что кодирование звука с низкой цифровой скоростью для звука высокого качества было испытано в Секторе радиосвязи МСЭ, и для некоторых применений были получены удовлетворительные эксплуатационные результаты (см. Дополнение 6 и Рекомендацию МСЭ-R BS.1548);
- k) что обычные объективные методы (например, для измерения отношения сигнал-шум и искажений), вероятно, больше не подходят для оценки качества систем, использующих схемы кодирования звука с низкой цифровой скоростью, и что методы объективных перцептивных испытаний, которые даются в Рекомендации МСЭ-R BS.1387, могут со временем дополнить или заменить обычные объективные методы оценки и, в конечном счете, дополнить методы субъективной оценки,

---

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

\*\* Данная Рекомендация должна быть доведена до сведения 9-й и 15-й Исследовательских комиссий по стандартизации электросвязи, Международной организации по стандартизации/Международной электротехнической комиссии (ИСО/МЭК) (JTC 1/SC 29/WG 11), Европейского союза радиовещания (ЕСР) и Общества инженеров-акустиков (AES).

*рекомендует,*

**1** чтобы для цифровых приложений звукового радиовещания, перечисленных в пп. 4–6 раздела *рекомендует*, использовалось кодирование звука с низкой цифровой скоростью, описанное в стандартах ИСО/МЭК 11172-3:1993 (см. Примечание 1), ИСО/МЭК 13818-7:2003 или АС-3 в соответствии с Приложением 2 к Рекомендации МСЭ-R BS.1196 (краткое описание этих стандартов дано в Дополнениях 1, 3 и 4);

ПРИМЕЧАНИЕ 1. – Иногда кодек ИСО/МЭК 11172-3 может быть указан как 13818-3, поскольку этот кодек включает 11172-3 посредством ссылки.

**2** чтобы для всех применений частота дискретизации равнялась либо 48 кГц, либо 32 кГц;

**3** чтобы входной сигнал кодера звука с низкой цифровой скоростью был свободен от предсказаний и чтобы в кодере также не применялось никаких предсказаний;

**4** чтобы для цифровой передачи звукового радиовещания (см. Примечание 1) использовались следующие кодеки:

- кодирование уровня II ИСО/МЭК 11172-3 со скоростью 128 кбит/с для монофонического сигнала и со скоростью 256 кбит/с (т. е.  $2 \times 128$  кбит/с) для стереофонического сигнала при независимом кодировании левого и правого компонентов стереосигнала (см. Примечание 2),
- кодирование ИСО/МЭК 13818-7 со скоростью не менее 144 кбит/с для стереосигнала,
- кодирование АС-3 со скоростью не менее 192 кбит/с для стереосигнала;

ПРИМЕЧАНИЕ 1. – Дополнение 1 к Приложению 2 к Рекомендации МСЭ-R BS.1548 содержит информацию о системах кодирования, которые удовлетворяют требованию к качеству и другим требованиям пользователей к передаче.

ПРИМЕЧАНИЕ 2. – Ожидается, что радиовещательные организации пожелают иметь в передаваемом сигнале некоторую емкость для дополнительной информации. Однако они должны быть поставлены в известность, что снижение скорости передачи звуковых сигналов обычно влияет на качество звука.

**5** чтобы для линий доставки и распределения использовался уровень II кодирования ИСО/МЭК 11172-3 при цифровой скорости, равной не менее 180 кбит/с на один звуковой сигнал (т. е. на монофонический сигнал или на одну составляющую независимо кодируемого стереофонического сигнала), исключая вспомогательную информацию (см. Дополнение 2 и Примечание 2), но чтобы для особого случая одной линии распределения без последующего каскадирования (см. Примечание 3) могла использоваться цифровая скорость, равная не менее 120 кбит/с на один звуковой сигнал;

ПРИМЕЧАНИЕ 1. – Дополнение 1 к Приложению 1 Рекомендации МСЭ-R BS.1548 содержит информацию о системах кодирования, которые удовлетворяют требованию к качеству и другим требованиям пользователей к линиям подачи и распределению программ.

ПРИМЕЧАНИЕ 2. – При применении настоящей Рекомендации необходимо учитывать, что некоторые сети электросвязи в настоящее время не позволяют неограниченно использовать каналы со скоростью передачи 64 кбит/с.

ПРИМЕЧАНИЕ 3. – Если, например, подача сигналов на аналоговый передатчик осуществляется через одну линию распределения или если сигнал, закодированный до конечного стандарта излучения в студии, подается на передатчик для его излучения без последующего декодирования и кодирования.

