



ФЕДЕРАЛЬНАЯ СЛУЖБА
ПО ИНТЕЛЛЕКТУАЛЬНОЙ СОБСТВЕННОСТИ

(12) ОПИСАНИЕ ИЗОБРЕТЕНИЯ К ПАТЕНТУ

(52) СПК

G10L 19/167 (2023.02); G10L 19/008 (2023.02); H04S 3/008 (2023.02)

(21)(22) Заявка: 2020130051, 02.07.2019

(24) Дата начала отсчета срока действия патента:
02.07.2019Дата регистрации:
30.08.2023

Приоритет(ы):

(30) Конвенционный приоритет:
02.07.2018 US 62/693,246

(43) Дата публикации заявки: 14.03.2022 Бюл. № 8

(45) Опубликовано: 30.08.2023 Бюл. № 25

(85) Дата начала рассмотрения заявки РСТ на
национальной фазе: 14.09.2020(86) Заявка РСТ:
US 2019/040271 (02.07.2019)(87) Публикация заявки РСТ:
WO 2020/010064 (09.01.2020)

Адрес для переписки:

129090, Москва, ул. Б.Спасская, 25, строение 3,
ООО "Юридическая фирма Городисский и
Партнеры"

(72) Автор(ы):

БРУН, Стефан (SE),
ТОРРЕС, Хуан Феликс (US)

(73) Патентообладатель(и):

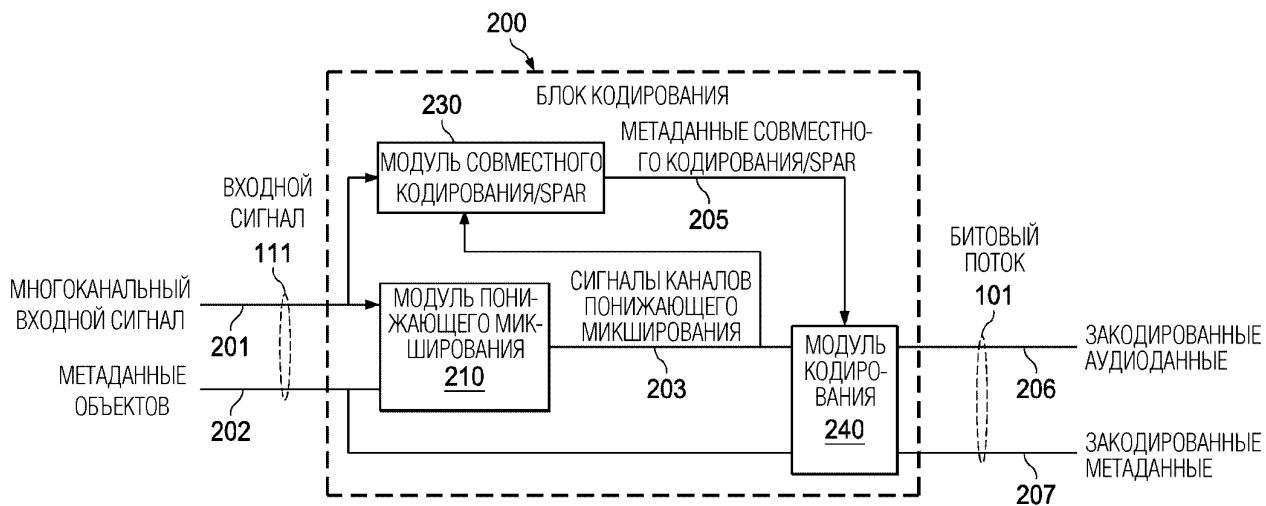
ДОЛБИ ЛЭБОРЕТЕРИЗ ЛАЙСЕНСИНГ
КОРПОРЕЙШН (US),
ДОЛБИ ИНТЕРНЭШНЛ АБ (NL)(56) Список документов, цитированных в отчете
о поиске: "Dolby AC-4: Audio Delivery for
Next-Generation Entertainment Services", с. 1-
30, июнь 2015. "ISO/IEC 23003-2, Information
technology - MPEG Audio Technologies, Part 2:
Spatial Audio Object Coding SAOC", октябрь
2010. US 2015/0348558 A1, 03.12.2015. US 2011/
0238426 A1, 29.09.2011. US 2014/0226823 A1,
14.08.2014. US 2018/0174594 A1, 21.06.2018. WO
(см. прод.)

(54) СПОСОБЫ И УСТРОЙСТВА ДЛЯ ФОРМИРОВАНИЯ ИЛИ ДЕКОДИРОВАНИЯ БИТОВОГО ПОТОКА, СОДЕРЖАЩЕГО ИММЕРСИВНЫЕ АУДИОСИГНАЛЫ

(57) Реферат:

Изобретение относится к средствам для кодирования и декодирования иммерсивных аудиосигналов. Технический результат заключается в повышении эффективности кодирования иммерсивных аудиосигналов. Битовый поток содержит последовательность суперкадров для последовательности кадров иммерсивного аудиосигнала. Повторяющимся образом для последовательности суперкадров:

вставляют закодированные аудиоданные для двух или более кадров одного или более сигналов каналов понижающего микширования, полученных из иммерсивного аудиосигнала, в поля данных суперкадра; и вставляют метаданные для воссоздания двух или более кадров иммерсивного аудиосигнала из закодированных аудиоданных в поле метаданных суперкадра. 4 н. и 9 з.п. ф-лы, 6 ил., 13 табл.



ФИГ. 2

(56) (продолжение):

2017/132082 A1, 03.08.2017. RU 2450440 C1, 10.05.2012.

RU 2802677 C2

RU 2802677 C2



FEDERAL SERVICE
FOR INTELLECTUAL PROPERTY

(51) Int. Cl.
G10L 19/16 (2013.01)
G10L 19/008 (2013.01)
H04S 3/00 (2006.01)

(12) **ABSTRACT OF INVENTION**

(52) CPC
G10L 19/167 (2023.02); *G10L 19/008* (2023.02); *H04S 3/008* (2023.02)

(21)(22) Application: **2020130051, 02.07.2019**

(24) Effective date for property rights:
02.07.2019

Registration date:
30.08.2023

Priority:

(30) Convention priority:
02.07.2018 US 62/693,246

(43) Application published: **14.03.2022 Bull. № 8**

(45) Date of publication: **30.08.2023 Bull. № 25**

(85) Commencement of national phase: **14.09.2020**

(86) PCT application:
US 2019/040271 (02.07.2019)

(87) PCT publication:
WO 2020/010064 (09.01.2020)

Mail address:
**129090, Moskva, ul. B.Spaskaya, 25, stroenie 3,
OOO "Yuridicheskaya firma Gorodisskij i
Partnery"**

(72) Inventor(s):

**BRUHN, Stefan (SE),
TORRES, Juan Felix (US)**

(73) Proprietor(s):

**DOLBY LABORATORIES LICENSING
CORPORATION (US),
DOLBY INTERNATIONAL AB (NL)**

(54) **METHODS AND DEVICES FOR FORMING OR DECODING A BITSTREAM CONTAINING IMMERSIVE AUDIO SIGNALS**

(57) Abstract:

FIELD: coding and decoding of audio signals.

SUBSTANCE: invention relates to means for coding and decoding immersive audio signals. The bitstream contains a sequence of superframes for a sequence of frames of an immersive audio signal. Repeated order for the superframe sequence: insert encoded audio data for two or more frames of one or more downmix channel signals obtained from the immersive audio

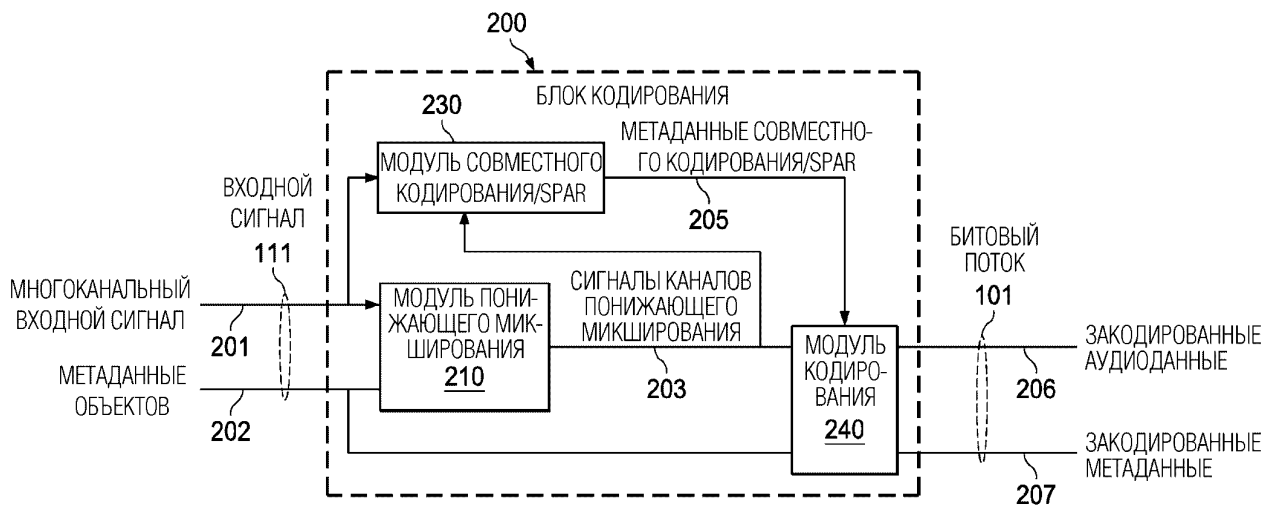
signal into the data fields of the superframe; and inserting metadata for recreating two or more frames of the immersive audio signal from the encoded audio data in the metadata field of the superframe.

EFFECT: increasing the efficiency of coding immersive audio signals.

13 cl, 6 dwg, 13 tbl

RU 2 802 677 C2

RU 2 802 677 C2



ФИГ. 2

RU 2802677 C2

RU 2802677 C2

ПЕРЕКРЕСТНАЯ ССЫЛКА НА РОДСТВЕННЫЕ ЗАЯВКИ

Настоящая заявка испрашивает приоритет согласно предварительной заявке на патент США № 62/693,246, поданной 2 июля 2018 года, содержание которой включено в настоящую заявку посредством ссылки.

ОБЛАСТЬ ТЕХНИКИ

Настоящий документ относится к иммерсивным аудиосигналам, которые могут содержать сигналы представления звукового поля, в частности сигналы амбиофонии (звукозаписи с эффектом присутствия, ambisonics). В частности, настоящий документ относится к формированию и декодированию битового потока, содержащего иммерсивный аудиосигнал.

УРОВЕНЬ ТЕХНИКИ

Звук или звуковое поле в окружающей среде слушателя, который размещен в позиции прослушивания, могут быть описаны с использованием сигнала амбиофонии. Сигнал амбиофонии может рассматриваться как многоканальный аудиосигнал, в котором каждый канал соответствует конкретной диаграмме направленности звукового поля в позиции прослушивания. Сигнал амбиофонии может быть описан с использованием трехмерной (3D) декартовой системы координат, где начало координат системы соответствует позиции прослушивания, ось X направлена вперед, ось Y направлена влево, и ось Z направлена вверх.

Посредством увеличения количества аудиосигналов или каналов и посредством увеличения количества соответствующих диаграмм направленности (и соответствующих функций панорамирования) может быть увеличена точность, с которой описано звуковое поле. В качестве примера сигнал амбиофонии первого порядка содержит 4 канала или формы сигнала, а именно, канал W, указывающий всенаправленный компонент звукового поля, канал X, описывающий звуковое поле с диаграммой направленности дипольного типа, соответствующей оси X, канал Y, описывающий звуковое поле с диаграммой направленности дипольного типа, соответствующей оси Y, и канал Z, описывающий звуковое поле с диаграммой направленности дипольного типа, соответствующей оси Z. Сигнал амбиофонии второго порядка содержит 9 каналов, включающих в себя 4 канала сигнала амбиофонии первого порядка (также называемый В-форматом) плюс 5 дополнительных каналов для других диаграмм направленности. В целом сигнал амбиофонии порядка L содержит $(L+1)^2$ каналов, включающих в себя L^2 каналов сигналов амбиофонии порядка (L-1) плюс $[(L+1)^2 - L^2]$ дополнительных каналов для дополнительных диаграмм направленности (при использовании трехмерного формата амбиофонии). Сигналы амбиофонии порядка L для $L > 1$ могут упоминаться как сигналы амбиофонии высшего порядка (НОА).