**6** чтобы для комментаторских линий использовался уровень III кодирования ИСО/МЭК 11172-3 при цифровой скорости, равной не менее 60 кбит/с, исключая вспомогательную информацию для монофонических сигналов (Примечание 1), и не менее 120 кбит/с, исключая вспомогательную информацию для стереофонических сигналов, используя совместное стереокодирование (см. Дополнение 2 и Примечание 2);

ПРИМЕЧАНИЕ 1. – Результаты, полученные при проведении испытаний в 1993 году, заключавшихся в прослушивании через наушники для оценки характеристик уровня III монофонического кодека для комментаторских линий со скоростью 60 кбит/с, показали, что он не удовлетворяет требуемому качеству при передаче речевых сигналов. Настоящее Примечание останется в Рекомендации до тех пор, пока характеристики комментаторского кодека не будут удовлетворять этим требованиям.

ПРИМЕЧАНИЕ 2. – При применении настоящей Рекомендации необходимо учитывать, что некоторые сети электросвязи в настоящее время не позволяют неограниченно использовать каналы со скоростью передачи 64 кбит/с.

7 чтобы при реализации кодеров и декодеров задержка сводилась к минимуму;

ПРИМЕЧАНИЕ 1. – Со стандартами ИСО/МЭК 11172-3 (звуковой сигнал MPEG-1) и 13818-7 (MPEG-2 AAC) в электронной форме можно ознакомиться по следующему адресу: <http://www.iso.org/itu>.

## Дополнение 1

### Описание метода кодирования ИСО/МЭК 11172-3 Кодирование движущихся изображений и связанных с ними звуковых сигналов со скоростью примерно до 1,5 Мбит/с для цифрового носителя хранения информации – Часть 3: Звуковой сигнал

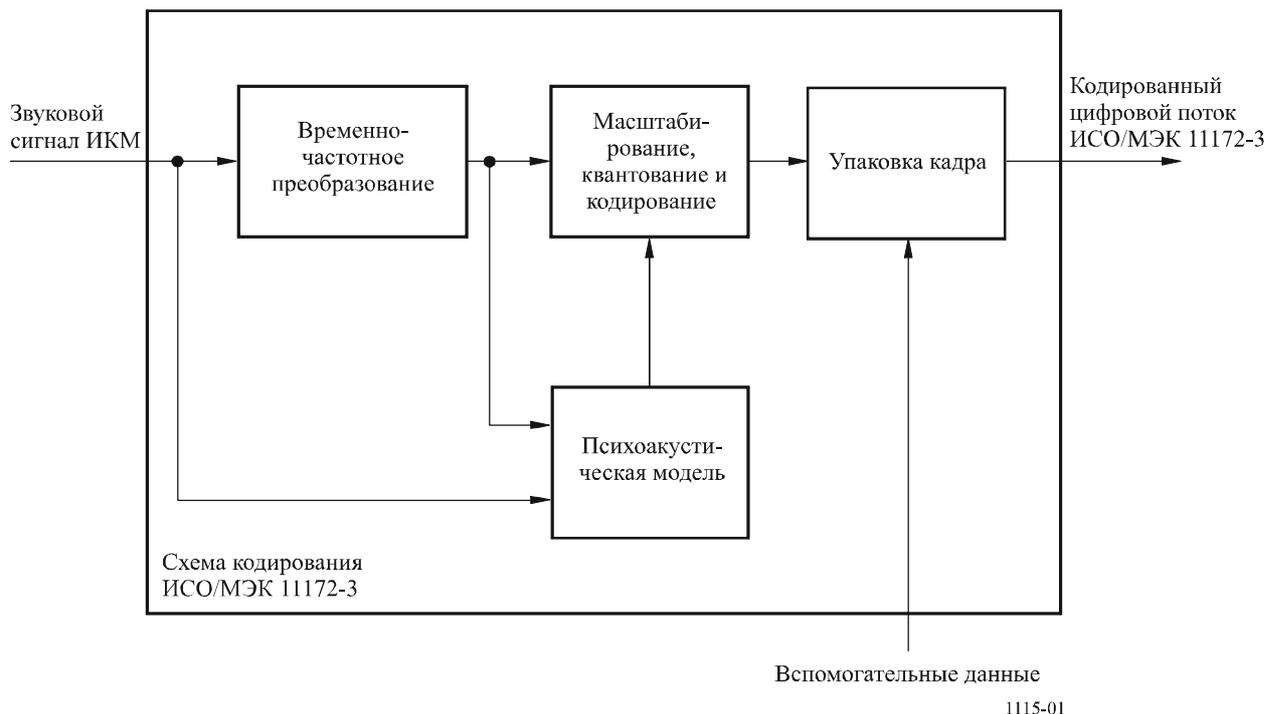
#### 1 Кодирование

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). Однако выход кодера должен быть таким, чтобы декодер, соответствующий данной 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 г., 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 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.

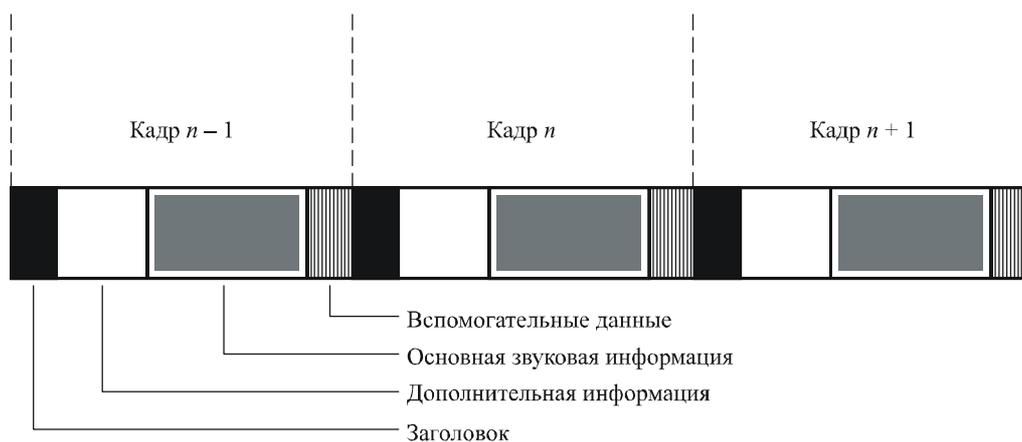
Существуют четыре различных режима, возможных на любом из уровней:

- одиночный канал;
- двойной канал (два независимых звуковых сигнала, 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

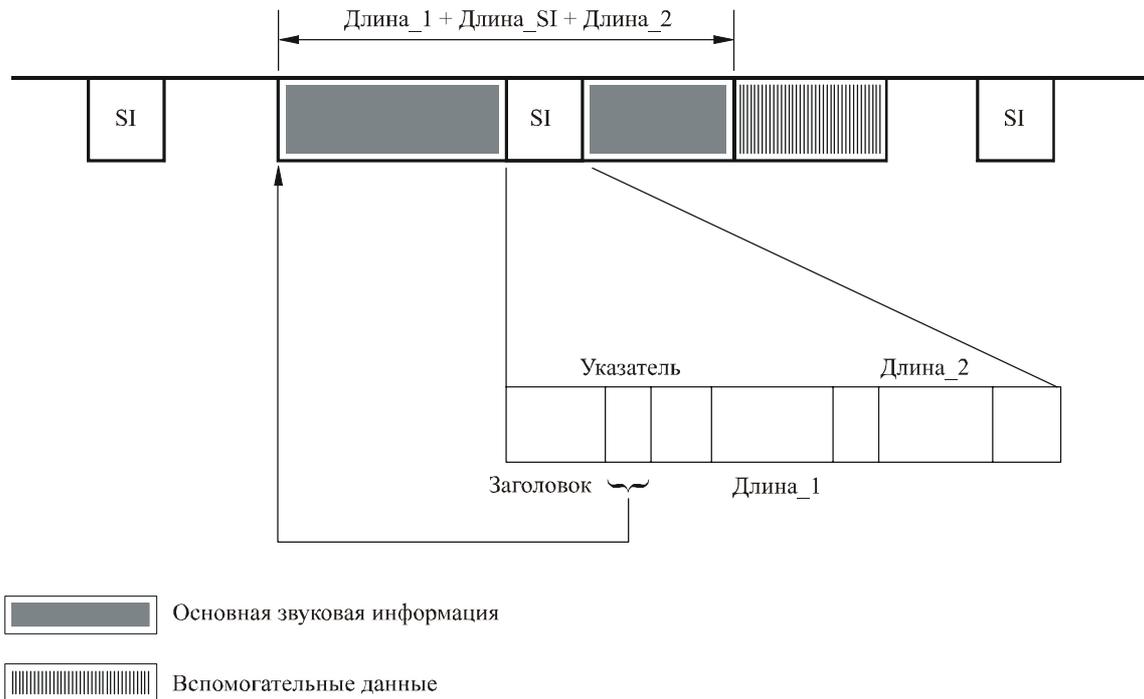


Уровень II:

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.