Сигнал НОА может использоваться для описания трехмерного звукового поля независимо от размещения динамиков, которое используется для рендеринга сигнала НОА. Иллюстративные размещения динамиков включают в себя наушники, или одно или более размещений громкоговорителей, или окружающую среду рендеринга виртуальной реальности. Следовательно, может оказаться полезным обеспечить сигнал НОА рендереру аудиоинформации, чтобы позволить рендереру аудиоинформации гибко адаптироваться к разным размещениям динамиков.

Сигналы представления звукового поля (SR), такие как сигналы амбиофонии, могут быть дополнены аудиообъектами и/или многоканальными сигналами, чтобы обеспечить сигнал иммерсивной аудиоинформации (IA). Настоящий документ направлен на решение технической проблемы передачи и/или хранения сигналов IA с высоким качеством восприятия эффективным образом с точки зрения использования ширины полосы. В

частности, настоящий документ направлен на решение технической проблемы обеспечения эффективного битового потока, который указывает сигнал IA. Данная техническая проблема решена посредством независимых пунктов формулы изобретения. Предпочтительные примеры описаны в зависимых пунктах формулы изобретения.

5 СУЩНОСТЬ ИЗОБЕТЕНИЯ

В соответствии с аспектом описан способ формирования битового потока, причем битовый поток содержит последовательность суперкадров для последовательности кадров иммерсивного аудиосигнала. Способ содержит, повторяющимся образом для последовательности суперкадров, вставку закодированных аудиоданных для одного или более кадров одного или более сигналов каналов понижающего микширования, полученных из иммерсивного аудиосигнала, в поля данных суперкадра. Кроме того, способ содержит вставку метаданных, в частности закодированных метаданных, для восстановления одного или более кадров иммерсивного аудиосигнала из закодированных аудиоданных в поле метаданных суперкадра.

В соответствии с другим аспектом описан способ для получения данных относительно иммерсивного аудиосигнала из битового потока, причем битовый поток содержит последовательность суперкадров для последовательности кадров иммерсивного аудиосигнала. Способ содержит, повторяющимся образом для последовательности суперкадров, извлечение закодированных аудиоданных для одного или более кадров одного или более сигналов каналов понижающего микширования, полученных из иммерсивного аудиосигнала, из полей данных суперкадра. Кроме того, способ содержит извлечение метаданных для восстановления одного или более кадров иммерсивного аудиосигнала из закодированных аудиоданных из поля метаданных суперкадра.

В соответствии с дополнительным аспектом описана программа. Программа может быть выполнена с возможностью исполнения на процессоре и выполнения этапов способа, изложенных в общих чертах в настоящем документе, при ее исполнении на процессоре.

В соответствии с другим аспектом описан носитель информации. Носитель информации может содержать программу, выполненную с возможностью исполнения на процессоре и выполнения этапов способа, изложенных в общих чертах в настоящем документе, при ее исполнении на процессоре.

В соответствии с дополнительным аспектом описан компьютерный программный продукт. Компьютерная программа может содержать исполняемые инструкции для выполнения этапов способа, изложенных в общих чертах в настоящем документе, при ее исполнении на компьютере.

В соответствии с дополнительным аспектом описан суперкадр битового потока, причем битовый поток содержит последовательность суперкадров для последовательности кадров иммерсивного аудиосигнала. Суперкадр содержит поля данных для закодированных аудиоданных для одного или более (в частности для нескольких) кадров одного или более сигналов каналов понижающего микширования, полученных из иммерсивного аудиосигнала. Кроме того, суперкадр содержит (одно) поле метаданных для метаданных, выполненных с возможностью воссоздавать один или более (в частности несколько) кадров иммерсивного аудиосигнала из закодированных аудиоданных.

В соответствии с другим аспектом описано устройство кодирования, выполненное с возможностью формировать битовый поток. Битовый поток содержит последовательность суперкадров для последовательности кадров иммерсивного аудиосигнала. Устройство кодирования выполнено с возможностью, повторяющимся

образом для последовательности суперкадров, вставлять закодированные аудиоданные для одного или более (в частности нескольких) кадров одного или более сигналов каналов понижающего микширования, полученных из иммерсивного аудиосигнала, в поля данных суперкадра; и вставлять метаданные для восстановления одного или более (в частности нескольких) кадров иммерсивного аудиосигнала из закодированных аудиоданных в поле метаданных суперкадра.

В соответствии с дополнительным аспектом описано устройство для декодирования, выполненное с возможностью получать данные относительно иммерсивного аудиосигнала из битового потока, причем битовый поток содержит последовательность суперкадров для последовательности кадров иммерсивного аудиосигнала. Устройство для декодирования выполнено с возможностью, повторяющимся образом для последовательности суперкадров, извлекать закодированные аудиоданные для одного или более (в частности нескольких) кадров одного или более сигналов каналов понижающего микширования, полученных из иммерсивного аудиосигнала, из полей данных суперкадра; и извлекать метаданные для восстановления одного или более (в частности нескольких) кадров иммерсивного аудиосигнала из закодированных аудиоданных из поля метаданных суперкадра.

Следует отметить, что способы, устройства и системы, в том числе их предпочтительные варианты осуществления, изложенные в общих чертах в настоящей заявке на патент, могут использоваться автономно или в сочетании с другими способами, устройствами и системами, раскрытыми в этом документе. Кроме того, все аспекты способов, устройств и систем, изложенных в общих чертах в настоящей заявке на патент, могут быть произвольным образом объединены. В частности, признаки пунктов формулы изобретения могут сочетаться друг с другом произвольным образом.

КРАТКОЕ ОПИСАНИЕ ЧЕРТЕЖЕЙ

Изобретение разъясняется ниже иллюстративным образом со ссылкой на следующие прилагаемые чертежи.

Фиг. 1 показывает иллюстративную систему кодирования;

Фиг. 2 показывает иллюстративный блок кодирования для кодирования иммерсивного аудиосигнала;

Фиг. 3 показывает другой иллюстративный блок декодирования для декодирования иммерсивного аудиосигнала;

Фиг. 4 показывает иллюстративную структуру суперкадра для иммерсивного аудиосигнала, в частности для закодированных данных, которые указывают иммерсивный аудиосигнал;

Фиг. 5 показывает блок-схему последовательности этапов иллюстративного способа для формирования битового потока, содержащего последовательность суперкадров, указывающих иммерсивный аудиосигнал; и

Фиг. 6 показывает блок-схему последовательности этапов иллюстративного способа для извлечения информации из битового потока, содержащего последовательность суперкадров, указывающих иммерсивный сигнал.

ПОДРОБНОЕ ОПИСАНИЕ

Как изложено в общих чертах выше, настоящий документ относится к эффективному кодированию иммерсивных аудиосигналов, таких как сигналы НОА, многоканальные аудиосигналы и/или аудиосигналы объектов, причем в частности сигналы НОА упоминаются в настоящем документе в более общем случае как сигналы представления звукового поля (SR). Кроме того, настоящий документ относится к хранению или передаче сигнала иммерсивной аудиоинформации (IA) по сети связи в битовом потоке.

Как изложено в общих чертах во вводном разделе, сигнал SR может содержать относительно большое количество каналов или форм сигналов, причем разные каналы относятся к разным функциям панорамирования и/или к разным диаграммам направленности. В качестве примера, трехмерный сигнал амбиофонии первого порядка (FOA) L-го порядка или сигнал НОА содержит $(L+1)^2$ каналов. Сигнал FOA представляет собой сигнал амбиофонии порядка $L=1$, содержащий 4 канала. Сигнал SR может быть представлен во всевозможных форматах.

Звуковое поле может рассматриваться как состоящее из одного или более звуковых событий, происходящих с произвольных направлений вокруг позиции прослушивания. Вследствие этого местоположения одного или более звуковых событий могут быть определены на поверхности сферы (с позицией прослушиванием или опорной позицией в центре сферы).

Формат звукового поля, такой как FOA или НОА, определен таким образом, чтобы позволить выполнять рендеринг звукового поля с произвольными размещениями динамиков (т.е. произвольными системами рендеринга). Однако системы рендеринга (такие как система Dolby Atmos) обычно ограничены в том смысле, что возможные вертикальные возвышения динамиков фиксированы определенным количеством плоскостей (например, плоскость высоты уха (горизонтальная), плоскость потолка или верхняя плоскость, и/или плоскость пола или нижняя плоскость). Следовательно, понятие идеального сферического звукового поля может быть модифицировано в звуковое поле, состоящее из аудиообъектов, которые расположены в разных кольцах на разных высотах на поверхности сферы (подобно сложенным в стопку кольцам, которые составляют пчелиный улей).

Как показано на фиг. 1, система 100 аудиокодирования содержит блок 110 кодирования и блок 120 декодирования. Блок 110 кодирования может быть выполнен с возможностью формировать битовый поток 101 для передачи блоку 120 декодирования на основе входного сигнала 111, причем входной сигнал 111 может содержать или может являться иммерсивным аудиосигналом (используемым, например, для приложений виртуальной реальности (VR)). Иммерсивный аудиосигнал 111 может содержать сигнал SR, многоканальный сигнал и/или множество объектов (каждый объект содержит сигнал объекта и метаданные объекта). Блок 120 декодирования может быть выполнен с возможностью обеспечивать выходной сигнал 121 на основе битового потока 101, причем выходной сигнал 121 может содержать или может являться воссозданным иммерсивным аудиосигналом.

Фиг. 2 иллюстрирует приведенный в качестве примера блок 110, 200 кодирования. Блок 200 кодирования может быть выполнен с возможностью кодировать входной сигнал 111, причем входной сигнал 111 может являться сигналом 111 иммерсивной аудиоинформации (IA). Сигнал 111 IA может содержать многоканальный входной сигнал 201. Многоканальный входной сигнал 201 может содержать сигнал SR и один или более сигналов объекта. Кроме того, метаданные 202 объекта для множества сигналов объекта могут быть обеспечены как часть сигнала 111 IA. Сигнал 111 IA может быть обеспечен посредством механизма создания контента, причем механизм создания контента может быть выполнен с возможностью получать объекты и/или сигнал SR из (сложного) контента IA, такого как контент виртуальной реальности, который может содержать сигнал SR, один или более многоканальных сигналов и/или один или более объектов.

Блок 200 кодирования содержит модуль 210 понижающего микширования, выполненный с возможностью выполнять понижающее микширование многоканального

входного сигнала 201 до множества сигналов 203 каналов понижающего микширования. Множество сигналов 203 каналов понижающего микширования может соответствовать сигналу SR, в частности сигналу амбиофонии первого порядка (FOA). Понижающее микширование может быть выполнено в области частотных подполос или области QMF (например, с использованием десяти или более частотных подполос).

Блок 200 кодирования дополнительно содержит модуль 230 совместного кодирования (в частности модуль SPAR), который выполнен с возможностью определять метаданные 205 совместного кодирования (в частности метаданные SPAR, воссоздания пространственного звучания), которые выполнены с возможностью воссоздавать многоканальный входной сигнал 201 из множества сигналов 203 каналов понижающего микширования. Модуль 230 совместного кодирования может быть выполнен с возможностью определять метаданные 205 совместного кодирования в области частотных подполос. В одном примере инструмент воссоздания пространственного звучания (SPAR) представляет собой инструмент кодирования для улучшенного кодирования относительно большого количества аудиоканалов и объектов. Чтобы достигнуть эффективности кодирования, этот инструмент поддерживает воссоздание аудиоканалов и объектов из меньшего количества совместных входных аудиоканалов и меньшего количества вспомогательной информации.

Для определения метаданных 205 совместного кодирования или SPAR множество сигналов 203 каналов понижающего микширования может быть преобразовано в область частотных подполос и/или может быть обработано в области частотных подполосы. Кроме того, многоканальный входной сигнал 201 может быть преобразован в область частотных подполос. Затем метаданные 205 совместного кодирования или SPAR могут быть определены для каждой частотной подполосы, в частности таким образом, чтобы посредством повышающего микширования сигнала частотной подполосы множества сигналов 203 каналов понижающего микширования с использованием метаданных 205 совместного кодирования или SPAR получалось приближение сигнала частотной подполосы многоканального входного сигнала 201. Метаданные 205 совместного кодирования или SPAR для разных частотных подполос могут быть вставлены в битовый поток 101 для передачи соответствующему блоку декодирования.