1115-02

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.

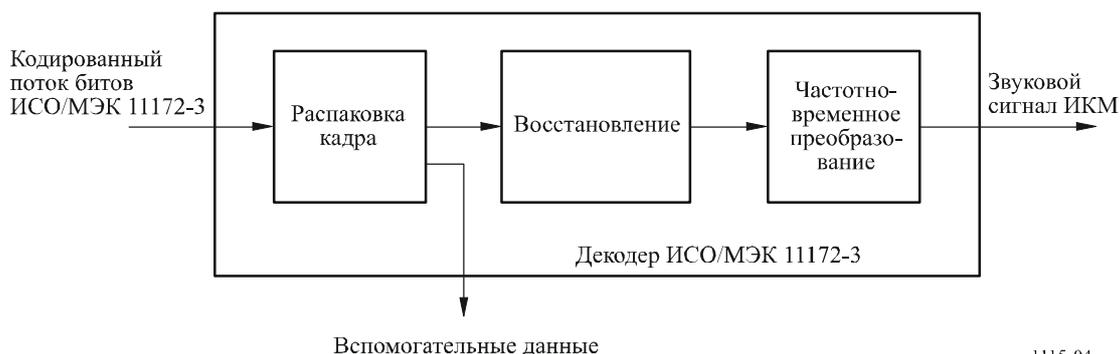
1115-03

#### 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



1115-04

## Дополнение 2

### Руководство по эксплуатации кодеков ИСО/МЭК 11172-3 при определенных цифровых скоростях

Настоящее Дополнение предназначено для предоставления некоторого руководства для приложений, в которых может понадобиться эксплуатация кодека ИСО/МЭК 11172-3 при определенных цифровых скоростях, кратко перечисленных в таблице цифровых скоростей стандарта ИСО/МЭК (таблица 1). Эти конкретные цифровые скорости битов, возможно, должны выбираться с учетом существующей скорости битов в канале передачи и соответствующей минимальной цифровой скорости битов, приведенной в настоящей Рекомендации.

Более подробное обсуждение приведено в Рекомендации МСЭ-T J.52 по цифровой передаче высококачественных сигналов звуковых программ с использованием одного, двух или трех каналов со скоростью 64 кбит/с на один монофонический сигнал (и до 6 на один стереофонический сигнал).

В стандарте ИСО/МЭК кратко перечислены следующие цифровые скорости битов:

ТАБЛИЦА 1

Уровень II (кбит/с)	Уровень III (кбит/с)
32	32
48	40
56	48
64	56
80	64
96	80
112	96
128	112
160	128
192	160
224	192
256	224
320	256
384	320

Значения, приведенные в таблице 1, указывают общую цифровую скорость на одну звуковую программу независимо от режима. Существует дополнительный формат, свободный формат, который может использоваться для фиксированных цифровых скоростей, не включенных в таблицу, при максимальном значении, равном 384 кбит/с для уровня II и 320 кбит/с для уровня III.

Звуковой цикл уровня II или уровня III ИСО/МЭК 11172-3 всегда соответствует 1152 звуковым ИКМ входным отсчетам на канал. Продолжительность цикла равна 24 мс при частоте дискретизации 48 кГц и 36 мс при частоте дискретизации 32 кГц. Цикл обычно состоит из целого числа байтов. Число байтов может быть рассчитано путем умножения продолжительности цикла на цифровую скорость и деления на 8. Цифровая скорость и частота дискретизации указываются в заголовке в начале кадра звукового сигнала.

Далее представлены три метода, которые могут использоваться для реализации цифровой скорости звукового сигнала, не вошедшей в таблицу:

### **1 Свободный формат**

Состояние свободного формата должно быть указано в заголовке. В этом случае декодер при запуске измеряет расстояние между последовательными словами синхронизации, после чего может применяться процедура инерционной синхронизации, как в случае определенной цифровой скорости. Длина цикла в байтах может быть рассчитана при помощи формулы, приведенной выше. Это значение должно учитываться кодером.

Например, если желательна цифровая скорость, равная 240 кбит/с для стереопрограммы с частотой дискретизации 48 кГц, длина цикла будет равна  $0,024 \times 240\,000/8 = 720$  байтам. Для цифровых скоростей, которые привели бы к нецелому количеству байтов в цикле, требуемая цифровая скорость может быть введена путем использования заполнения. Подробности по этому вопросу приведены в подпункте 2.4.2.3 ИСО/МЭК 11172-3.