Кроме того, блок 200 кодирования может содержать модуль 240 кодирования, который выполнен с возможностью выполнять кодирование формы сигнала из множества сигналов 203 каналов понижающего микширования, тем самым обеспечивая закодированные аудиоданные 206. Каждый из сигналов 203 каналов понижающего микширования может быть закодирован с использованием моно-кодера формы сигнала (например, кодирование 3GPP EVS), тем самым обеспечивая эффективное кодирование. Дополнительными примерами для кодирования множества сигналов 203 каналов понижающего микширования являются MPEG AAC, MPEG HE-AAC и другие аудиокодеки MPEG, кодеки 3GPP, Dolby Digital/Dolby Digital Plus (AC-3, eAC-3), Opus, LC-3 и другие сходные кодеки. В качестве дополнительного примера, инструменты кодирования, содержащиеся в кодеке AC-4, могут быть выполнены с возможностью выполнять операции блока 200 кодирования.

Кроме того, модуль 240 кодирования может быть выполнен с возможностью выполнять энтропийное кодирование метаданных 205 совместного кодирования (т.е. метаданных SPAR) и метаданных 202 объекта, тем самым обеспечивая закодированные метаданные 207. Закодированные аудиоданные 206 и закодированные метаданные 207 могут быть вставлены в битовый поток 101. Битовый поток 101 может представлять

структуру суперкадров, которая описана в настоящем документе. Способ 500, который описан в настоящем документе, может быть выполнен модулем 240 кодирования.

Фиг. 3 показывает иллюстративный блок 120, 350 декодирования. Блок 120, 350 декодирования может включать в себя приемник, который принимает битовый поток 101, который может включать в себя закодированные аудиоданные 206 и закодированные метаданные 207. Блок 120, 350 декодирования может включать в себя процессор и/или демультимплексор, который демультимплексирует закодированные аудиоданные 206 и закодированные метаданные 207 из битового потока 101. Блок 350 декодирования содержит модуль 360 декодирования, который выполнен с возможностью получать множество воссозданных сигналов 314 каналов из закодированных аудиоданных 206. Модуль 360 декодирования также может быть выполнен с возможностью получать метаданные 205 совместного кодирования или SPAR и/или метаданные 202 объектов из закодированных метадаанных 207. Способ 600, который описан в настоящем документе, может быть выполнен модулем 360 декодирования.

Кроме того, блок 350 декодирования одержит модуль 370 воссоздания, который выполнен с возможностью получать воссозданный многоканальный сигнал 311 из метадаанных 205 совместного кодирования или SPAR и из множества воссозданных сигналов 314 каналов. Метаданные 205 совместного кодирования или SPAR могут переносить изменяющиеся по времени и/или частоте элементы матрицы повышающего микширования, которая позволяет воссоздавать многоканальный сигнал 311 из множества воссозданных сигналов 314 каналов. Процесс повышающего микширования может быть выполнен в области частотных подполос QMF (квадратурный зеркальный фильтр). В качестве альтернативы, другое частотно-временное преобразование, в частности преобразование на основе FFT (быстрое преобразование Фурье) может использоваться, чтобы выполнить процесс повышающего микширования. В общем случае может быть применено преобразование, которое обеспечивает выборочный по частоте анализ и обработку (повышающего микширования). Процесс повышающего микширования также может включать в себя декорреляторы, которые обеспечивают улучшенное воссоздание ковариации воссозданного многоканального сигнала 311, причем декорреляторами могут управлять дополнительные метаданные 205 совместного кодирования или SPAR.

Воссозданный многоканальный сигнал 311 может содержать воссозданный сигнал SR и один или более воссозданных сигналов объектов. Воссозданный многоканальный сигнал 311 и метаданные объектов могут сформировать выходной сигнал 121 (также известный как воссозданный сигнал 121 IA). Воссозданный сигнал 121 IA может использоваться для рендеринга 331 динамиков, для рендеринга 332 наушников и/или, например, для рендеринга 333 контента виртуальной реальности с использованием звукового поля.

Таким образом, описан блок 110, 200 кодирования, который выполнен с возможностью кодировать входной сигнал IA 111 в

- микшированный с понижением сигнал, содержащий множество сигналов 203 каналов понижающего микширования, причем микшированный с понижением сигнал 203 может представлять собой сигнал представления звукового поля (SR); и

- метаданные 202, 205, включающие в себя метаданные 205 SPAR или совместного кодирования и/или метаданные 202 объектов для одного или более объектов.

Метаданные 202, 205, в частности метаданные 205 SPAR, могут представлять другое временное разрешение по сравнению с микшированным с понижением сигналом. В частности, метаданные 202, 205 могут использоваться для множества кадров (например,

для двух кадров) микшированного с понижением сигнала. Ввиду этого суперкадр может быть определен для битового потока 101, причем суперкадр содержит множество кадров микшированного с понижением сигнала плюс метаданные 202, 205 для множества кадров микшированного с понижением сигнала SR.

5 Фиг. 4 показывает иллюстративный суперкадр 400. Суперкадр 400 может содержать поле 401 основного заголовка (ВН) и/или поле 402 информации конфигурации (СІ), которое может содержать данные, которые достоверны для всего суперкадра 400. Кроме того, суперкадр 400 содержит поля 411, 412, 421, 422 данных сигнала для закодированных аудиоданных 206 для одного или более (в частности для множества) 10 кадров микшированного с понижением сигнала. В частности, для каждого сигнала 203 канала понижающего микширования может быть обеспечено одно или более (в частности множество) полей 411, 412, 421, 422 данных сигнала, например, поля 411, 421 данных сигнала для двух кадров первого сигнала 203 канала понижающего микширования и поля 412, 422 данных сигнала для двух кадров N-го сигнала 203 канала 15 понижающего микширования. Поля 411, 412, 421, 422 данных сигнала также упоминаются в настоящем документе как битовые поля EVS (для примера, в котором кодер EVS используется для кодирования сигналов 203 каналов понижающего микширования).

Кроме того, суперкадр 400 содержит поле 403 метаданных (MDF). Поле 403 20 метаданных может быть выполнен с возможностью обеспечивать метаданные 205 SPAR или совместного кодирования и/или прогнозирующие коэффициенты (PC). Следовательно, поле метаданных 403 может являться битовым полем SPAR или битовым полем PC (в зависимости от режима кодирования, который используется). Кроме того, суперкадр 400 может содержать поле 404 расширителя кадра (FE).

25 Следовательно, суперкадр 400 может содержать сигнальные элементы, выполненные с возможностью

- указывать один или более режимов кодека (EVS), которые использовались для кодирования (EVS) N сигналов каналов понижающего микширования; по умолчанию может быть N=4 каналов, это означает, что существует 4 сигнала каналов понижающего 30 микширования кодека (EVS): W, X', Y', Z';

- указывать выбранный режим работы поддерживаемого метаданными кодека (EVS);
- указывать битовую скорость метаданных;
- обеспечивать возможность сигнализировать о потенциальных будущих расширениях.

Один или более сигнальных элементов (таких как поле 402 CI) могут быть обеспечены 35 только по условию внутри полосы в суперкадре 400. Если обеспечен опциональный или условный сигнальный элемент, этот сигнальный элемент может быть динамически адаптирован и/или включен в суперкадр 400. Один или более сигнальных элементов могут поддерживаться статическими и/или могут быть обеспечены только один раз, например, как внеполосное сообщение. Один или более сигнальных элементов могут 40 быть полудинамическими, в этом случае один или более сигнальных элементов обеспечиваются только внутри полосы в выбранных суперкадрах 400.

Суперкадр 400 может быть выполнен с возможностью обеспечивать один или более из следующих признаков:

- Полное декодирование и рендеринг суперкадров, закодированных поддерживаемым 45 метаданными кодеком EVS;
- Частичное моно-декодирование суперкадров, закодированных поддерживаемым метаданными кодеком EVS;
- Извлечение с низкой сложностью информации о размере суперкадра из

последовательности последовательно соединенных суперкадров без необходимости декодировать суперкадры, например, для размещения информации о размере суперкадра во вторичный формат (такой как ISOBMFF, ISO базовый формат медиа-файла), который

- 5 - Определение битовой скорости с низкой сложностью без необходимости декодировать данные суперкадра;
- Продвижение вперед с низкой сложностью и пропуск суперкадров без необходимости декодировать данные суперкадра;
- Продвижение назад с низкой сложностью без необходимости декодировать данные суперкадра (в частности в случае операции с постоянной битовой скоростью);
- 10 - Простая повторная синхронизация и пропуск суперкадра в случае битовых ошибок в участках битового потока с EVS и/или метаданными, закодированными посредством арифметического или статистического кодирования;
- Редактируемые суперкадры, которые позволяют заменять метаданные или кадры
- 15 данных EVS.

Закодированный битовый суперкадр 400 поддерживаемого метаданными кодека EVS может соответствовать шагу кодирования 40 мс (например, содержащему два кадра по 20 мс). Он может состоять из следующих элементарных битовых полей:

- 20 - Поле 401 основного заголовка (BH): Это поле может нести индикатор наличия поля конфигурации (CPI), индикатор корректировки размера поля метаданных (MDA) и индикатор расширения (EI). Индикатор CPI может указывать, обеспечено ли поле информации конфигурации (CI) в текущем суперкадре 400. Индикатор MDA может сигнализировать о разности между сигнализируемым максимальным размером кадра метаданных и фактическим размером кадра метаданных. Индикатор EI может
- 25 сигнализировать о том, расширен ли суперкадр 400 посредством расширителя кадров 404 (FE);
- Поле 402 информации конфигурации (CI): Это поле может нести сигнальную информацию, относящуюся к конфигурациям EVS, SPAR и инструментов кодирования прогнозирующих коэффициентов, которые использовались, например, тип кадра (режим
- 30 кодирования), кодовая скорость и другие параметры конфигурации, которые описаны в настоящем документе;
- Битовые поля 411, 421, 412, 422 EVS: Каждое поле может нести биты одного кадра EVS (в частности без заголовка полезной нагрузки EVS), как определено в документе 3GPP TS 26.445: "Codec for Enhanced Voice Services (EVS); Detailed algorithmic description",
- 35 раздел 7, который включен в настоящий документ по ссылке;
- Битовое поле 403 SPAR (SPAR) 403: Это поле может нести части одного кадра метаданных SPAR, возможно заполненного нулями в конце для выравнивания с байтом;
- Битовое поле 403 прогнозирующего коэффициента (PC): Это поле может нести биты одного кадра метаданных прогнозирующего коэффициента, возможно
- 40 заполненного нулями в конце для выравнивания с байтом;
- Расширитель 404 кадра (FE): Это поле может быть определено для будущего использования и может нести дополнительные данные. За исключением элемента размера, содержащегося в FE, любые другие данные, которые несет FE, могут быть зарезервированы для будущего использования (RFU).
- 45 Все элементарные битовые поля могут быть выровнены с байтом и, если необходимо, заполнены нулями в конце до их заданного размера.

Вышеупомянутые элементарные поля могут быть включены в (один) суперкадр 400 в следующей последовательности. Суперкадр может содержать:

- один основной заголовок 401 (BH) 401 содержащий:
 - индикатор наличия поля конфигурации (CPI),
 - индикатор корректировки размера поля метаданных (MDA), и
 - индикатор расширения (EI);
- одно опциональное поле 402 информации конфигурации (CI); о наличии поля 402 CI может оповестить индикатор CPI,
 - данные N сигналов S_1, \dots, S_N каналов понижающего микширования, закодированных кодеком EVS, два последовательных кадра для каждого сигнала канала понижающего микширования, которые могут быть перенесены посредством $2 \cdot N$ элементарных битовых полей 411, 421, 412, 422 EVS (далее EVS(.)); в операции по умолчанию с 4 сигналами каналов понижающего микширования имеется 8 последовательных битовых полей 411, 421, 412, 422 EVS, представляющих два кадра сигналов W, X', Y', Z' каналов понижающего микширования,
 - одно поле 403 кадра метаданных (MDF) для SPAR или прогнозирующих коэффициентов, таким образом, оно представляет собой:
 - одно элементарное поле бита SPAR или
 - одно элементарное поле бита PC,
 - один опциональный расширитель 404 кадра (FE); наличие поля FE может быть указано индикатором EI.

Таблица 1 показывает иллюстративную структуру суперкадра 400.

Биты (старший-младший)	Имя	Описание
8	BH	Основной заголовок, содержащий CPI, MDA, EI
Переменное кол-во	CI	Поле информации конфигурации (опциональное, в зависимости от CPI)
Переменное кол-во	EVS ($S_{1,1}$)	Данные кадра EVS для первого кадра первого канала понижающего микширования
Переменное кол-во	EVS ($S_{2,1}$)	Данные кадра EVS для первого кадра второго канала понижающего микширования
Переменное кол-во	EVS ($S_{\dots,1}$)	...
Переменное кол-во	EVS ($S_{N,1}$)	Данные кадра EVS для первого кадра N-го канала понижающего микширования
Переменное кол-во	EVS ($S_{1,2}$)	Данные кадра EVS для второго кадра первого канала понижающего микширования
Переменное кол-во	EVS ($S_{2,2}$)	Данные кадра EVS для второго кадра второго канала понижающего микширования
Переменное кол-во	EVS ($S_{\dots,2}$)	...
Переменное кол-во	EVS ($S_{N,2}$)	Данные кадра EVS для второго кадра N-го канала понижающего микширования
Переменное кол-во	MDF	Кадр метаданных, содержащий либо поле SPAR, либо поле PC
Переменное кол-во	FE	Расширитель кадра (опциональный, в зависимости от EI)

Таблица 1

В ситуации по умолчанию имеются четыре закодированных кодеком EVS сигнала каналов понижающего микширования. Структура суперкадра для ситуации по умолчанию показана в таблице 2.

Биты (старший-младший)	Имя	Описание
8	BH	Основной заголовок, содержащий CPI, MDA, EI
72	CI	Поле информации конфигурации (опциональное, в зависимости от CPI)
Переменное кол-во	EVS (W(1))	Данные кадра EVS для первого кадра сигнала W
Переменное кол-во	EVS (X'(1))	Данные кадра EVS для первого кадра сигнала X'
Переменное кол-во	EVS (Y'(1))	Данные кадра EVS для первого кадра сигнала Y'
Переменное кол-во	EVS (Z'(1))	Данные кадра EVS для первого кадра сигнала Z'
Переменное кол-во	EVS(W(2))	Данные кадра EVS для второго кадра сигнала W
Переменное кол-во	EVS (X'(2))	Данные кадра EVS для второго кадра сигнала X'

Переменное кол-во	EVS (Y'(2))	Данные кадра EVS для второго кадра сигнала Y'
Переменное кол-во	EVS (Z'(2))	Данные кадра EVS для второго кадра сигнала Z'
Переменное кол-во	MDF	Кадр метаданных, содержащий либо поле SPAR, либо поле PC
Переменное кол-во	FE	Расширитель кадра (опциональный, в зависимости от EI)

5 Таблица 2

Более подробная информация относительно различных элементарных битовых полей предоставлена ниже.

Поле 401 основного заголовка (ВН) может нести индикатор наличия поля конфигурации (CPI), индикатор корректировки размера поля метаданных (MDA) и индикатор расширения (EI). Это байтовое поле всегда может быть первым элементом в суперкадре 400.

Структура поля 401 ВН показана в таблице 3.

Биты (старший-младший)	Имя	Описание
1	CPI	Индикатор наличия поля конфигурации
6	MDA	Корректировка размера поля метаданных
1	EI	Индикатор расширения

15 Таблица 3

Индикатор наличия поля конфигурации (CPI) может являться одним битом, используемым для сигнализации о наличии поля информации конфигурации (CI) в текущем суперкадре 400. Индикатор CPI может означать следующее.

-- CPI = '0': Это указывает, что поле информации конфигурации не обеспечено в текущем суперкадре 400. Следует отметить, что информация конфигурации вместо этого может быть обеспечена как статическая внеполосная информация или из самого последнего принятого суперкадра 400, содержащего поле 402 информации конфигурации.

-- CPI = '1': Это указывает, что поле информации конфигурации обеспечено в текущем суперкадре 400. Информация конфигурации, которая обеспечена в поле 402 CI, является достоверной для этого суперкадра 400 и для любого будущего суперкадра 400 до тех пор, пока не будет обеспечен следующий суперкадр 400, содержащий поле 402 информации конфигурации.

Индикатор корректировки размера поля метаданных (MDA) может быть обеспечен непосредственно после бита CPI. Этот 6-битный индикатор может сигнализировать о разности между длиной поля 403 MDF, сообщенной элементом MDR (который определен ниже), и фактическим размером поля 403 MDF. С использованием индикатора MDA в качестве индекса указанная разность может быть получена посредством таблицы поиска, показанной в таблице 4. Последовательность значений корректировки в таблице 4 определена в стиле Matlab: start-value:step-size:end-value (начальное-значение:размер-шага:конечное-значение). Непостоянные размеры шага параметра корректировки, показанные в таблице 4, могут быть предусмотрены в соответствии с моделью аппроксимации распределения полной длины статистического кода метаданных. Это позволяет минимизировать количество неиспользованных битов в поле 403 MDF и тем самым накладных расходов передачи.

MDA	0...47	48...55	56...59	60...61	62	63
Значение корректировки	0:1:47	49:2:63	67:4:79	87:8:95	111	143

45 Таблица 4

В зависимости от максимального размера поля MDF значение корректировки представляет однобайтные или двухбайтные блоки. Для максимальных размеров поля MDF до 275 байтов значение корректировки представляет однобайтные блоки, в ином

случае двухбайтные блоки.

За индикатором MDA может следовать один бит индикатора расширения (EI). Если этот бит установлен равным 1, к текущему суперкадру 400 прилагается элемент расширителя кадра (FE).

5 Опционально поле 402 информации конфигурации (CI) может нести сигнальные элементы, проиллюстрированные в таблице 5. Поле 402 CI может состоять или содержать 8 байтов данных (для случая двух кадров EVS на каждый сигнал канала понижающего микширования и N=4 каналов понижающего микширования).

Биты (старший-младший)	Имя	Описание
3	N-1	Индикатор количества каналов с понижающим микшированием кодека EVS
1	MDT	Индикатор типа метаданных
11	MDC	Конфигурация кодирования метаданных
5	MDR	Сигнализация битовой скорости метаданных
15 3	BND	Количество частотных полос кодирования метаданных
1	RES	Резерв для будущего использования
6	FT-1,1	EVS FT для первого кадра первого канала понижающего микширования
6	FT-2,1	EVS FT для первого кадра второго канала понижающего микширования
6
20 6	FT-N,1	EVS FT для первого кадра N-го канала понижающего микширования
6	FT-1,2	EVS FT для второго кадра первого канала понижающего микширования
6	FT-2,2	EVS FT для второго кадра второго канала понижающего микширования
6
25 6	FT-N,2	EVS FT для второго кадра N-го канала понижающего микширования
Переменное кол-во	zero-pad	Заполнение нулями для выравнивания с байтом

Таблица 5

30 Таблица 6 иллюстрирует опциональное поле 402 информации конфигурации для случая по умолчанию с четырьмя закодированными кодеком EVS сигналами каналов понижающего микширования. В этом случае поле CI состоит из 9 байтов данных.

Биты (старший-младший)	Имя	Описание
3	N-1	'011' указывает 4 канала понижающего микширования кодека EVS
1	MDT	Индикатор типа метаданных
11	MDC	Конфигурация кодирования метаданных
5	MDR	Сигнализация битовой скорости метаданных
3	BND	Количество частотных полос кодирования метаданных
1	RES	Резерв для будущего использования
40 6	FT-1,1	EVS FT для кодирования W(1)
6	FT-2,1	EVS FT для кодирования X'(1)
6	FT-3,1	EVS FT для кодирования Y'(1)
6	FT-4,1	EVS FT для кодирования Z'(1)
6	FT-1,2	EVS FT для кодирования W(2)
6	FT-2,2	EVS FT для кодирования X'(2)
45 6	FT-3,2	EVS FT для кодирования Y'(2)
6	FT-4,2	EVS FT для кодирования Z'(2)

Таблица 6

Индикатор для количества N закодированных кодеком EVS сигналов каналов

понижающего микширования EVS (N-I) может являться 3-битным элементом, который кодирует количество N закодированных кодеком EVS сигналов каналов понижающего микширования. N получается из индикатора N-I посредством увеличения числа, представленного 3-битным элементом, на 1. Для достижения операции по умолчанию с 4 сигналами каналов понижающего микширования EVS элемент N-I может быть установленным равным 3 ('011').

Бит индикатора типа метаданных (MDT) может означать следующее:

-- MDT = '0': указывает, что поле MDF несет битовое поле PC.

-- MDT = '1': указывает, что поле MDF несет битовое поле SPAR.

Поле конфигурации кодирования метаданных (MDC), может содержать либо информацию конфигурации используемого инструмента прогнозирующего коэффициента, либо инструмента кодирования SPAR? в зависимости от индикатора бита MDT. Поле MDC может являться 11-битным элементом поля 402 CI. Назначение его битов может зависеть от бита MDT поля 402 CI. В зависимости от значения бита MDT биты MDC могут означать следующее.

-- MDT = '0': Если бит MDT равен нулю, 3 старших значащих бита (MSB) поля MDC кодируют параметр конфигурации схемы кодирования прогнозирующего коэффициента. Остальные 8 битов поля MDC не используются и заполняются нулями. Структура и содержание поля MDC в этом случае показаны в таблице 7a.

-- MDT = '1': Если бит MDT равен единице, 11 битов поля MDC кодируют конфигурацию кодека SPAR, как проиллюстрировано в таблице 7b. Порядок НОА может быть вычислен посредством увеличения `hoa_order_idx` на 1.

Биты (старший-младший)	Имя	Описание	Диапазон значений
3	<code>coding_strategy_idx</code>	Индекс в таблице Хаффмана	0-7
8	Заполнение нулями		

Таблица 7a

Биты (старший-младший)	Имя	Описание	Диапазон значений
4	<code>n_obj</code>	Количество аудиообъектов	1-15
2	<code>hoa_order_idx</code>	Идентификатор порядка НОА	0-2
2	<code>n_deco</code>	Количество декорреляторов	0-3
3	<code>coding_strategy_idx</code>	Индекс в таблице Хаффмана	0-7

Таблица 7b

Поле сигнализации битовой скорости метаданных (MDR) может содержать 5 битов и может использоваться, чтобы закодировать максимальный размер поля MDF. Максимальный размер поля MDF может быть получен посредством поиска в таблице с использованием таблицы 8, причем значение поля MDR является индексом таблицы 8. Кроме того, таблица 8 указывает (максимальную) битовую скорость метаданных в Кбит/с. В таблице 8 фактический размер поля MDF сообщается как максимальный размер поля MDF минус количество/значение корректировки, указанное посредством индикатора MDA (из поля 401 BH). Это позволяет сигнализировать о фактическом размере поля MDF с высоким разрешением (обычно с разрешением байта). Также следует отметить, что любой неиспользованный бит в поле MDF может быть заполнен нулями, что может произойти в случае, если фактический размер поля MDF обеспечивает больше места, чем необходимо для закодированных метаданных.

Значение MDR	Битовая скорость метаданных (Кбит/с)	Максимальный размер MDF (байты)
0	4	20
1	5	25
2	6	30
3	7	35
4	8	40
5	10	50
6	12	60
7	15	75
8	18	90
9	21	105
10	24	120
11	28	140
12	32	160
13	36	180
14	40	200
15	42	210
16	45	225
17	47	235
18	50	250
19	55	275
20	60	300
21	65	325
22	70	350
23	75	375
24	80	400
25	85	425
26	90	450
27	95	475
28	100	500
29	105	525
30	110	550
31	128	640

30 Таблица 8

Поле количества частотных полос (BND) может являться 3-битным числом и может указывать количество частотных подполос, используемых при кодировании метаданных. Количество частотных полос получается из значения поля BND посредством поиска в таблице 9. В операции по умолчанию поле BND может быть установлено равным 5 ('101'), что указывает 12 частотных подполос.

BND (3 бита)	Количество частотных полос
0	1
1	3
2	5
3	7
4	9
5	12
6	15
7	23

45 Таблица 9

Использование резервного бита (RES) может быть зарезервировано для будущего использования. В операциях по умолчанию этот бит может быть установлен равным '0' и может быть проигнорирован приемником.