Этот метод может использоваться как для уровня II, так и для уровня III.

### **2 Использование поля вспомогательных данных (для уровня II)**

Может выбираться одна из перечисленных цифровых скоростей (выше или равная требуемой цифровой скорости). Путем ограничения распределения битов в кодере может устанавливаться резерв определенного количества битов для вспомогательных данных, так чтобы реальная цифровая скорость звукового сигнала была равна требуемой цифровой скорости. Вспомогательные данные для уровня II расположены в самом конце каждого кадра звукового сигнала.

Биты вспомогательных данных могут быть удалены или переписаны перед передачей. Если эти биты удаляются, соответствующее количество ложных битов должно быть снова вставлено перед декодированием.

Например, если желательна цифровая скорость, равная 240 кбит/с для стереопрограммы с частотой дискретизации 48 кГц, может выбираться цифровая скорость, равная 256 кбит/с, и в соответствии со стандартом ИСО/МЭК указываться в заголовке цикла. Кодер должен быть установлен таким образом, чтобы зарезервировать для вспомогательных данных 48 байтов на один цикл, что соответствует пропускной способности вспомогательных данных в 16 кбит/с.

### **3 Динамическое переключение цифровой скорости (для уровня III)**

На уровне III может использоваться динамическое переключение цифровой скорости для получения цифровых скоростей, не указанных в таблице 1. Если, например, требуется цифровая скорость, равная 120 кбит/с, то кодер уровня III может работать в режиме, где цифровые скорости, подробно определенные в стандарте ИСО/МЭК и равные 128 кбит/с и 112 кбит/с, выбираются в чередующейся последовательности. Кодер, работающий в этом режиме, может быть установлен на число распределения битов, полученное как среднее из этих чередующихся цифровых скоростей. Метод запаса битов на уровне III поддерживает постоянную цифровую скорость, подходящую для процесса кодирования. При помощи этого метода расстояние в битах между словами синхронизации колеблется между двумя различными значениями.

## Дополнение 3

### Описание метода кодирования ИСО/МЭК 13818-7 Общее кодирование движущихся изображений и связанной с ними звуковой информации – Часть 7: Перспективное звуковое кодирование (AAC)

См. ИСО/МЭК 13818-7.

#### 1 Введение

В стандарте ИСО/МЭК 13818-7 описываются звуковые стандарты, не имеющие обратной совместимости, 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 обратной совместимости 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 поток битов, закодированный 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 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 2, система AAC поддерживает 12 типов частот дискретизации в диапазоне от 8 до 96 кГц и до 48 звуковых каналов. В таблице 3 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 2  
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 3  
Конфигурации каналов по умолчанию