Поле EVS FT (FT-х, у) может представлять тип кадра EVS (FT), который применяется для кодирования у-го кадра х-го сигнал канала понижающего микширования, где х= 1, ..., N и у=1, 2. Тип кадра EVS может быть определен в документе 3GPP TS 26.445, раздел A2.2.1.2, который включен в настоящий документ по ссылке. Следует отметить, что за последним поле EVS FT в поле 402 CI может следовать до 7 заполняющих нулевых битов, что гарантирует выравнивание октета. В случае, если конец последнего поля EVS FT выровнен по октету, заполняющие нулевые биты не применяются. Заполняющие нулевые биты должны быть проигнорированы приемником.

Элементарные битовые поля 411, 421, 412, 422 EVS могут быть заданы, как определено в документе 3GPP TS 26.445, раздел 7 (который включен в настоящий документ по ссылке), для соответствующим образом используемого режима кодирования EVS. Как определено по процитированной ссылке, никакие дополнительные сигнальные биты не задаются как часть элементарного поля кадра EVS, чтобы указать битовую скорость или режим работы EVS. Эта информация может являться частью опционального поля 402 CI текущего или предыдущего суперкадра 400, или также может быть обеспечена вне полосы.

Подробные сведения о распределении коэффициентов для метаданных SPAR показаны в таблице 10. Таблица 10 показывает порядок битов, в котором они вставлены в кадре. Следует отметить, что старший значащий бит (MSB) каждого параметра всегда вставляется сначала. Поскольку каждое поле динамически квантуется, распределение битов является переменным.

Имя	Описание	Кодирование
obj_pos	Матрица позиций объектов	Зависимое от позиции квантование
Mspar	Матрица воссоздания HOA SPAR	Код Хаффмана
Pspar	Матрица SPAR	Код Хаффмана

Таблица 10

Подробные сведения о распределении коэффициентов для метаданных PC показаны в таблице 11. Таблица 11 показывает порядок битов, в котором они вставляются в суперкадре 400. Следует отметить, что старший значащий бит (MSB) каждого параметра всегда вставляется сначала. Поскольку каждое поле динамически квантуется, распределение битов является переменным.

Имя	Описание	Кодирование
Gfoa	Матрица прогнозирующих коэффициентов	Код Хаффмана

Таблица 11

Элемент 404 расширителя кадра (FE) обычно несет в своих первых двух байтах 16-битное целое число без знака, которое указывает размер поля 404 FE в байтах. Этот элемент упоминается как FE-size. Таким образом, значение FE-size больше или равно 2. Содержание и значение остальной части FE-data поля 404 FE могут быть зарезервированы для будущего использования. В операции по умолчанию элемент FE-size может быть проанализирован, и элемент FE-data может быть пропущен и проигнорирован. Структура и содержание поля 404 FE показаны в таблице 12.

Биты (старший-младший)	Имя	Описание
16	FE-size	Размер поля FE
Переменное кол-во	FE-data	Биты данных в поле FE

Таблица 12

Таким образом, описана структура суперкадра, который позволяет сигнализировать информацию конфигурации поддерживаемого метаданными кодека EVS. Структура суперкадра позволяет приемнику декодировать данные поддерживаемого метаданными

5

кодека EVS. На общем уровне поддерживаемый метаданными кодек EVS является многорежимной и/или мультискоростной кодирующей системой. Базовый кодек EVS может быть выполнен с возможностью работать во множестве различных режимов кодирования и/или битовых скоростей. Кроме того, пространственный кодек метаданных может

10

предложить различные режимы кодирования и/или битовые скорости кодирования. Пространственный кодек метаданных использует статистическое кодирование, которое обычно приводит к непостоянной скорости передачи. Это означает, что фактически используемая битовая скорость обычно ниже, чем заданная целевая битовая скорость. Для некоторых кадров это отклонение от номинальной битовой скорости может быть

15

меньшим, и для некоторых других кадров оно может быть большим. Чтобы декодер 120 мог правильно декодировать переданный битовый поток 101, могут быть обеспечены точный режим кодирования и битовая скорость, используемая кодером 110. Для статистически закодированного фрагмента метаданных точно используемая битовая скорость может не потребоваться, поскольку используемые

20

коды Хаффмана являются непрерывными и уникально декодируемыми. Тем не менее, приемнику битового потока 101 может быть предоставлено количество битов, используемых для кодирования кадра (или суперкадра 400). Это желательно, например, если декодеру 120 требуется пропустить несколько принятых кадров, не имея необходимости декодировать эти кадры. В настоящем документе была описана

25

структура суперкадра, который поддерживает следующие функции:

- Декодирование полного кадра;
 - Декодирование только части, необходимой для моно-воспроизведения;
 - Извлечение информации о длине из кадра 400, чтобы поместить ее во вторичный формат (ISOBMFF), который обеспечивает и/или нуждается в этой информации;
- 30 - В конкатенации кадров 400 только частичное декодирование с эффективным пропуском первых нескольких кадров;
- Если существует битовая ошибка, поиск начала следующего кадра (повторная синхронизация);
 - Быстрое и эффективное определение битовой скорости без необходимости
- 35 декодировать кадры;
- Редактирование кадров (замена метаданные или части кадров EVS);
 - Быстрая операция продвижения вперед без декодирования кадров;
 - Поддержка эффективного переноса полезных нагрузок данных кодека постоянной и переменной длины.

35

40 Кроме того, описана структура суперкадра, содержащего все необходимые сигнальные элементы, чтобы

- указывать режимы кодека EVS (в том числе битовые скорости), используемые для кодирования EVS N сигналов каналов понижающего микширования. По умолчанию может быть $N=4$, это означает, что существуют 4 канала понижающего микширования

45

кодека EVS: W, X', Y', Z';

- указывать выбранный режим работы поддерживаемого метаданными кодека;
- указывать битовую скорость метаданных при высоком разрешении без значительных накладных расходов;

- обеспечивать возможность оповещать о потенциальных будущих расширениях.

Некоторые из сигнальных элементов суперкадра 400 могут не изменяться часто в течение сеанса кодирования или даже являться статическими. Некоторые другие сигнальные элементы, такие как битовая скорость метаданных, могут изменяться для каждого суперкадра. По этой причине некоторые сигнальные элементы обеспечиваются только по условию внутри полосы в суперкадре 400 (такие как поле 402 CI). Если они обеспечены, эти сигнальные элементы могут быть динамически адаптированы на основе суперкадра. Существует также возможность поддерживать эти сигнальные элементы статическими и обеспечивать их только один раз, например, как внеполосное сообщение. Сигнальные элементы также могут быть полудинамическими, в этом случае они обеспечиваются только внутри полосы в некоторых суперкадрах.

Относительно сигнализации битовой скорости метаданных основная проблема состоит в том, что количество требуемых битов (или байтов) на суперкадр 400 может изменяться в относительно большом диапазоне. Сигнализация только максимального возможного количества битов на кадр может оставить неиспользованным относительно большое количество битов в случае, если статистический код значительно короче, чем максимальная длина. С другой стороны, обеспечение непосредственного сигнального элемента для показателя фактически используемого количества битов (или байтов) в суперкадре 400 потребовало бы относительно большого количества сигнальных битов. В настоящем документе описана схема, которая поддерживает минимальное количество сигнальных битов для фактически используемого количества битов (или байтов) в суперкадре 400, при этом по-прежнему позволяя покрывать относительно большой диапазон возможных битовых скоростей метаданных.

С точки зрения системы, суперкадры 400 поддерживаемого метаданными кодека EVS формируются в головном узле кодирования. Это может быть сервер в сети, имеющий доступ к незакодированным иммерсивным аудиоданным или аудиоданным виртуальной реальности (VR). Это также может быть мобильный телефон, захватывающий иммерсивные аудиосигналы. Закодированные кадры 400 могут быть вставлены в файл, который загружается на принимающий терминал или передается в соответствии с текущим протоколом, таким как стандарт DASH (Dynamic Adaptive Streaming over HTTP, динамическая адаптивная потоковая передача по HTTP) или RTSP/RTP (Real-Time Streaming Protocol/Real-time Transport Protocol, протокол потоковой передачи в реальном времени/транспортный протокол в реальном времени). Если закодированные суперкадры 400 сохранены в файле, суперкадры 400 могут быть вставлены в файл, отформатированный в соответствии с ISO/BMFF. В случае, если некоторая информация конфигурации является статической, и в случае, если она не передается как часть суперкадра 400, она может быть вместо этого обеспечена от стороны кодирования стороне декодирования внеполосными средствами, такими как протокол описания сеанса (SDP).

Схемы, изложенные в общих чертах в настоящем документе, могут использовать кодек EVS как базовый кодек и могут обеспечить многорежимные сообщения/многоскоростные сообщения (тип кадра) внутри полосы в суперкадре 400 или вне полосы, например, с использованием SDP. Это может быть объединено с инфраструктурой многорежимного кодирования иммерсивных метаданных, которая может быть эффективно сконфигурирована с помощью набора параметров конфигурации, которые также могут быть переданы внутри полосы или вне полосы. Кроме того, существует возможность объединить многорежимное кодирование иммерсивных метаданных со схемой, обеспечивающей ассоциированные максимальные

битовые скорости (или количество битов в кадре/суперкадре) внутри полосы или вне полосы.

Структура суперкадра, описанная в настоящем документе, сигнализировать о фактическом используемом размере поля метаданных как о максимальном количестве (которое опционально сообщается вне полосы) минус параметр корректировки, 5 показатель для которого передается как часть каждого суперкадра 400. Кодирование параметров корректировки предпочтительно выполняется с непостоянным размером шага, что позволяет покрывать увеличенный диапазон возможных корректировок с использованием сокращенного количества сигнальных битов для параметров 10 корректировки. Кроме того, непостоянные размеры шага параметра корректировки могут быть предусмотрены с использованием аппроксимирующей модели распределения полной длины статистического кода метаданных. Это позволяет минимизировать количество неиспользованных битов в поле метаданных и тем самым минимизировать накладные расходы. Кроме того, накладные расходы для битовой скорости метаданных 15 (размер) может быть сообщен с поддержанием минимального количества неиспользованных битов в поле метаданных. Таким образом, полная битовая скорость передачи сокращается.

Информация конфигурации (CI) в поле 402 CI может относиться к выбранным типам кадра EVS для EVS-кодирования четырех сигналов каналов понижающего 20 микширования W, X', Y', Z'. Информация конфигурации может также относиться (i) к выбранному режиму работы поддерживаемого метаданными кода EVS, FOA или HIQ; (ii) к битовой скорости метаданных SPAR в случае операции HIQ; (iii) к битовой скорости метаданных коэффициента предсказания в случае операции FOA. Показатель того, является ли информация конфигурации (1) динамической и обеспеченной внутри полосы 25 вместе с полезной нагрузкой; (2) полудинамической и обеспеченной внутри полосы вместе с предыдущей полезной нагрузкой; или (3) статической и обеспеченной вне полосы как шестнадцатеричная строка вместе с атрибутом кодека из набора адаптации стандарта DASH.

Режим FOA (амбиофония первого порядка) представляет собой режим работы с 30 низкой битовой скоростью (например, на уровне ~128 Кбит/с), который полагается на метаданные прогнозирующего коэффициента. FOA обычно представляет относительно ограниченное качество вследствие относительно низкого пространственного разрешения. Режим HIQ (High Immersive Quality, высокое иммерсивное качество) представляет собой режим работы со средней и высокой скоростью (например, на 35 уровне 128-512 Кбит/с). Он полагается на метаданные SPAR и способен предложить очень высокое иммерсивное качество, поскольку нацелен на воссоздание первоначального сигнала SR.

Фиг. 5 показывает способ 500 для формирования битового потока 101, в котором битовый поток 101 содержит последовательность суперкадров 400 для 40 последовательности (базовых) кадров иммерсивного аудиосигнала 111. Сигнал 111 иммерсивной аудиоинформации (IA) может содержать сигнал представления звукового поля (SR), который может описывать звуковое поле в опорной позиции. Опорная позиция может являться позицией прослушивания и/или позицией захвата микрофона. Сигнал SR может содержать множество каналов (или форм сигнала) для множества 45 разных направлений прибытия звукового поля в опорной позиции. В качестве альтернативы или в дополнение, сигнал 111 IA может содержать один или более аудиообъектов и/или многоканальный сигнал.

Сигнал 111 IA, в частности сигнал SR, содержащийся в сигнале IA, может содержать

или может являться сигналом амбиофонии порядка L , где L больше или равно 1. В качестве альтернативы или в дополнение, сигнал SR может представлять формат "пчелиного улья" (BH) с множеством направлений прибытия, размещенных в множестве разных колец на сфере вокруг опорной позиции. Множество колец может содержать среднее кольцо, верхнее кольцо, нижнее кольцо и/или зенит. В качестве альтернативы или в дополнение, сигнал SR может представлять промежуточный пространственный формат, называемый ISF, в частности формат ISF, определенный в технологии Dolby Atmos.

Следовательно, сигнал 111 IA может содержать множество разных каналов. Каждый канал, содержащийся в сигнале 111 IA, обычно содержит последовательность отсчетов аудиоданных для последовательности моментов времени или для последовательности кадров. Другими словами, "сигналы", описанные в настоящем документе, обычно содержат последовательность отсчетов аудиоданных для соответствующей последовательности моментов времени или кадров (например, на временном расстоянии 20 мс или меньше).

Способ 500 может содержать извлечение одного или более аудиообъектов из сигнала 111 IA. Аудиообъект обычно содержит сигнал объекта (с последовательностью отсчетов аудиоданных для соответствующей последовательности моментов времени или кадров). Кроме того, аудиообъект обычно содержит метаданные 202 объекта, указывающие позицию аудиообъекта. Позиция аудиообъекта может изменяться во времени, в результате чего метаданные 202 объекта для аудиообъекта могут указывать последовательность позиций для последовательности моментов времени или кадров.

Кроме того, способ 500 может содержать определение разностного сигнала на основе сигнала 111 IA и на основе одного или более аудиообъектов. Разностный сигнал может описывать первоначальный сигнал IA, из которого один или более аудиообъектов 103, 303 были извлечены и/или удалены. Разностный сигнал может являться сигналом SR, содержащимся в сигнале 111 IA. В качестве альтернативы или в дополнение, разностный сигнал может содержать или может являться многоканальным аудиосигналом и/или множеством аудиосигналов. В качестве альтернативы или в дополнение, разностный сигнал может содержать множество аудиообъектов в местоположениях или позициях фиксированного объекта (например, аудиообъекты, которые назначены конкретным динамикам определенного размещения динамиков).

Кроме того, способ 500 может содержать формирование и/или обеспечение микшированного с понижением сигнала на основе сигнала 111 IA (например, с использованием модуля 210 понижающего микширования). Количество каналов микшированного с понижением сигнала обычно меньше, чем количество каналов сигнала 111 IA. Кроме того, способ 500 может содержать определение метаданных 205 совместного кодирования или SPAR, которые обеспечивают возможность повышающего микширования микшированного с понижением сигнала (т.е. одного или более сигналов 203 каналов понижающего микширования) с получением сигналов объектов одного или более воссозданных аудиообъектов для соответствующего одного или более аудиообъектов. Кроме того, метаданные 205 совместного кодирования или SPAR могут обеспечить возможность повышающего микширования микшированного с понижением сигнала с получением воссозданного разностного сигнала для соответствующего разностного сигнала.

Микшированный с понижением сигнал, содержащий один или более сигналов 203 каналов понижающего микширования, метаданные 205 SPAR и метаданные объекта 202, может быть вставлены в битовый поток 101. В частности, способ 500 может

содержать выполнение кодирования формы сигнала микшированного с понижением сигнала, чтобы обеспечить закодированные аудиоданные 206 для последовательности кадров одного или более сигналов 203 каналов понижающего микширования 203.

Кодирование формы сигнала может быть выполнено, например, с использованием кодирования методом улучшенных речевых услуг (EVS). Кроме того, способ 500 может содержать выполнение статистического кодирования метаданных 205 SPAR и/или метаданных 202 объектов одного или более аудиообъектов, чтобы обеспечить (закодированные) метаданные 207 для вставки в битовый поток 101.

Способ 500 может содержать повторяющуюся для последовательности суперкадров 400 вставку 501 закодированных аудиоданных 206 для одного или более (в частности нескольких) кадров (например, для двух или более кадров) одного или более сигналов 203 каналов понижающего микширования, полученных из иммерсивного аудиосигнала 111, в поля 411, 421, 412, 422 данных суперкадра 400. (Базовый) кадр сигнала 203 канала понижающего микширования может охватывать 20 мс сигнала 203 канала понижающего микширования. Суперкадр 400 может охватывать кратное число длины (базового) кадра, например, 40 мс.

Кроме того, способ 500 может содержать вставку 502 метаданных 202, 205 (в частности закодированных метаданных 207) для восстановления одного или более (в частности нескольких) кадров иммерсивного аудиосигнала 111 из закодированных аудиоданных 206 в (одно) поле 403 метаданных суперкадра 400. Таким образом, суперкадр 400 может обеспечить метаданные 202, 205 для одного или более (в частности нескольких) кадров одного или более сигналов 203 каналов понижающего микширования, тем самым обеспечивая возможность для эффективной передачи сигнала 111 А.

В частности, кадр сигнала 203 канала понижающего микширования может быть сформирован с использованием многорежимного и/или многоскоростного речевого кодека или аудиокодека. Кроме того, метаданные 202, 205 могут быть сформированы с использованием многорежимной и/или многоскоростной схемы иммерсивного кодирования метаданных. Информация конфигурации, указывающая операции многорежимного и/или многоскоростного речевого кодека или аудиокодека (который использовался для сигнала 203 канала понижающего микширования) и/или операции многорежимной и/или многоскоростной схемы иммерсивного кодирования метаданных может содержаться в поле 402 информации конфигурации (текущего) суперкадра 400, может содержаться в поле 402 информации конфигурации предыдущего суперкадра 400 из последовательности суперкадров 400, или может быть передана с использованием внеполосной схемы сигнализации. В результате этого может быть обеспечена эффективная и гибкая схема кодирования иммерсивного аудиосигнала 111.

Суперкадр 400 может содержать закодированные аудиоданные 206, ассоциированные с множеством сигналов 203 каналов понижающего микширования. Закодированные аудиоданные 206 кадра первого сигнала 203 канала понижающего микширования могут быть сформированы с использованием первого экземпляра многорежимного и/или многоскоростного речевого кодека или аудиокодека. Кроме того, закодированные аудиоданные 206 кадра второго сигнала 203 канала понижающего микширования могут быть сформированы с использованием второго экземпляра многорежимного и/или многоскоростного речевого кодека или аудиокодека, причем первый и второй экземпляры многорежимного и/или многоскоростного речевого кодека или аудиокодека могут отличаться. Информация конфигурации (содержащаяся в текущем суперкадре 400, предыдущем суперкадре 400 или переданная вне полосы) может указывать операцию

первого и второго экземпляров (в частности каждого экземпляра) многорежимного и/или многоскоростного речевого кодека или аудиокодека. Посредством выполнения этого гибкость и эффективность для кодирования иммерсивного аудиосигнала 111 могут быть дополнительно увеличены.

5 Другими словами, способ 500 может содержать вставку закодированных аудиоданных 206 для одного или более кадров первого сигнала 203 канала понижающего микширования и второго сигнала 203 канала понижающего микширования, полученных из иммерсивного аудиосигнала 111, в одно или более первых полей 411, 421 данных и в одно или более вторых полей 412, 422 данных суперкадра 400, соответственно. Первый
10 сигнал 203 канала понижающего микширования может быть закодирован с использованием первого (аудио или речевого) кодера, и второй сигнал канала понижающего микширования может быть закодирован с использованием второго (аудио или речевого) кодера. Первый и второй кодеры могут отличаться или могут управляться с использованием разной конфигурации. Кроме того, способ 500 может
15 содержать обеспечение информации конфигурации относительно первого кодера и второго кодера в суперкадре 400, в предыдущем суперкадре 400 из последовательности суперкадров 400, или с использованием внеполосной схемы сигнализации. Посредством выполнения этого гибкость и эффективность для кодирования иммерсивного аудиосигнала 111 могут быть дополнительно увеличены.

20 Способ 500 может содержать вставку поля 401 заголовка в суперкадр 400. Поле 401 заголовка может указывать размера поля 403 метаданных суперкадра 400, тем самым обеспечивая возможность гибкой адаптации размера суперкадра 400 к переменной длине метаданных 207 (закодированных посредством статистического кодирования и/или кодирования без потерь).

25 Поле 403 метаданных может представлять максимальный возможный размер (который может быть указан, например, в опциональном поле 402 информации конфигурации суперкадра 400). Поле 401 заголовка может указывать значения корректировки, и размер поля 403 метаданных суперкадра 400 может соответствовать
30 максимальному возможному размеру минус значение корректировки, тем самым обеспечивая возможность точной и эффективной сигнализации размера поля 403 метаданных.

Поле 401 заголовка может содержать индикатор размера (например, значение корректировки) для размера поля 403 метаданных. Индикатор размера может представлять разное разрешение или размер шага (относительно интервалов размера)
35 для разных диапазонов размера поля 403 метаданных. Разрешение и/или размер шага индикатора размера могут зависеть от статистического распределения размера (статистически закодированных) метаданных. Посредством обеспечения индикатора размера с переменным разрешением эффективность битовой скорости для сигнализации о размере поля 403 метаданных может быть повышена.

40 Поле 401 заголовка может указывать, содержит ли суперкадр 400 поле 402 информации конфигурации. Другими словами, поле 401 заголовка может указывать наличие поля 402 информации конфигурации. Поле 402 информации конфигурации может быть вставлено в суперкадр 400 только в случае необходимости (например, если конфигурация кодера сигнала 111 IA изменилась). В результате этого может быть
45 повышена эффективность битовой скорости последовательности суперкадров 400.

Поле 401 заголовка может указывать, что поле 402 информации конфигурации не присутствует в текущем суперкадре 400. Способ 500 может содержать передачу информации конфигурации в предыдущем суперкадре 400 из последовательности

суперкадров 400 или использование внеполосной схемы сигнализации. В результате этого информация конфигурации (который по меньшей мере является статической во времени) может быть передана эффективным образом.

5 В качестве альтернативы или в дополнение, поле 401 заголовка может указывать, содержит ли суперкадр 400 поле 404 расширения для дополнительной информации относительно иммерсивного аудиосигнала 111. В результате этого структура суперкадра может быть гибко адаптирована к будущим расширениям.

10 Способ 500 может содержать вставку поля 402 информации конфигурации в суперкадр 400 (в случае необходимости). Поле 402 информации конфигурации может указывать количество сигналов 203 каналов понижающего микширования, содержащихся в полях 411, 421, 412, 422 данных суперкадра 400. В качестве альтернативы или в дополнение, поле 402 информации конфигурации может указывать порядок сигналов представления звукового поля, содержащихся в сигнале 111 IA. В результате этого может быть закодировано и передано множество различных типов сигналов 111 IA (со множеством
15 различных типов сигналов SR).

Поле 402 информации конфигурации может указывать максимальный возможный размер поля 403 метаданных. В качестве альтернативы или в дополнение, поле 402 информации конфигурации может указывать типа кадра и/или режим кодирования, используемые для кодирования каждого одного или более сигналов 203 каналов
20 понижающего микширования. Предоставление этой информации обеспечивает возможность использования разных схем кодирования для кодирования сигнала 111 IA.

Закодированные аудиоданные 206 кадра сигнала 203 канала понижающего микширования могут быть сформированы с использованием многорежимного и/или
25 многоскоростного речевого кодека или аудиокодека. В качестве альтернативы или в дополнение, (закодированные) метаданные 207 могут быть сформированы с использованием многорежимной и/или многоскоростной схемы иммерсивного кодирования метаданных. В результате сигнал 111 IA может быть закодирован с относительно высоким качеством и на относительно низких скоростях передачи данных.

30 Суперкадр 400 из последовательности суперкадров 400 может содержать по меньшей мере часть элемента данных, переданного с использованием протокола передачи, в частности стандартов DASH, RTSP или RTP, или сохраненного в файле в соответствии с форматом хранения, в частности ISO/BMFF. Другими словами, битовый поток 101, содержащий последовательности суперкадров 400, может использовать один или более
35 элементов данных протокола передачи или формата хранения. Тем самым обеспечивая возможность передачи или хранения битового потока 101 эффективным и надежным образом.

Фиг. 6 показывает блок-схему последовательности этапов иллюстративного способа 600 для получения данных относительно иммерсивного аудиосигнала 111 из битового
40 потока 101. Битовый поток 101 содержит последовательность суперкадров 400 для последовательности кадров иммерсивного аудиосигнала 111. В предпочтительном примере несколько (базовых) кадров сигнала 111 IA содержатся в одном суперкадре 400. Следует отметить, что все признаки, описанные в контексте способа 500 для формирования битового потока 101, аналогичным образом применимы для способа
45 600 для получения данных из битового потока 101.