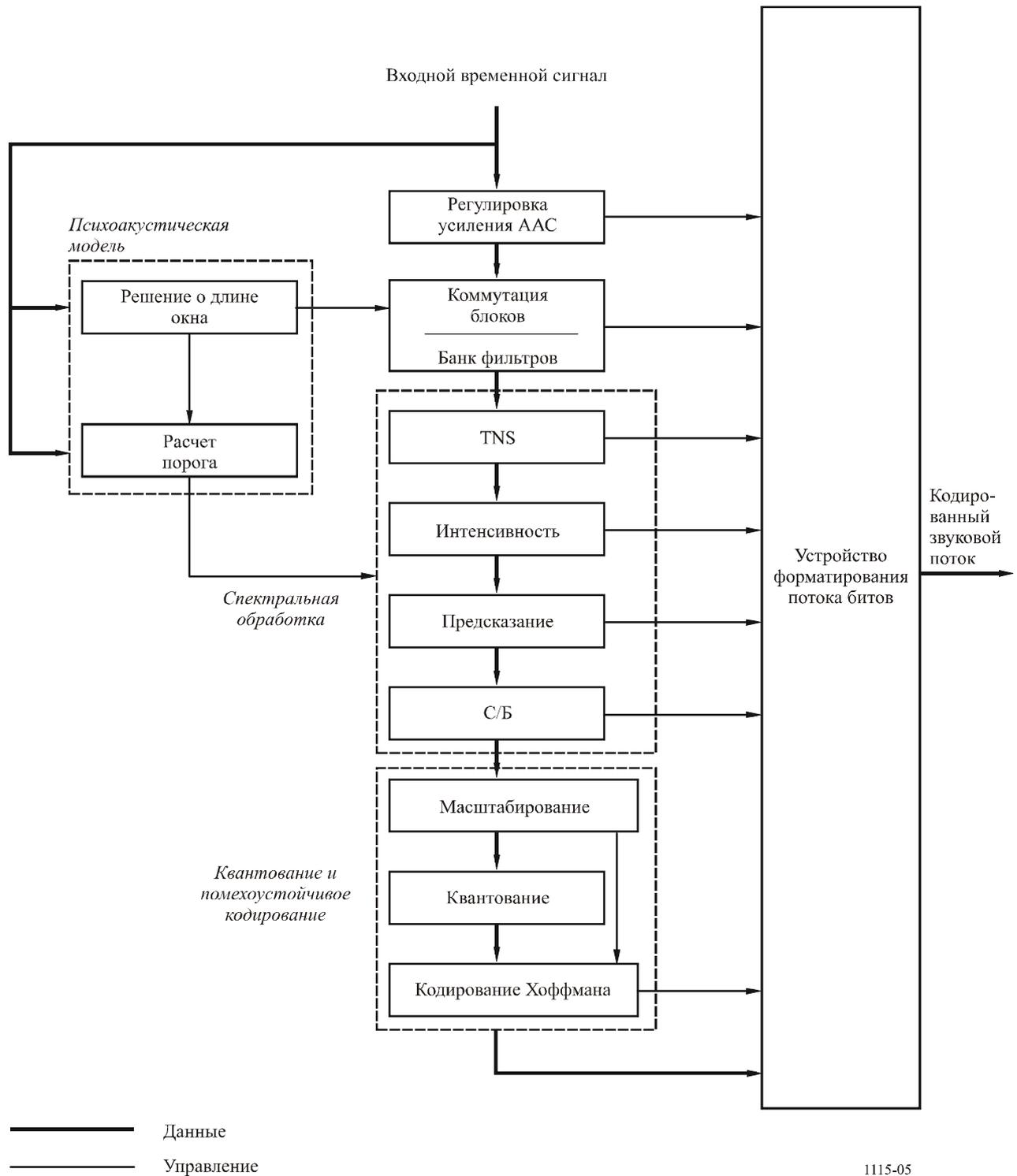
Число громкоговорителей	Звуковые синтаксические элементы, перечисленные в порядке приема	Отображение по умолчанию элемента на громкоговоритель
1	single_channel_element одноканальный элемент	Основной центральный громкоговоритель
2	channel_pair_element элемент пары каналов	Левый и правый основные громкоговорители
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
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 квантатор 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