Сигнал 111 IA может содержать сигнал SR, многоканальный сигнал и/или один или более аудиообъектов. Аспекты и/или признаки, которые описаны в контексте способа 500 и/или в контексте устройства 110 кодирования, также аналогичным и/или

взаимодополняющим образом применимы для способа 600 и/или для устройства 120 декодирования (и наоборот).

Способ 600 содержит повторяющееся для последовательности суперкадров 400 извлечение 601 закодированные аудиоданных 206 для одного или более (в частности
5 нескольких) кадров одного или более сигналов 203 каналов понижающего микширования, полученных из иммерсивного аудиосигнала 111, из полей 411, 421, 412, 422 данных суперкадра 400. Кроме того, способ 600 содержит извлечение 602 (закодированных) метаданных 207 для восстановления одного или более (в частности
10 нескольких) кадров иммерсивного аудиосигнала 111 из закодированных аудиоданных 206 из поля метаданных 403 суперкадра 400.

Способ 600 может содержать получение одного или более воссозданных аудиообъектов из закодированных аудиоданных 206 и из метаданных 207 (в частности из метаданных 202 объектов). Как указано выше, аудиообъект обычно содержит сигнал
15 объекта и метаданные 202 объекта, которые указывают (изменяющуюся во времени) позицию аудиообъекта. Кроме того, способ 600 может содержать получение воссозданного разностного сигнала из закодированных аудиоданных 206 и из метаданных 202, 205. Один или более воссозданных аудиообъектов и воссозданный разностный сигнал могут описывать и/или могут указывать сигнал 111 IA. В частности,
20 данные (такие как порядок сигнала SR, содержащегося в сигнале 111 IA) могут быть извлечены из битового потока 101, что позволяет определить воссозданный сигнал 121 IA, причем воссозданный сигнал 121 IA является приближением первоначального сигнала 111 IA.

Как указано выше, способ 600 для получения данных относительно иммерсивного аудиосигнала 111 из битового потока 101 может содержать признаки, соответствующие
25 способу 500 для формирования битового потока 101. В частности, способ 600 может содержать извлечение поля 401 заголовка из заданного суперкадра 400. Размер поля 403 метаданных заданного суперкадра 400 может быть получен из поля 401 заголовка.

Размер поля 403 метаданных может быть указан, как изложено в общих чертах в контексте способа 500. Поле 403 метаданных может представлять максимальный
30 возможный размер, и поле 401 заголовка может указывать значения корректировки, причем размер 403 поля метаданных суперкадра 400 может соответствовать максимальному возможному размеру минус значение корректировки. В частности, поле 401 заголовка может содержать индикатор размера для размера поля 403 метаданных, причем индикатор размера может представлять разное разрешение для
35 разных диапазонов размера поля 403 метаданных. В результате этого размер поля 403 метаданных может быть сигнализирован эффективно относительно битовой скорости.

Способ 600 может содержать определение на основе поля 401 заголовка, содержит ли суперкадр 400 поле 402 информации конфигурации, и/или присутствует ли поле 402 информации конфигурации в суперкадре 400. Если поле 402 информации конфигурации
40 отсутствует, информация конфигурации, которая была обеспечена в предыдущем суперкадре 400 и/или которая была обеспечена вне полосы, может использоваться для обработки одного или более кадров одного или более сигналов 203 каналов понижающего микширования, содержащихся в суперкадре 400. Если поле 402 информации конфигурации присутствует, то информация конфигурации, содержащаяся
45 в суперкадре 400, может использоваться для обработки одного или более кадров одного или более сигналов 203 каналов понижающего микширования, содержащихся в суперкадре 400.

Кроме того, способ 600 может содержать определение на основе поля 401 заголовка,

содержит ли суперкадр 400 поле 404 расширения для дополнительной информации относительно иммерсивного аудиосигнала 111, тем самым обеспечивая эффективные и гибкие средства для передачи информации в битовом потоке 101.

5 Способ 600 может содержать извлечение поля 402 информации конфигурации из суперкадра 400. Кроме того, способ 600 может содержать определение на основе поля 402 информации конфигурации количества сигналов 203 каналов понижающего микширования, представленных посредством полей 411, 421, 412, 422 суперкадра 400, тем самым обеспечивая возможность точной обработки одного или более кадров
10 одного или более сигналов 203 каналов понижающего микширования, содержащихся в суперкадре 400.

Кроме того, способ 600 может содержать определение на основе поля 402 информации конфигурации максимального возможного размера поля 403 метаданных.

Кроме того, способ 600 может содержать определение на основе поля 402 информации конфигурации порядка иммерсивного аудиосигнала 111 для обеспечения возможности
15 точного воссоздания сигнала 111 IA.

Способ 600 также может содержать определение на основе поля 402 информации конфигурации типа кадра и/или режима кодирования, используемых для кодирования каждого одного или более сигналов каналов понижающего микширования, тем самым обеспечивая возможность точной обработки одного или более кадров одного или более
20 сигналов 203 каналов понижающего микширования, содержащихся в суперкадре 400.

Различные иллюстративные варианты осуществления настоящего изобретения могут быть реализованы в аппаратных средствах или специализированных схемах, программном обеспечении, логических схемах или любой их комбинации. Некоторые аспекты могут быть реализованы в аппаратных средствах, в то время как другие аспекты
25 могут быть реализованы в программно-аппаратном обеспечении или программном обеспечении, которое может быть исполнено контроллером, микропроцессором или другим вычислительным устройством. В общем случае настоящее раскрытие также охватывает устройство, подходящее для выполнения описанных выше способов, например, устройство (пространственный рендерер), имеющее память и процессор,
30 соединенный с памятью, причем процессор выполнен с возможностью исполнять инструкции и выполнять способы в соответствии с вариантами осуществления раскрытия.

Хотя различные аспекты иллюстративных вариантов осуществления настоящего изобретения проиллюстрированы и описаны как блок-схемы, блок-схемы
35 последовательности этапов, или с использованием некоторого другого иллюстрированного представления, очевидно, что блоки, устройство, системы, методики или способы, описанные в настоящем документе, могут быть реализованы, в качестве неограничивающих примеров, в аппаратных средствах, программном обеспечении, программно-аппаратном обеспечении, специализированных схемах или логических
40 схемах, аппаратных средствах или контроллере общего назначения, других вычислительных устройствах или их некоторой комбинации.

Кроме того, различные блоки, показанные в блок-схемах последовательности этапов, могут быть рассмотрены как этапы способа и/или как операции, которые являются результатом работы компьютерного программного кода, и/или как множество
45 соединенных элементов логической схемы, построенных для выполнения соответствующей функции (функций). Например, варианты осуществления настоящего изобретения включают в себя компьютерный программный продукт, содержащий компьютерную программу, физически воплощенную на машиночитаемом носителе, в

котором компьютерная программа, содержащая программные коды, выполнена с возможностью выполнять описанные выше способы.

В контексте раскрытия машиночитаемый носитель может являться любым материальным носителем, который может содержать или хранить программу для использования системой или устройством для исполнения инструкций или в связи с ними. Машиночитаемый носитель может являться машиночитаемым сигнальным носителем или машиночитаемым носителем хранения информации. Машиночитаемый носитель может включать в себя, но без ограничения, электронные, магнитные, оптические, электромагнитные, инфракрасные или полупроводниковые системы или устройства или любую подходящую их комбинацию. Более конкретные примеры машиночитаемого носителя информации включают в себя электрическое соединение, имеющее один или более проводов, дискету портативного компьютера, жесткий диск, оперативное запоминающее устройство (ОЗУ; RAM), постоянное запоминающее устройство (ПЗУ; ROM), стираемое программируемое постоянное запоминающее устройство (СППЗУ; EPROM, или флэш-память), оптоволокну, переносное ПЗУ на компактном диске (CD-ROM), оптическое запоминающее устройство, магнитное запоминающее устройство или любую подходящую их комбинацию.

Компьютерный программный код для выполнения способов настоящего изобретения может быть написан на любой комбинации из одного или более языков программирования. Эти компьютерные программные коды могут быть обеспечены процессору компьютера общего назначения, специализированного компьютера или другого программируемого устройства обработки данных, в результате чего программные коды при их исполнении процессором компьютера или другого программируемого устройства обработки данных реализуют функции/операции, определенные в блок-схемах последовательности этапов и/или блок-схемах. Программный код может исполняться полностью на компьютере, частично на компьютере, как автономный пакет программного обеспечения, частично на компьютере и частично на удаленном компьютере, или полностью на удаленном компьютере или сервере.

Кроме того, хотя операции изображены в конкретном порядке, это не должно быть истолковано как требование, что такие операции должны быть выполнены в конкретном показанном порядке или в последовательном порядке, или что все проиллюстрированные операции должны быть выполнены для достижения желаемых результатов. При определенных обстоятельствах может иметь преимущество многозадачная и параллельная обработка. Аналогичным образом, хотя в приведенных выше обсуждениях содержатся некоторые подробности конкретной реализации, они должны быть истолкованы не как ограничения объема изобретения или заявленного предмета, а как описания признаков, которые могут быть заданы для конкретных вариантов осуществления конкретных изобретений. Некоторые признаки, которые описаны в этом описании в контексте отдельных вариантов осуществления, также могут быть реализованы в комбинации в единственном варианте осуществления. С другой стороны, различные признаки, которые описаны в контексте единственного варианта осуществления, также могут быть реализованы в нескольких вариантах осуществления отдельно или в любой подходящей комбинации.

Следует отметить, что описание и чертежи лишь иллюстрируют принципы предложенных способов и устройств. Таким образом, будет очевидно, что специалисты в области техники смогут создавать различные структуры, которые, хотя явно не описаны и не показаны в настоящем документе, воплощают принципы изобретения и

включены в пределы его сущности и объема. Кроме того, все примеры, приведенные в настоящем документе, преимущественно явно предназначены иметь место лишь в целях обучения, чтобы помочь читателю в понимании принципов предложенных способов и устройств, и концепций, внесенных изобретателями в развитие области техники, и должны быть истолкованы как не являющиеся ограничениями для таких специальным образом приведенных примеров и условий. Кроме того, предполагается, что все утверждения в настоящем документе, излагающие принципы, аспекты и варианты осуществления изобретения, а также их конкретные примеры, охватывают его эквиваленты.

10

(57) Формула изобретения

1. Способ (500) формирования битового потока (101), в котором битовый поток (101) содержит последовательность суперкадров (400) для последовательности кадров иммерсивного аудиосигнала (111), причем способ (500) содержит этапы, на которых, повторяющимся образом для последовательности суперкадров (400):

15

- вставляют (501) закодированные аудиоданные (206) для двух или более кадров одного или более сигналов (203) каналов понижающего микширования, полученных из иммерсивного аудиосигнала (111), в поля (411, 421, 412, 422) данных суперкадра (400); и

20

- вставляют (502) метаданные (202, 205) для воссоздания двух или более кадров иммерсивного аудиосигнала (111) из закодированных аудиоданных (206) в поле (403) метаданных суперкадра (400).

2. Способ (500) по п. 1, причем

25

- способ (500) содержит этап, на котором вставляют поле (401) заголовка в суперкадр (400); и

- поле (401) заголовка указывает размер поля (403) метаданных суперкадра (400), причем опционально:

- поле (403) метаданных представляет максимальный возможный размер;

- поле (401) заголовка указывает значение корректировки; и

30

- размер поля (403) метаданных суперкадра (400) соответствует максимальному возможному размеру минус значение корректировки.

3. Способ (500) по п. 2, в котором

- поле (401) заголовка содержит индикатор размера для размера поля (403) метаданных; и

35

- индикатор размера представляет разное разрешение для разных диапазонов размера для размера поля (403) метаданных,

причем опционально:

- метаданные (202, 205) для восстановления одного или более кадров иммерсивного аудиосигнала (111) представляют статистическое распределение размера для размера метаданных (202, 205); и

40

- разрешение индикатора размера зависит от распределения размера метаданных (202, 205).

4. Способ (500) по любому из предыдущих пунктов, причем

- способ (500) содержит этап, на котором вставляют поле (401) заголовка в суперкадр (400); и

45

- поле (401) заголовка указывает на то, содержит ли суперкадр (400) поле (402) информации конфигурации, и/или

- поле (401) заголовка указывает наличие поля (402) информации конфигурации, и/

или

- поле (401) заголовка указывает на то, содержит ли суперкадр (400) дополнительное поле (404) для дополнительной информации относительно иммерсивного аудиосигнала (111).

- 5 5. Способ (500) по любому из предыдущих пунктов, причем
- способ (500) содержит этап, на котором вставляют поле (402) информации конфигурации в суперкадр (400); и
 - поле (402) информации конфигурации указывает количество сигналов (203) каналов понижающего микширования, представленных полями (411, 421, 412, 422) данных суперкадра (400), и/или
 - 10 - поле (402) информации конфигурации указывает максимальный возможный размер поля (403) метаданных, и/или
 - поле (402) информации конфигурации указывает порядок сигнала представления звукового поля, содержащегося в иммерсивном аудиосигнале (111), и/или
 - 15 - поле (402) информации конфигурации указывает тип кадра и/или режим кодирования, используемые для кодирования каждого из упомянутого одного или более сигналов (203) каналов понижающего микширования.

6. Способ (500) по любому из предыдущих пунктов, в котором закодированные аудиоданные (206) кадра сигнала (203) канала понижающего микширования
- 20 закодированы с использованием кодера улучшенных речевых услуг.

7. Способ (500) по любому из предыдущих пунктов, в котором суперкадр (400) составляет по меньшей мере часть элемента данных, переданного с использованием протокола передачи, в частности стандартов DASH, RTSP или RTP, или сохранен в файле в соответствии с форматом хранения, в частности ISO/BMFF.

- 25 8. Способ (500) по любому из предыдущих пунктов, причем
- поле (401) заголовка указывает отсутствие поля (402) информации конфигурации;
 - и
 - способ (500) содержит этап, на котором передают информацию конфигурации в предыдущем суперкадре (400) из последовательности суперкадров (400) или используют
 - 30 внеполосную схему сигнализации.

9. Способ (500) по любому из предыдущих пунктов, причем способ (500) содержит этапы, на которых

- вставляют закодированные аудиоданные (206) для одного или более кадров первого сигнала (203) канала понижающего микширования и второго сигнала (203) канала
- 35 понижающего микширования, полученных из иммерсивного аудиосигнала (111), в одно или более первых полей (411, 421) данных и одно или более вторых полей (412, 422) данных суперкадра (400), соответственно; причем первый сигнал (203) канала понижающего микширования закодирован с использованием первого кодера, и причем второй сигнал (203) канала понижающего микширования закодирован с использованием
- 40 второго кодера; и
- обеспечивают информацию конфигурации относительно первого кодера и второго кодера в данном суперкадре (400), в предыдущем суперкадре (400) из последовательности суперкадров (400) или с использованием внеполосной схемы сигнализации.

- 45 10. Способ (500) по любому из предыдущих пунктов, причем способ (500) содержит этапы, на которых

- извлекают один или более аудиообъектов из иммерсивного аудиосигнала (111) IA; причем аудиообъект содержит сигнал объекта и метаданные (202) объекта, указывающие

позицию аудиообъекта;

- определяют разностный сигнал (201) на основе сигнала (111) IA и на основе упомянутого одного или более аудиообъектов;

5 - обеспечивают микшированный с понижением сигнал на основе сигнала (111) IA, в частности, таким образом, что количество сигналов (230) каналов понижающего микширования микшированного с понижением сигнала меньше, чем количество сигналов каналов сигнала (111) IA;

10 - определяют метаданные (205) совместного кодирования для обеспечения возможности повышающего микширования микшированного с понижением сигнала в один или более воссозданных сигналов аудиообъекта, соответствующих упомянутому одному или более аудиообъектам и/или в воссозданный разностный сигнал (311), соответствующий разностному сигналу (201);

15 - выполняют кодирование формы сигнала микшированного с понижением сигнала, чтобы обеспечить закодированные аудиоданные (206) для последовательности кадров упомянутого одного или более сигналов (203) каналов понижающего микширования;

и

20 - выполняют статистическое кодирование метаданных (205) совместного кодирования и метаданных (202) объектов упомянутого одного или более аудиообъектов, чтобы обеспечить метаданные (202, 205) для вставки в поля (403) метаданных последовательности суперкадров (400).

11. Способ (600) получения данных относительно иммерсивного аудиосигнала (111) из битового потока (101), причем битовый поток (101) содержит последовательность суперкадров (400) для последовательности кадров иммерсивного аудиосигнала (111), причем способ (600) содержит этапы, на которых, повторяющимся образом для

25 последовательности суперкадров (400):

- извлекают (601) закодированные аудиоданные (206) для двух или более кадров одного или более сигналов каналов (203) понижающего микширования, полученных из иммерсивного аудиосигнала (111) из полей (411, 421, 412, 422) данных суперкадра (400); и

30 - извлекают (602) метаданные (202, 205) для восстановления двух или более кадров иммерсивного аудиосигнала (111) из закодированных аудиоданных (206) из поля (403) метаданных суперкадра (400).

12. Устройство (110) кодирования, выполненное с возможностью формировать битовый поток (101), причем битовый поток (101) содержит последовательность суперкадров (400) для последовательности кадров иммерсивного аудиосигнала (111), причем устройство (110) кодирования выполнено с возможностью, повторяющимся образом для последовательности суперкадров (400):

40 - вставлять закодированные аудиоданные (206) для двух или более кадров одного или более сигналов (203) каналов понижающего микширования, полученных из иммерсивного аудиосигнала (111), в поля (411, 421, 412, 422) данных суперкадра (400);

и

- вставлять метаданные (202, 205) для восстановления двух или более кадров иммерсивного аудиосигнала (111) из закодированных аудиоданных (206) в поле (403) метаданных суперкадра (400).

45 13. Устройство (120) декодирования, выполненное с возможностью получать данные относительно иммерсивного аудиосигнала (111) из битового потока (101), причем битовый поток (101) содержит последовательность суперкадров (400) для последовательности кадров иммерсивного аудиосигнала (111), причем устройство (120)

декодирования выполнено с возможностью, повторяющимся образом для последовательности суперкадров (400):

- 5 - извлекать закодированные аудиоданные (206) для двух или более кадров одного или более сигналов (203) каналов понижающего микширования, полученных из иммерсивного аудиосигнала, из полей (411, 421, 412, 422) данных суперкадра (400); и
- извлекать метаданные (202, 205) для восстановления двух или более кадров иммерсивного аудиосигнала (111) из закодированных аудиоданных (206) из поля (403) метаданных суперкадра (400).

10

15

20

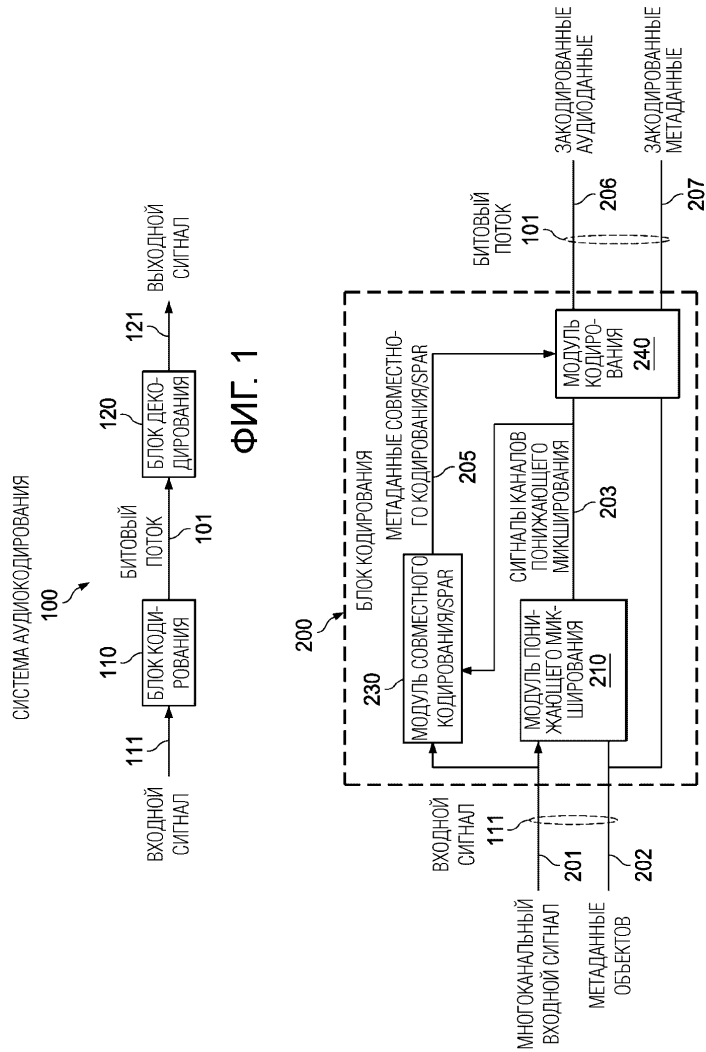
25

30

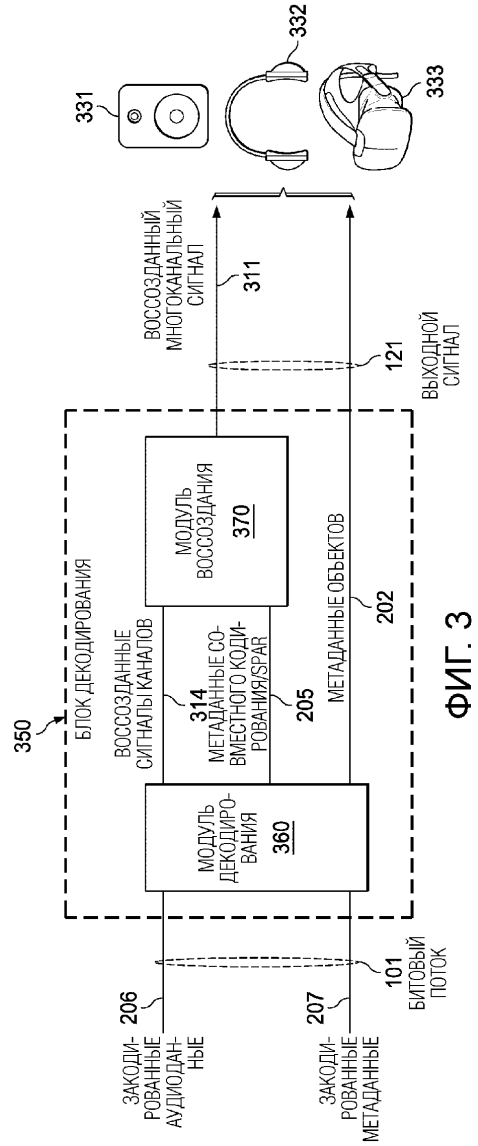
35

40

45

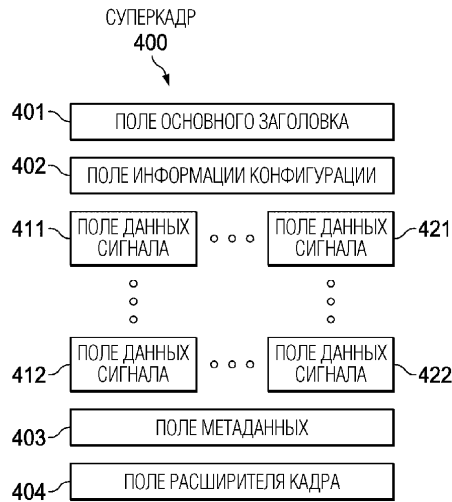


2/3



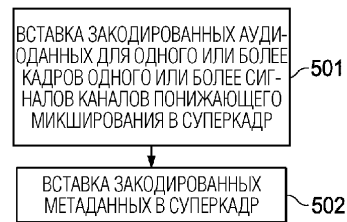
ФИГ. 3

3/3



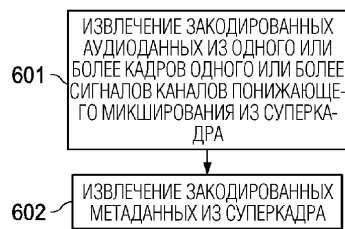
ФИГ. 4

ФОРМИРОВАНИЕ
БИТОВОГО ПОТОКА
500



ФИГ. 5

ВЫЯВЛЕНИЕ ДАННЫХ ИЗ
БИТОВОГО ПОТОКА
600



ФИГ. 6