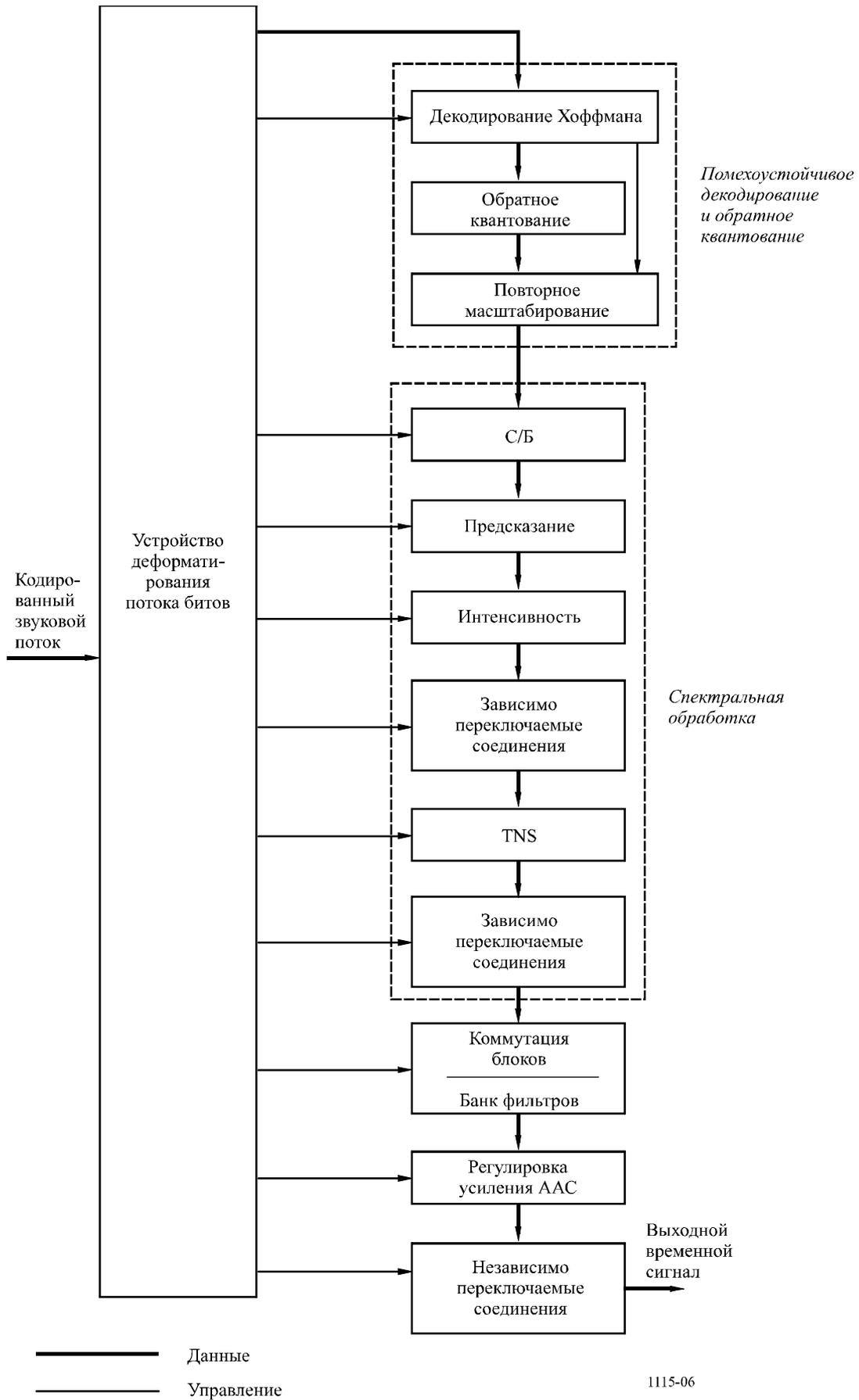
### 3 Декодирование

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

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.

РИСУНОК 6

Функциональная схема декодера AAC MPEG-2



## Дополнение 4

## Описание метода кодирования AC-3

См. Приложение 2 к Рекомендации МСЭ-R BS.1196 и Приложение В к стандарту ATSC A/53В.

## 1 Кодирование

Цифровой алгоритм сжатия 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 AC-3



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

- присоединяется заголовок кадра, содержащий информацию (скорость передачи данных, частота квантования, 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 составляющую и квантованные мантиссы. Результаты работы стандартной программы распределения битов используются для распаковки и деквантования мантисс. Спектральная огибающая кодируется для создания экспонент. Экспоненты и мантиссы обратно преобразуются во временную область для создания декодированных временных выборок ИКМ.

РИСУНОК 8

Декодер АС-3



Реальный декодер АС-3 гораздо 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.

## Дополнение 5

### Перцептивные объективные методы

Перцептивные объективные методы использовались экспериментально для измерения характеристик кодека, и некоторые опубликованные результаты показали хорошую корреляцию с результатами субъективных испытаний.

Далее представлены методы, экспериментально используемые в Секторе радиосвязи:

- отношение шум/маска (NMR);
- перцептивная модель оценки (PERCEVAL);
- перцептивное измерение качества звука (РАQM);
- перцептивная объективная модель (POM 620).

В Рекомендации МСЭ-R BS.1387 определяется метод для объективного измерения перцептивного качества звукового сигнала.

## Дополнение 6

### Методика выбора рекомендуемых звуковых кодеков с низкой цифровой скоростью в испытаниях, проведенных в 1992–1993 годах

Были проведены субъективные испытания для различных приложений одиночного канала и обычных 2-канальных приложений, результаты испытаний сравнивались с требованиями. Проводятся дальнейшие исследования многоканальных систем кодирования.

Для трех приложений – передачи, распределения и доставки – основное качество звукового сигнала кодеков, воспроизводимого после декодирования, должно быть эквивалентно качеству воспроизведения компакт-диска, т. е. субъективно неотличимо для большинства типов материала звуковых программ. Необходимо поддерживать качество картины стереофонического звука.

При излучении наиболее критичный материал для кодеков должен быть таким, чтобы ухудшение могло быть "заметным, но не раздражающим" (оценка 4). (Субъективные испытания были проведены при использовании пятибалльной шкалы ухудшений. Рекомендация МСЭ-R BS.562.) В испытаниях, проведенных в 1992 году, этим требованиям отвечали следующие системы:

независимые каналы со скоростью  $2 \times 128$  кбит/с (Примечание 1):

- ИСО/МЭК 11172-3 уровень II;
- ИСО/МЭК 11172-3 уровень III;
- Dolby AC-2;

комбинированный стереосигнал со скоростью 192 кбит/с:

- ИСО/МЭК 11172-3 уровень II;
- ИСО/МЭК 11172-3 уровень III.

При распределении сигнал после прохождения трех последовательных кодеков со скоростью 120 кбит/с на независимый канал должен оцениваться не более чем на 0,5 ниже, чем первоначальный сигнал источника (Примечание 1). Перезаполнение двоичных кодов было проведено путем снижения коэффициента усиления на 0,1 дБ в линейной области ИКМ. Кодек уровня II ИСО/МЭК 11172-3 был единственной системой, удовлетворяющей данным требованиям, и, следовательно, является основой Рекомендации для этого приложения.

ПРИМЕЧАНИЕ 1. – При этой цифровой скорости эти кодеки оценивались на 4,5 или выше для наиболее критических случаев. Поскольку методики испытания и кодеки продолжают развиваться, испытания в будущем могут дать в результате различные оценки.

При доставке сигнал после прохождения через пять последовательных кодеков со скоростью 180 кбит/с на независимый канал должен быть субъективно неотличим от первоначального сигнала источника. Последовательность прохождения была реализована при помощи разрешения 18-битовой линейной ИКМ. Перезаполнение двоичных кодов было проведено путем снижения коэффициента усиления на 0,1 дБ в линейной области ИКМ. Кодек уровня II ИСО/МЭК 11172-3 был единственной системой, удовлетворяющей данным требованиям, и, следовательно, является основой Рекомендации для этого приложения.

При передаче комментариев основное качество звуковых сигналов речи, воспроизводимых после декодирования, должно быть эквивалентно качеству звука 14-битового оригинала линейной ИКМ. В реальных испытаниях использовался 16-битовый формат линейной ИКМ. Допускаются заметные уровни ухудшения для музыкальных записей качества компакт-дисков. В испытаниях, проведенных в 1992 году только с использованием громкоговорителей, лишь кодек уровня III ИСО/МЭК 11172-3 удовлетворительно работал на речевом сигнале, и он был, как правило, лучшим для музыкальных сигналов. Для речевых сигналов оценки были обычно выше 4,0 (заметно, но не раздражает) для режимов как моно (60 кбит/с), так и стерео (комбинированный стереосигнал со скоростью 120 кбит/с). Этот кодек, следовательно, является основой Рекомендации для данного приложения. Результаты, полученные в ходе последующих испытаний в 1993 году с использованием прослушивания через наушники для оценки характеристик комментаторского кодера моно сигнала уровня III со скоростью 60 кбит/с, показали оценку ниже 4,0, что не удовлетворяло требуемому качеству речевых сигналов. Требуется дальнейшие исследования для скорости 60 кбит/с.

Оценка сложности подтвердила, что декодеры уровня II ИСО/МЭК 11172-3 и Dolby AC-2 характеризуются наименьшей сложностью. Основываясь на следующих соображениях:

- декодер небольшой сложности;
- совпадение с приложениями распределению и доставке;
- гибкость с целью усовершенствования кодера;
- кодек испытывался для обеих цифровых скоростей.

Формат уровня II ИСО/МЭК 11172-3 и декодер были выбраны в качестве основы Рекомендации для приложений в случае передачи со скоростью  $2 \times 128$  кбит/с. В испытаниях, проведенных в 1992 году, уровень II ИСО/МЭК 11172-3 только незначительно удовлетворял требованиям для скорости 192 кбит/с, и было принято решение, что необходимы усовершенствования для этой цифровой скорости, для того чтобы явно удовлетворить основным требованиям по качеству звука (как было сделано для уровня III ИСО/МЭК 11172-3). Последующие испытания, проведенные в 1993 году, не показали каких-либо улучшений на этой стадии. Для пересмотра использования цифровой скорости 192 кбит/с потребуются дальнейшие исследования с использованием режима комбинированной стереофонии. Просьба к администрациям делать вклады по этому вопросу.

Испытания по проверке сети для полного канала радиовещания были проведены в 1993 году, включая пять последовательных кодеков доставки со скоростью 180 кбит/с, три последовательных кодека распределения со скоростью 120 кбит/с и один кодек для целей передачи. Результаты этих испытаний показали, что основное качество звука в крайних точках такого канала не является удовлетворительным.

Также проведены испытания восьми последовательных кодеков на скорости 180 кбит/с, которые показали, что такая конфигурация является удовлетворительной для поддержания приемлемого качества.

---